

# Machine Learning Methods in Climate Finance: a Systematic Review



By Andrés Alonso-Robisco, José Manuel Carbó and José Manuel Marqués, Banco de España

*Keywords: climate change, machine learning, review, climate finance*

*JEL codes: Q54, G10*

*Preventing the materialization of climate change is one of the main challenges of our time. The involvement of the financial sector is a fundamental pillar in this task, which has led to the emergence of a new field in the literature, climate finance. In turn, the use of Machine Learning (ML) as a tool to analyze climate finance is on the rise, due to the need to use big data to collect new climate-related information and model complex non-linear relationships. We propose a review of the academic literature to assess how ML is enabling climate finance to scale up. Using topic modeling (Latent Dirichlet Allocation) we uncover representative thematic clusters. This allows us to statistically identify granular application domains where ML is playing a significant role in climate finance literature. Finally, we do an analysis highlighting publication trends and key analytical insights.*

## 1. The role of Machine Learning in Climate Finance

One characteristic of climate finance literature is how fragmented the research is. This is not only a bibliographic concern, as it also makes it difficult to join efforts from different academic profiles in order to develop specific research. In a literature review performed by Cunha et al. (2021) the authors conclude that *“it is difficult to identify what constitutes the field and what differentiates it from traditional finance”*, due to the poor theorization of the concept of “sustainability”, an opinion shared by several others like Capelle-Blancard & Monjon (2012), Zhang et al. (2019), Talan & Sharma (2019), Liang & Renneboog (2021) and Giglio et al. (2021). Therefore, for the purpose of this study we will rely exclusively on the definition provided by Giglio et al. (2021) as *“the tools of financial economics designed for valuing and managing risk which can help society assess and respond to climate change”*.<sup>1</sup>

Another feature of the field of climate finance is the difficulty researchers face in conducting robust empirical analysis. To name two key challenges:

First, the increasing amount of climate data available and the uncertainty about its reliability. The good news is that the sudden explosion of micro-level datasets offers an unparalleled insight into the inner workings of the economy and financial systems. The bad news is that datasets are increasingly more complex to deal with (López de Prado, 2019). As an example with implications for climate finance, we can point out, for example, the great variation that exists between the temperature predictions made by the 20 research teams that report to the Intergovernmental Panel on Climate Change (IPCC), with data for over 100 years (Monteleoni et al. 2011). In fact, some of the most interesting datasets in climate finance are not only high-dimensional, but also unstructured, including news articles, voice recordings or satellite images, which along with the complexity of the phenomena they measure, means that many of these datasets are beyond the grasp of econometric analysis.

Second, big datasets may allow for more flexible relationships between the variables than simple linear models. It has been largely recognized that ML techniques such as decision trees, support vector machines, neural nets, deep learning, and so on, may allow for more effective ways to model complex financial and economic relationships (Varian 2014, Athey 2018, Athey & Imbens 2019). The key advantage and one common feature of many ML methods is that they use data driven model selection, treating the data generating process (DGP) as unknown, allowing researches to deal with large datasets without imposing restrictive assumptions.

These type of problems increase the complexity when making inferences about the real climate (Stephenson et al. 2012) and its relationship with the economy. In fact, Diaz-Rainey et al. (2017) conclude that methodological constraints could explain previous lack of climate finance research in top finance and business journals.

In this context, the use of Machine Learning (ML) by researchers and experts seems to be justified, since it is particularly well suited to deal with these issues. This motivates us to understand the potential of this technology to assist climate finance to scale-up. To this purpose, in Alonso-Robisco et al. (2023) we systematically assemble a corpus of relevant articles in climate finance which harness ML to come up with a solution, and we estimate a Latent Dirichlet Allocation (LDA) model to uncover latent topics in this literature, offering academics, market experts and policy makers a structured guide to assess publication trends, emerging topics, knowledge gaps, and types of models used, aiming to facilitate a better knowledge of this innovative field.

