PROCEEDINGS OF SCIENCE

The impact of archival research with ESA Space Science missions

Guido De Marchi*

European Space Research and Technology Centre,
Keplerlaan 1, 2201 AZ Noordwijk, Netherlands

E-mail: gdemarchi@esa.int

We analysed more than 68,000 peer-reviewed publications derived from data collected by 25 missions within the ESA Science Programme, along with 11 additional missions in which ESA has a supporting role. These publications span nearly five decades and cover the disciplines of astronomy, planetary science, and heliophysics, with data ranging from each mission's launch year up to the end of 2021. Our study explored the temporal evolution of publication output and several bibliometric indicators, including citation counts and related indices. We also assessed the geographical distribution of contributing authors and, for ESA Member States, examined the correlation between scientific output and national financial contributions to the ESA Science Programme. Our findings suggest that scientific engagement across Member States generally aligns with their gross domestic products, although certain countries — regardless of size — demonstrate notable efficiency in translating mission data into scientific advancement. Additionally, we distinguished between publications authored by mission-affiliated scientists and those produced independently using mission data, known as "archival papers". Remarkably, archival papers account for over half of all ESA mission-based literature and show comparable citation impact to those authored by mission collaborators. This underscores the critical value of open access to and long-term preservation of scientific mission data.

Frontier Research in Astrophysics – IV (FRAPWS2024)
9–14 September 2024
Mondello, Palermo, Italy

*Speaker

https://pos.sissa.it/

PoS(FRAPWS2024)004

1. Introduction

This paper presents an analysis of the scientific impact of research based on data from missions of the European Space Agency (ESA) Science Programme. The work, which I conducted in collaboration with Arvind Parmar [1] as part of a broader study published in a recent volume [2], aims to assess the outcomes and effectiveness of ESA's investment in space science over the past five decades.

The ESA Science Programme includes a wide range of missions, beginning in the 1970s, and spanning diverse fields, not only astronomy but also planetary science and heliophysics. Among these latter are prominent missions such as Giotto [3], SOHO [4], and Rosetta [5]. The overarching objective of this study was to address a fundamental question: how can we evaluate whether a scientific mission — often involving significant financial and institutional commitment — has been successful? More importantly, how should "success" be defined in the context of scientific exploration?

A well-established method for assessing the productivity and impact of astronomical facilities involves the quantitative analysis of scholarly publications that make use of their data. In this context, we examined a comprehensive set of bibliometric indicators — such as the number of publications, citations, the h-index [6], and related metrics — to provide a data-driven evaluation of ESA's scientific output. In order to compare the performance of different missions meaningfully, we have applied consistent and objective inclusion criteria to identify relevant publications and construct comparable bibliographic datasets.

Beyond standard bibliometric analysis, we also explored the broader landscape of scientific engagement with ESA mission data. This includes examining the geographical and institutional distribution of authors, their affiliations, and the nature of their involvement with the missions. A particularly noteworthy finding is that a substantial portion of the publications based on ESA mission data — more than half — are authored by scientists who were not directly involved in the development of the instruments or the execution of the experiments. Rather, these "archival users" have accessed the mission data through publicly available archives, highlighting the enduring value and impact of open data policies in space science.

This paper presents these findings in some detail, contributing to a deeper understanding of the scientific returns of ESA missions and offering insights into how such returns can be effectively measured and possibly enhanced.

2. Searching the literature

A key component of this study was the construction of comprehensive and consistent bibliographic libraries for each ESA mission. To ensure objectivity and reproducibility in the selection process, we applied clearly defined criteria for including refereed publications in each mission's library. A paper qualifies for inclusion if it meets at least one of the following conditions:

1. Use of Mission Data: The publication must utilise data products from the mission — such as images, spectra, light curves, or derived catalogues — and perform analyses that lead to scientific conclusions.

2. Predictive Analysis: The paper presents quantitative predictions or interpretations related to the mission's expected results, often involving analysis of pre-existing data from other sources.

3. Mission Documentation: The publication provides technical descriptions of the mission, including its instruments, operational procedures, or calibration methods. These papers are typically few in number and usually published around the time of the mission's launch.

The vast majority of the publications fall under the first category, with the second category accounting for roughly 10% of the total. The third category is even less frequent. All mission bibliographies developed as part of this work are publicly available on the ESA web pages[1] and are also shared with NASA's Astrophysics Data System (ADS) [7]. Some of these bibliographies extend back nearly 50 years, covering missions such as COS–B [8] and IUE [9].