---

<sup>1</sup>Although we will use from now on the term climate finance, we acknowledge that three concepts are used indistinctively in the academic literature, namely green finance, climate finance and carbon finance (Zhang et al., 2019).

## 2. Methodology and data

Our final collection of documents adds up to 217 research articles, from which we extract the abstracts, which will comprise the corpus of our study. Our goal will be to discover the hidden or latent (unobservable) topics in the corpus of documents (observable), using a ML-technique, Latent Dirichlet Allocation or LDA (Blei et al. 2003). This will help us understand documents analyzing the presence of words. Often the term “topic” is used in a technical, statistical sense, but ultimately the last phase of any topic modeling approach involves expert analysis to uncover through inspection a more economic meaningful name. In addition, we rank the topics according to their prevalence (Sievert & Shirley, 2014), which we find to be a convenient visualization tool for the exploration and presentations of the topics.

We use the LDA model to locate 10 latent topics present in the articles. Once the model finds all 10 topics, we can inspect the keywords in each topic in order to label it. As shown in **Figure 1**, as a result of that inspection, we found the following research areas in climate finance and ML: (i) natural hazards, (ii) biodiversity, (iii) carbon markets, (iv) agricultural risk, (v) ESG factors & investing, (vi) energy economics, and (vii) climate data. We discard three topics because we find that their composition is either mainly comprised of methodological terms or repetitive with other topics.

**Figure 1: Topics and tokens (word stems)**

LDA Topic	1		2		3		4		5	
	Tokens	Probability of token	Tokens	Probability of token	Tokens	Probability of token	Tokens	Probability of token	Tokens	Probability of token
	activ	0.026	risk	0.026	sustain	0.018	biodivers	0.012	carbon	0.028
	csr	0.017	flood	0.023	chang	0.011	financi	0.012	soil	0.024
	valu	0.012	algorithm	0.012	studi	0.01	green	0.011	invest	0.016
	flood	0.01	predict	0.011	method	0.009	base	0.011	predict	0.011
	storag	0.01	price	0.01	financ	0.009	develop	0.01	power	0.011
	correl	0.01	term	0.009	polic	0.009	invest	0.01	polic	0.011
	corpor	0.01	differ	0.007	research	0.008	resourc	0.01	emiss	0.01
	signific	0.01	impact	0.007	topic	0.008	cost	0.009	studi	0.01
	avail	0.009	provid	0.007	inform	0.008	conserv	0.008	result	0.009
	base	0.009	studi	0.007	train	0.008	research	0.008	forecast	0.009
<b>Econ. Label</b>	<i>*discarded*</i>		<b>Natural hazards</b>		<i>*discarded*</i>		<b>Biodiversity</b>		<i>*discarded*</i>	
LDA Topic	6		7		8		9		10	
	Tokens	Probability of token	Tokens	Probability of token	Tokens	Probability of token	Tokens	Probability of token	Tokens	Probability of token
	carbon	0.026	chang	0.027	esg	0.07	energi	0.03	compani	0.02
	price	0.023	crop	0.024	invest	0.024	predict	0.019	corpor	0.019
	market	0.02	yield	0.019	rat	0.022	emiss	0.016	report	0.018
	emiss	0.018	futur	0.014	social	0.021	carbon	0.015	financi	0.018
	firm	0.016	agricultur	0.013	portfolio	0.021	forest	0.012	disclosur	0.017
	green	0.015	adapt	0.011	compani	0.013	result	0.01	csr	0.016
	financ	0.013	product	0.011	perform	0.013	chang	0.009	perform	0.014
	paper	0.012	hybrid	0.011	stock	0.013	use	0.008	risk	0.013
	stock	0.01	project	0.01	risk	0.012	random	0.008	relat	0.012
	sector	0.01	suitabl	0.01	score	0.012	impact	0.008	environment	0.011
<b>Econ. Label</b>	<b>Carbon markets</b>		<b>Agricultural risk</b>		<b>ESG factors &amp; investing</b>		<b>Energy economics</b>		<b>Climate data</b>	

### 3. Publication Trends and Analysis

From our results, we extract some stylized facts.