Traditionally, the identification of relevant publications has been a manual process. For many years, members of mission teams — most often the Project Scientists — have systematically reviewed the scientific literature on a monthly basis to identify qualifying papers. The ADS has been an invaluable resource in this effort, offering keyword-based search capabilities and, more recently, full-text search functionality. Despite these tools, accurate classification still requires detailed reading of each candidate paper to confirm whether it meets the inclusion criteria. This approach, while effective, is labor-intensive and time-consuming.

In recent years, advances in machine learning and natural language processing (NLP) have opened new opportunities for automating parts of this workflow. ESA now employs NLP techniques to assist in paper classification, using supervised learning algorithms trained on well-established mission-specific datasets. These training datasets consist of previously curated publication lists divided into two categories: those that use mission data (positive examples) and those that merely mention the mission without utilising its data (negative examples). These datasets serve as high-quality training inputs for developing classification algorithms.

Although only a few mission libraries currently incorporate NLP-based methods, all automated classifications are validated by a domain expert to ensure accuracy. Encouragingly, algorithms trained on one mission can often be adapted to others with only minor modifications. This is largely due to the structural and linguistic consistency of scientific literature across missions, which makes classification tasks relatively transferable.

3. Bibliometric trends

A quantitative overview of publication trends reveals significant insights into the scientific lifecycle of ESA missions. The waterfall diagram in Figure 1 illustrates the number of publications per year over the past 20 years. It demonstrates a sustained high output even after mission operations have concluded. Missions such as Herschel [10] and Planck [11], which ceased operations in 2013, continue to generate a large number of annual publications today, comparable to or even exceeding those produced during their active phases. Legacy missions such as IUE, Hipparcos [12], and ISO [13], all of which ended in the 1990s, still collectively produce around 100 papers annually.

[1] https://www.cosmos.esa.int/web/guest/mission-publications

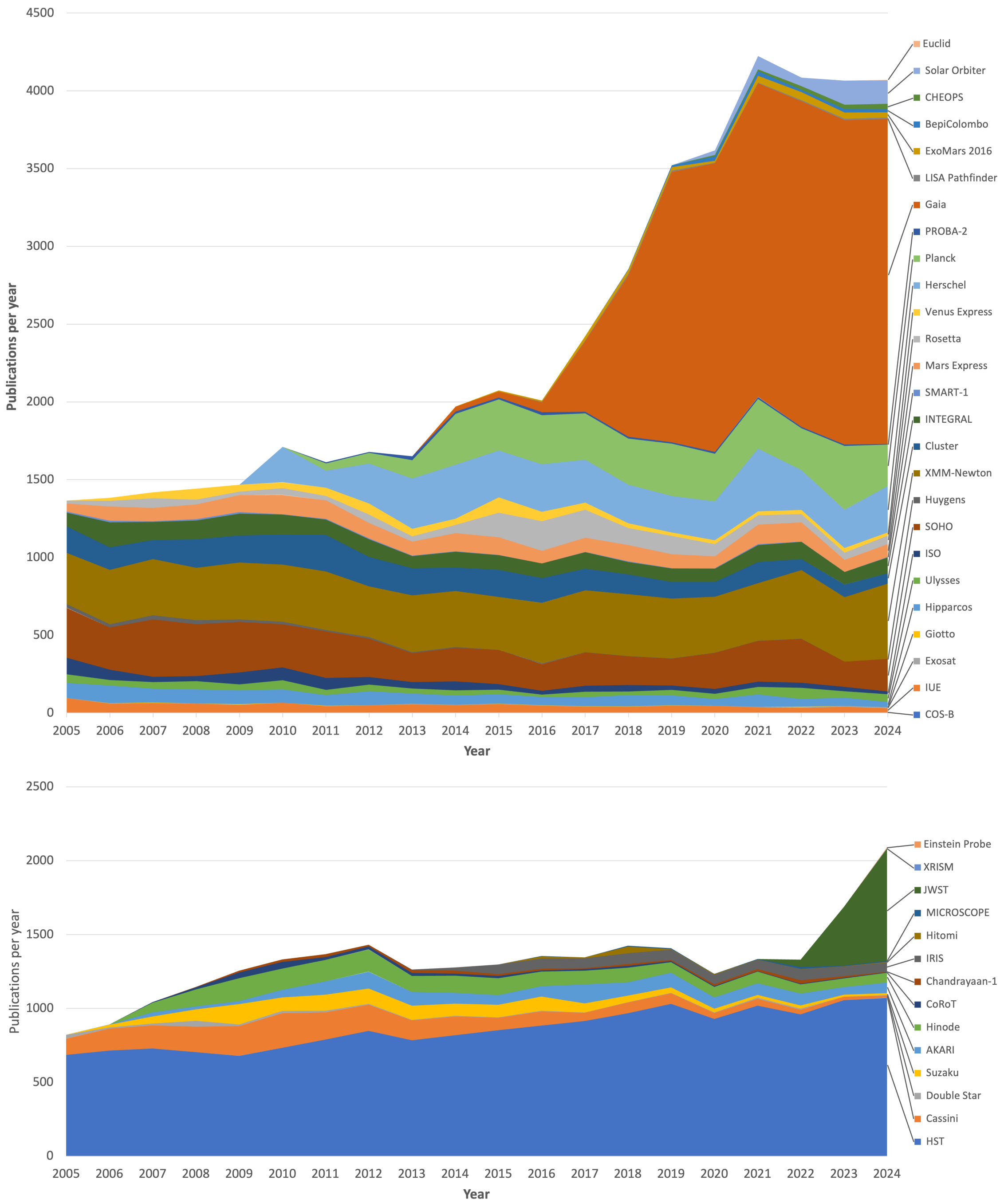

Figure 1: A "waterfall" diagram of the ESA Science Programme Mission refereed publications per year between 2005 and 2024. The missions are arranged in order of launch date. The top panel refers to missions led by ESA, while the bottom panel is for missions in which ESA is a junior partner. The large contribution from Gaia after 2016 (shown in dark orange in the top panel) is striking.

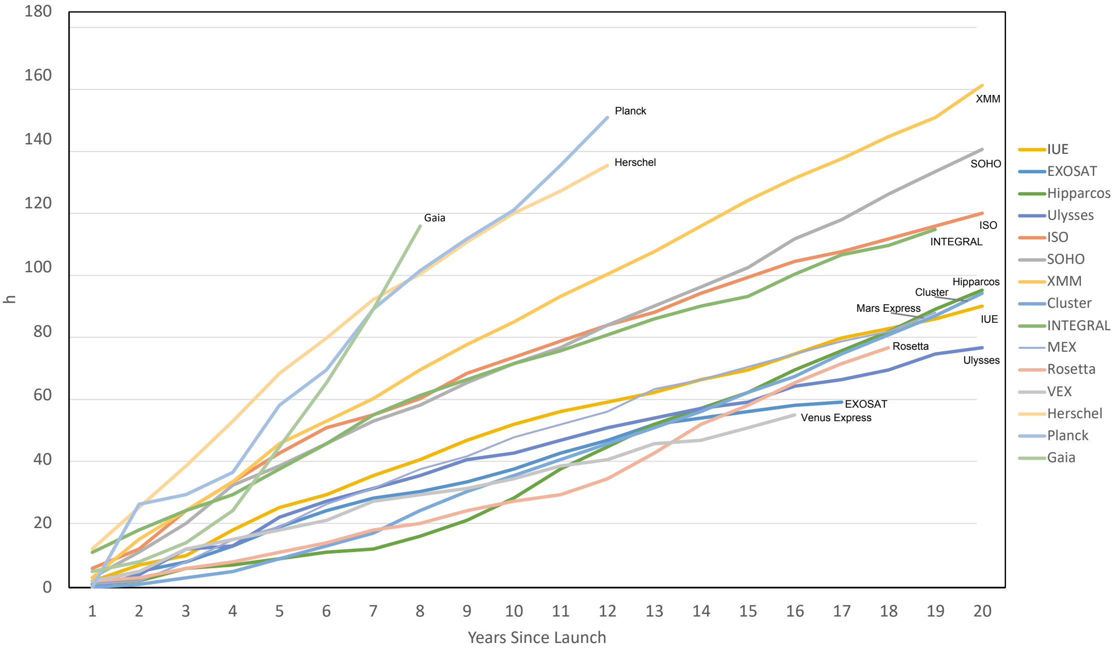

Figure 2: Evolution of the h-index metric with years after launch for a sample of ESA-led missions. Different missions are represented by different coloured lines.

Ongoing missions with long operational lifespans, including XMM–Newton [14], INTEGRAL [15], and SOHO, maintain robust scientific output, while collaborative missions like the Hubble Space Telescope (HST) — where ESA plays a supporting role — continue to yield over 1,000 publications per year. A particularly striking example is the Gaia [16] mission, which currently contributes more than 2,000 refereed publications annually. This exceptional productivity has led to a marked increase in overall publication rates across ESA missions, reflecting the growing impact of high-quality, publicly accessible data in enabling a wide range of scientific investigations.

While publication counts offer a measure of scientific productivity, they are not sufficient to compare the overall impact of different missions. This is due to the wide variety of scientific domains covered, the differing sizes of their respective research communities, and the substantial variation in mission lifespans. For example, the Huygens probe [17] operated for only a few hours during its descent to Titan, whereas Hubble has been active for more than 34 years. As a result, citation-based metrics offer a more meaningful basis for comparison, as they better reflect both the usage and influence of mission data over time.