#### **Finding #1 "ML covers most climate finance topics"**

We observe that ML is currently applied for a majority of topics related to climate change in finance. For instance, we identify relevant studies covering five out of the seven topics listed in Kumar et al. (2022), and four out of six topics identified in Debrah, Darko, and Chan (2022) which could serve as a benchmark survey describing the field of sustainable finance as a whole.

#### **Finding #2 "From physical risk to market-related topics"**

While ML was initially applied to physical risks problems, like weather and natural hazards forecasting, and issues related to energy economics, a relevant number of studies are now using ML for responsible investing, ESG factors and measuring corporate's compliance with climate data regulatory disclosures.

#### **Finding #3 "Mature vs emerging research topics"**

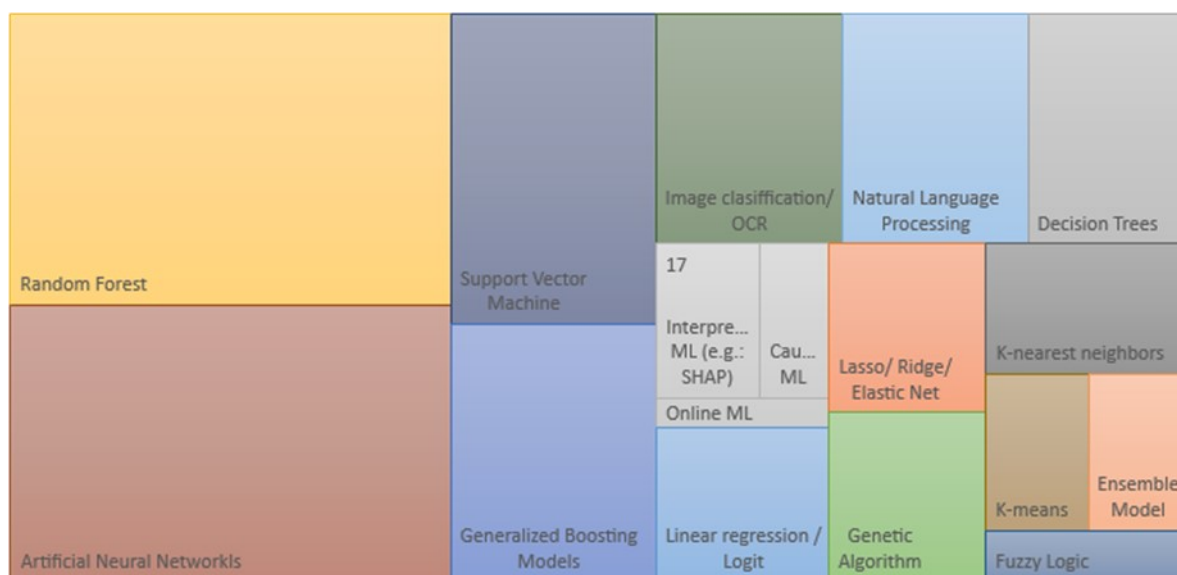
As evidenced by the higher proportions of peer-reviewed publications versus working papers format, topics like Agricultural risk, Natural hazards, Biodiversity, and Energy economics are more mature. Though, Climate data and ESG factors & investing are emerging, younger topics.

#### **Finding #4 "Undeserved attention to physical risk in Economic journals"**

We identify publications in very heterogeneous knowledge domains, like journals from environmental sciences, computer sciences, or economics and finance journals. We observe that Economic and Finance journals still pay more attention to topics related to CSR and Transitions risks, lagging behind other scientific journals that publish more work on Physical risk and its socioeconomic impact.