One commonly used metric is the Hirsch index [6], or h-index, which captures both the quantity and impact of a body of work. A mission with an h-index of 30, for instance, has produced at least 30 papers that have each been cited a minimum of 30 times. This metric is widely used by review panels and funding agencies to assess individual researchers, with a value of 40 or higher often considered typical for senior scientists. By this standard, many ESA missions would be considered highly impactful: as shown in Figure 2, several missions — including SOHO, XMM–Newton, Planck, Herschel, and Gaia — have achieved h-indices exceeding 120. In the case of Gaia, the h-index has continued to rise steeply and at the time of writing (early 2025) has surpassed 150,

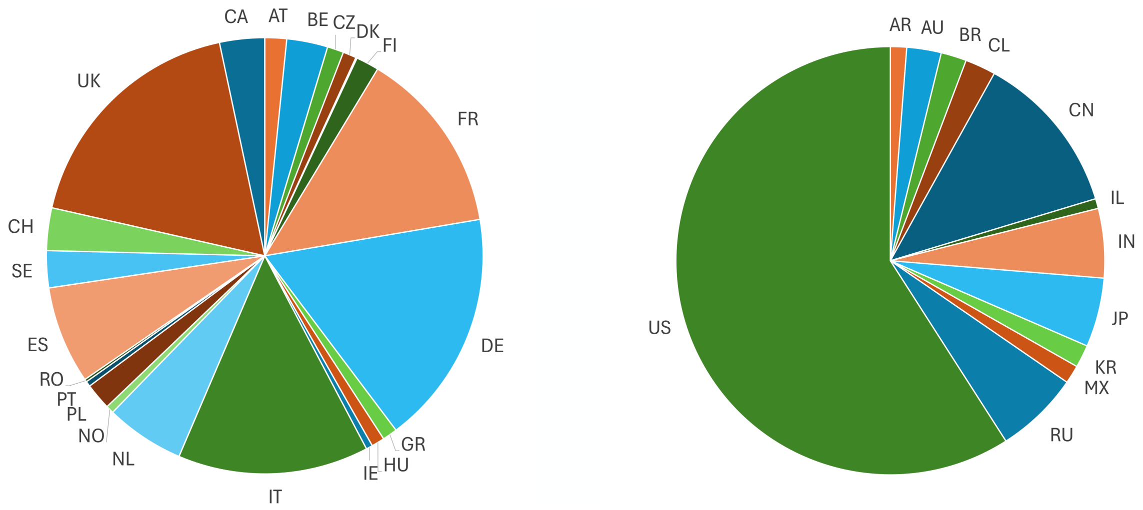

Figure 3: Distribution of the affiliations of the first authors of publications based on data from the ESA-led missions. The left-hand panel shows the distribution of the affiliations across ESA Member States, while on the right-hand side all other countries are presented.

underscoring its exceptional influence across multiple fields of astrophysics.

Many other bibliometric indicators were considered, such as the total number of citations or the number of publications with over 100 citations (see [1]), but the overall conclusions remain consistent: ESA missions have enabled a substantial and growing body of high-impact research, with Gaia currently leading in terms of citation-based scientific influence.

An obvious question to consider is whether more expensive missions result in a greater number of scientific publications, or whether it is possible to establish a correlation between a mission's cost and the volume or impact of its scientific output. In practice, however, such an analysis proves challenging, primarily due to the limited availability of comprehensive cost data. ESA typically funds and manages the development of the spacecraft, launch services, and scientific operations. In contrast, the payloads are usually provided and financed by institutions within the ESA Member States. Detailed information on these national contributions is generally not available to ESA and, as noted by [18], would in any case be difficult to compare across countries due to differences in accounting practices, funding structures, and institutional frameworks.

4. Geographical distribution

An important aspect of evaluating the scientific impact of ESA missions is understanding the geographical distribution of the resulting publications. Our study encompasses a comprehensive dataset of almost 70,000 refereed papers published between 1976 and 2022. Of these, roughly half are led by authors affiliated with institutions in ESA Member States, while the remaining half have first authors based in countries that are not part of ESA.

The right-hand panel in Figure 3 shows the distribution of papers by first-author affiliation outside of Europe. Notably, researchers based in the United States account for a substantial portion of the total output, publishing as many first-author papers as the three largest ESA contributors — Germany, France, and Italy — combined. Significant engagement with ESA data is also observed

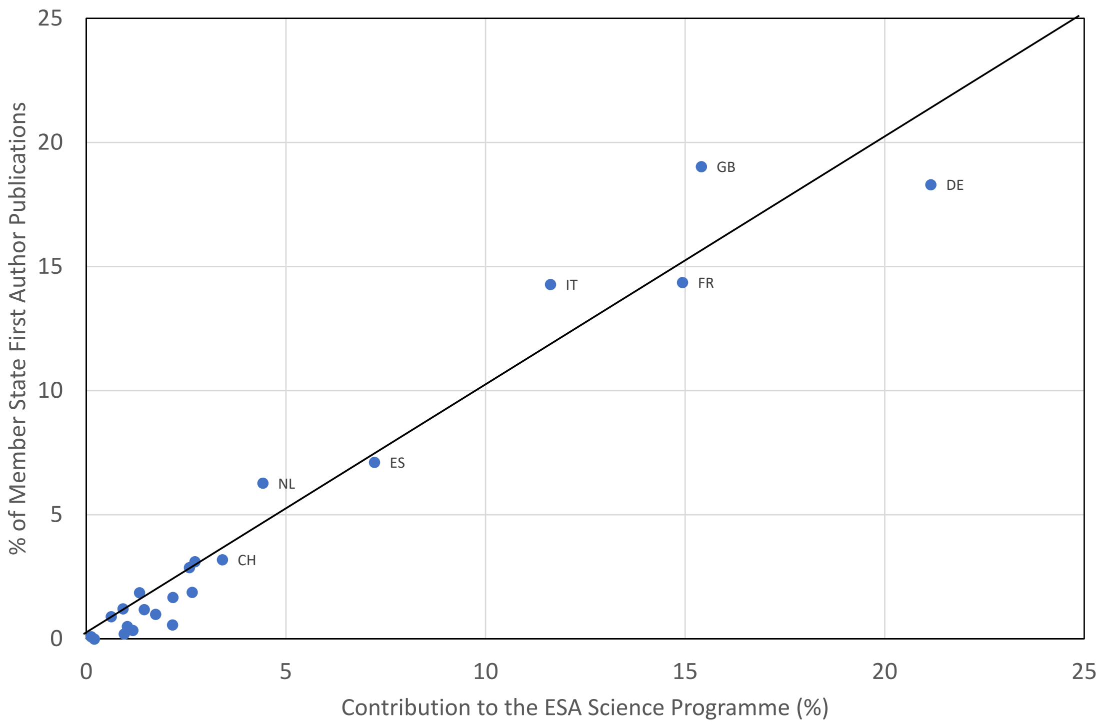

Figure 4: Percentages of first author publications for each ESA Member State as a function of the 2000–2021 financial contributions of those countries to the Science Programme. The seven countries making the largest financial contributions to the ESA Science Programme are labelled. Error bars are too small to be visible.

in countries such as China, Japan, Russia, and India, underscoring the global relevance of ESA's missions.

The left-hand panel of Figure 3 focuses on contributions within the ESA Member States. As expected, the four largest countries — France, Germany, Italy, and the United Kingdom — are the primary European users of ESA science data. However, the figure also reveals that countries such as the Netherlands and Spain produce a comparable volume of publications, despite notable differences in population size. This suggests that factors beyond national population or economic scale, such as scientific tradition and infrastructure, also play a role in scientific productivity. It is worth noting that similar trends are observed whether considering only first authors or all contributing authors.

To explore the relationship between national scientific output and financial support to ESA, Figure 4 plots the fraction of papers produced in each Member State against its relative contribution to the ESA Science Programme. This funding model, established at the founding of ESA and inherited from its predecessor, the European Space Research Organisation, mandates that contributions be proportional to each country's Gross Domestic Product (GDP). For example, Germany — Europe's largest economy — is the principal contributor to the programme.

The general trend in Figure 4 shows that countries with higher financial contributions tend to produce a greater share of publications. However, there are noteworthy deviations from this trend. In particular, the United Kingdom and Italy appear to generate more scientific output than their proportional financial input would predict, slightly outperforming larger contributors such as