#### **Finding #5 "Artificial neural networks do not always lead"**

Overall, as shown in **Figure 2**, Random forests and Artificial Neural Networks are the mostly used methods, but for instance, in Physical risk we appreciate a strong usage of image recognition tools, usually associated with the need to handle newly available (unstructured) data from remote sensing, text, and satellites, relying therefore heavily on Convolutional Neural Networks and Random forests. However, in Transition risks, Artificial Neural Networks dominate within our subset of documents, usually benefiting from access to a big datasets to study energy-related topics. Finally, in CSR, interestingly the access to bigger amounts of data is still challenging, and the requirements on the specifications of the models and the interpretability of results push towards more linear techniques like Ridge and/or Elastic net regularization in multiple types of regressions, together with a notable share of studies introducing techniques from explainable AI (xAI) like Shapley values (Lundberg and Lee 2017).

**Figure 2: Type of ML model used. Overall.**

#### 4. Green AI

Finally, we feel responsibly obliged to bring to this discussion the other side of the impact of ML on climate change, as well. New technologies do not only bring us opportunities. Kaack et al. (2020) explain ways in which AI and ML can be detrimental to efforts addressing climate change, warning of those uses that might harm our planet. AI or AI-driven technologies can become pollutants and net emitters of greenhouse emissions, depending on the types of applications and the circumstances of their deployment. For example, remote sensing algorithms for satellite image analysis can be used to gather information on agricultural productivity, but can also be used to accelerate oil and gas exploration. Self-driving cars can make driving more efficient, but they could also increase the amount people drive. And finally, ML include computationally expensive programming, which is an energy intensive activity. This final concern has minted the term “Green AI”, referring to responsible and low carbon intensive coding and good practices relating the training and deployment of complex algorithms in the academic industry (e.g.: Strubell et al. 2019, or Hershovich et al. 2022). ■

## References

- Alonso-Robisco, A., Carbó, J. M., & Marqués, J. M. (2023). Machine Learning methods in climate finance: a systematic review. *Documentos de Trabajo/Banco de España*, 2310.
- Athey, S. (2018). The impact of machine learning on economics. *The economics of artificial intelligence: An agenda*, 507-547.
- Athey, S., & Imbens, G. (2019). Machine learning methods economists should know about. *arXiv preprint arXiv:1903.10075*.
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, 3 (Jan), 993-1022.
- Capelle-Blancard, G., & Monjon, S. (2012). Trends in the literature on socially responsible investment: Looking for the keys under the lamppost. *Business ethics: a European review*, 21(3), 239-250.
- Castle, J. L., & Hendry, D. F. (2022). Econometrics for modelling climate change. In *Oxford Research Encyclopedia of Economics and Finance*.
- Debrah, C., Darko, A., & Chan, A. P. C. (2022). A bibliometric-qualitative literature review of green finance gap and future research directions. *Climate and Development*, 1-24.
- Diaz-Rainey, I., Robertson, B., & Wilson, C. (2017). Stranded research? Leading finance journals are silent on climate change. *Climatic Change*, 143(1), 243-260.
- Giglio, S., Kelly, B., & Stroebel, J. (2021). Climate finance. *Annual Review of Financial Economics*, 13, 15-36.
- Hershcovich, D., Webersinke, N., Kraus, M., Bingler, J. A. Leippold, M., (2021). Towards Climate Awareness in NLP Research. *arXiv:2205.05071*.
- Kumar, S., Sharma, D., Rao, S., Lim, W. M., & Mangla, S. K. (2022). Past, present, and future of sustainable finance: insights from big data analytics through machine learning of scholarly research. *Annals of Operations Research*, 1-44.
- Liang, H., & Renneboog, L. (2021). Corporate Social Responsibility and Sustainable Finance. In *Oxford Research Encyclopedia of Economics and Finance*.
- López de Prado, M. (2019). Beyond econometrics: A roadmap towards financial machine learning. Available at SSRN 3365282.
- Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30.
- Monteleoni, C., Schmidt, G. A., Saroha, S., & Asplund, E. (2011). Tracking climate models. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 4(4), 372-392.
- Strubell, E., Ganesh, A., & McCallum, A. (2019). Energy and policy considerations for deep learning in NLP. *arXiv preprint arXiv:1906.02243*.
- Sievert, C., & Shirley, K. (2014, June). LDAvis: A method for visualizing and interpreting topics. In *Proceedings of the workshop on interactive language learning, visualization, and interfaces* (pp. 63-70).
- Talan, G., & Sharma, G. D. (2019). Doing well by doing good: A systematic review and research agenda for sustainable investment. *Sustainability*, 11(2), 353.
- Varian, H. R. (2014). Big data: New tricks for econometrics. *Journal of Economic Perspectives*, 28(2), 3-28.
- Zhang, D., Zhang, Z., & Managi, S. (2019). A bibliometric analysis on green finance: Current status, development, and future directions. *Finance Research Letters*, 29, 425-430.