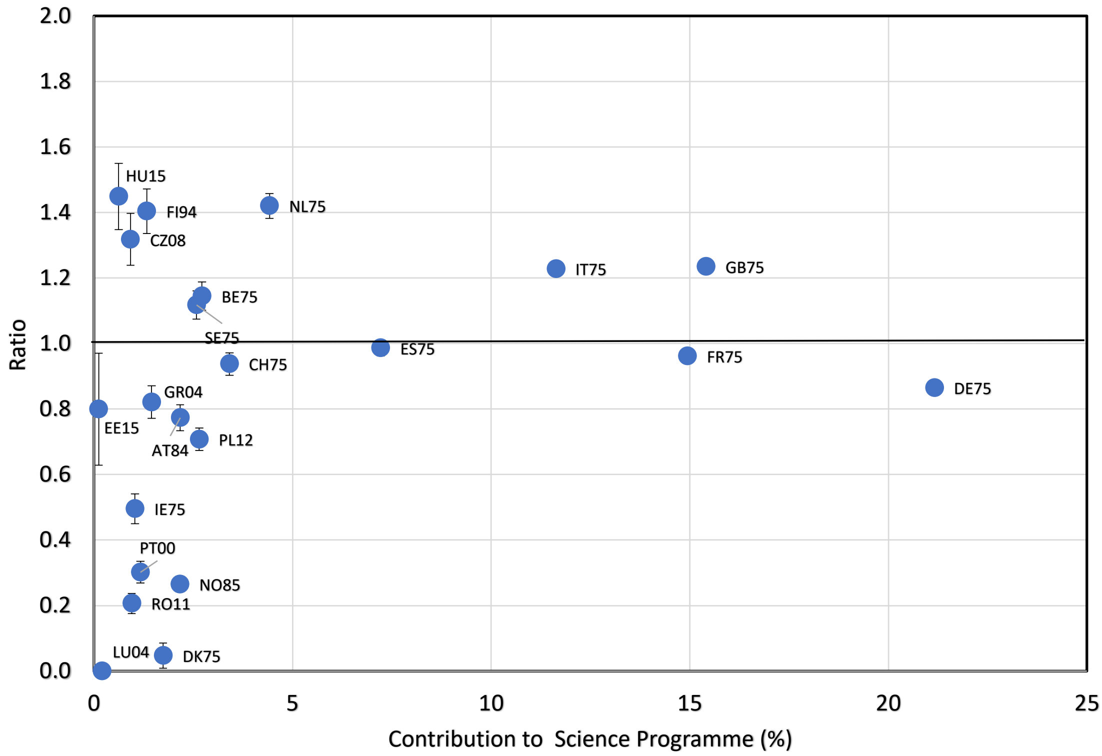

Figure 5: Ratio between the number of first-author publications affiliated with institutions in each ESA Member State and the country's average financial contribution to the ESA Science Programme over the period 2000–2021. Error bars are included for visual guidance only and are based on the assumption of Poisson statistics for publication counts. The ISO 3166 country codes are shown alongside the final two digits of the year in which each Member State joined ESA.

Germany and France.

This becomes clearer in Figure 5, which shows the ratio between each country's share of publications and its share of financial contribution. In an ideal scenario of perfectly proportional output, all data points would lie on a horizontal line at unity. While several countries — including France, Spain, and Switzerland — conform closely to this expectation, others, such as the United Kingdom and Italy, exceed it by approximately 20%, indicating a higher-than-expected return on investment in scientific terms.

Some smaller Member States such as the Netherlands, Finland, the Czech Republic, and Hungary exhibit particularly strong performance. The latter two, which joined ESA relatively recently (as indicated by the digits following their country codes in the figure), have rapidly established themselves as effective contributors to ESA's scientific output. Conversely, countries such as Denmark and Ireland, despite long-standing membership, appear to produce fewer publications relative to their financial input.

It is important to interpret these figures with nuance. While GDP-scaled contributions provide a useful reference point, many other factors influence national scientific productivity. These include strategic research priorities, the size and disciplinary focus of national scientific communities, the

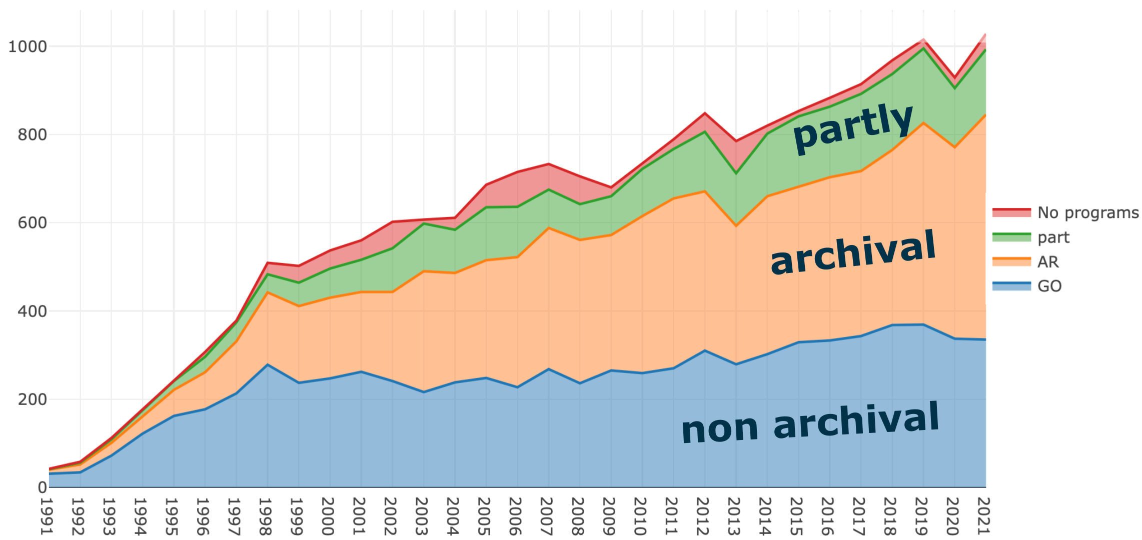

Figure 6: Growth of the fraction of archival papers based on HST observations (source: STScI Library).

level of institutional support, and historical engagement with space science. As such, the observed deviations should be considered as indicative rather than definitive assessments of performance, serving as a prompt for reflection on how ESA Member States can best leverage their participation in the Science Programme.

5. Archival Papers

A final and increasingly important aspect of evaluating the scientific impact of ESA missions concerns the role of archival data. This discussion begins with a simple but fundamental question: how do researchers perceive working with data acquired by others — specifically, data collected through observing time awarded to different principal investigators (PIs) or generated by instruments designed and operated by teams other than their own? When posed at scientific conferences, this question is typically met with broad acceptance. Indeed, in practice, a significant proportion of European scientists working with ESA mission data do so without having been directly involved in the acquisition of those data.

This introduces the concept of "archival papers", namely scientific publications that make use of mission data retrieved from public archives, without involvement in the original proposals, instruments, or experiments that generated those data. While the idea is straightforward in the context of observatories, it extends naturally to survey missions and PI-led experiments, which are common within the ESA Science Programme. In these cases, PIs have initially exclusive access to the data, but these are subsequently released to the scientific community via mission archives.

For the purposes of this study, a publication is defined as fully archival if none of the authors is the PI of the relevant proposal, instrument, or experiment that produced the data utilized in the paper. Conversely, a paper is non-archival if the PI is among the authors. In the occasional case where a publication draws upon multiple datasets, and only a subset of the corresponding PIs are co-authors, the work is classified as partly archival, although such cases are relatively rare.

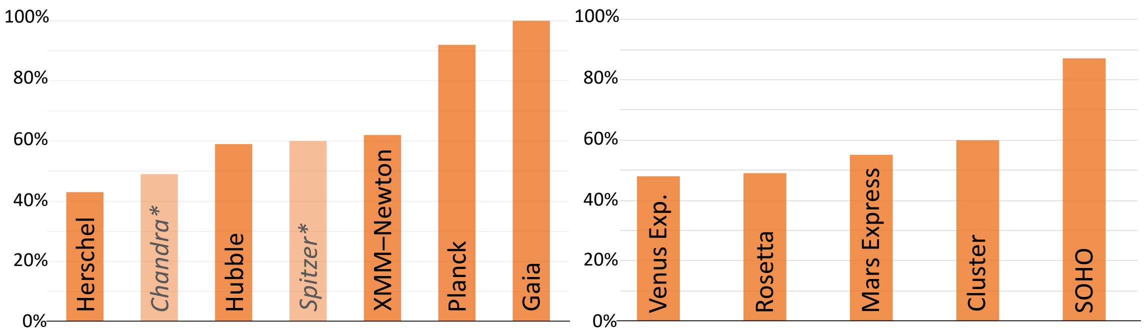

Figure 7: Cumulative fractions of archival papers for a sample of ESA Astronomy (left) and Solar System (right) missions. (*) Data about Chandra and Spitzer are from [19].

A useful benchmark in this domain is the Hubble Space Telescope, which has long been recognised for its significant volume of archival research. As illustrated in Figure 6, archival papers now constitute more than 50% of all publications using Hubble data, a figure that continues to grow steadily over time. Encouraged by this trend, we extended the analysis to all ESA missions. As shown in Figure 7, archival papers represent more than 50% of the total output across the ESA Science Programme, a result that holds true not only for astronomy but also for heliophysics and planetary science. In fact, some ESA missions exceed this threshold by a substantial margin. Among these, Planck stands out with an archival publication rate exceeding 90%, and Gaia — whose data are made publicly available immediately upon release — has an archival rate of 100% by definition. These findings are comparable with, and in some cases exceed, those for other NASA's Great Observatories, such as Chandra and Spitzer, which also hover around the 50–60% range.

Understandably, for missions where PIs initially have restricted access to the data it is to be expected that non-archival papers will dominate the scientific productivity for a certain period of time. As demonstrated in Figure 8, it typically takes several years after launch for archival publications to become the majority. However, this shift has occurred consistently across all ESA missions. The cases of XMM–Newton, Planck, and Cluster [20] clearly show a growth trajectory in which archival research steadily becomes the dominant mode of scientific output.