## About the authors

**Andrés Alonso Robisco** joined Banco de España in 2019 as senior economist in the Financial Innovation Division where he analyses the latest trends in financial innovation. Specifically, he studies the impact of machine learning on credit risk modelling, and different topics related to climate finance innovation. Previously he had been working in the Single Resolution Board (SRB), an agency of the European Commission, and beforehand, in the treasury and capital markets teams of Instituto de Crédito Oficial (ICO), the Spanish financial agency. He has published articles in several journals like *International Review of Financial Analysis*, and *Financial Innovation*.

**Jose Manuel Carbó** joined Banco de España in 2019 as a senior economist in the Financial Innovation Division, where he analyses the latest trends in financial innovation. Specifically, he studies the impact of machine learning on credit risk modelling, and different topics related to cryptocurrencies and financial stability. Prior to this, he was a consultant in ScanmarQED London and Research Associate in Imperial College London. He has a PhD in economics from Universidad Carlos III de Madrid. His research interests are machine learning, financial innovation, and policy analysis. He has published articles in several journals like *Journal of Applied Statistics*, *Economics of Transportation*, *Financial Innovation*, and *Regional Science and Urban Economics*, among others.

**José Manuel Marqués** is Head the Financial Innovation and Market Infrastructures Department Bank of Spain (Banco de España). In this position, he coordinates the contribution from Bank of Spain to Infrastructures related with the TARGET Services (Target 2 Securities and Eurosystem Collateral Management System) and Market Activities Processing System. He also coordinates the analysis of the main trends transforming the financial system: Big tech and Fintech, Cryptocurrencies, CBDC, Artificial Intelligence and Machine Learning, Blockchain and DLT. Marqués has joined Bank of Spain in 1996 and has written a number of articles on financial markets and is chair of the Working Group of Green Finance in the BIS Innovation Network and represent Bank of Spain on the High Level Taskforce on the digital euro, the CEMLA Foro Fintech. He has a BA in Economics and Business from University of Zaragoza, and an MSc from University Pompeu Fabra, Barcelona.

## SUERF Publications

Find more **SUERF Policy Briefs** and **Policy Notes** at [www.suerf.org/policynotes](http://www.suerf.org/policynotes)



**SUERF** is a network association of central bankers and regulators, academics, and practitioners in the financial sector. The focus of the association is on the analysis, discussion and understanding of financial markets and institutions, the monetary economy, the conduct of regulation, supervision and monetary policy.

SUERF's events and publications provide a unique European network for the analysis and discussion of these and related issues.

**SUERF Policy Briefs (SPBs)** serve to promote SUERF Members' economic views and research findings as well as economic policy-oriented analyses. They address topical issues and propose solutions to current economic and financial challenges. SPBs serve to increase the international visibility of SUERF Members' analyses and research.

The views expressed are those of the author(s) and not necessarily those of the institution(s) the author(s) is/are affiliated with.

All rights reserved.

### Editorial Board

Ernest Gnan  
Frank Lierman  
David T. Llewellyn  
Donato Masciandaro  
Natacha Valla

SUERF Secretariat  
c/o OeNB  
Otto-Wagner-Platz 3  
A-1090 Vienna, Austria  
Phone: +43-1-40420-7206  
[www.suerf.org](http://www.suerf.org) • [suerf@oenb.at](mailto:suerf@oenb.at)