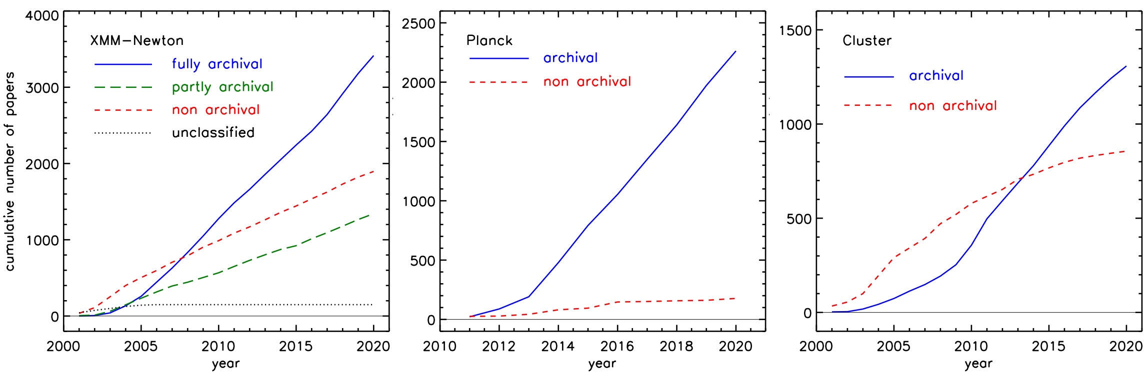

Figure 8: Cumulative distribution of refereed papers as a function of time for different categories and missions (see legend).

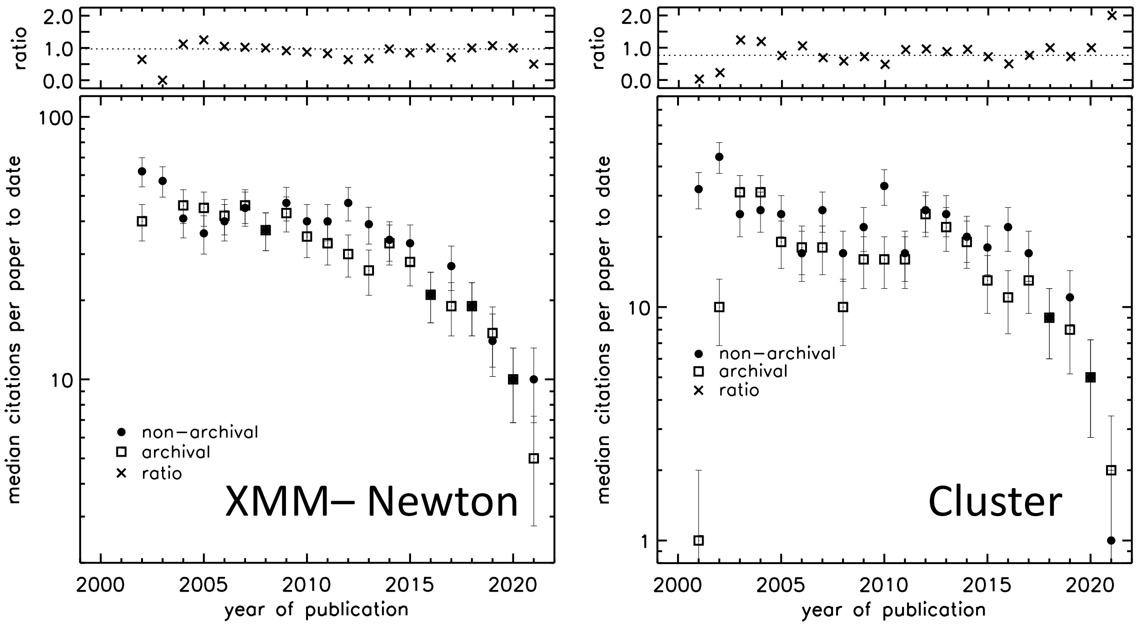

Figure 9: Median number of citations to date for two representative ESA Science Programme missions (XMM–Newton and Cluster). Non-archival papers are indicated with dots, archival papers with squares. Error bars assume Poisson statistics. Crosses in the upper panels display the ratio of archival and non-archival papers.

An important follow-up question concerns the scientific impact of these archival papers: are they cited as frequently as non-archival works? Analysis of citation data provides a clear answer. Using XMM–Newton as an example, we examined the median number of citations per paper for archival and non-archival publications across multiple years. The results, illustrated in Figure 9 (left-hand panel), show two data series: squares represent the median number of citations received by archival papers, and dots represent the same for non-archival papers, plotted by publication year. As expected, earlier papers accumulate more citations due to longer exposure times. However, the critical metric is the ratio of citations between the two types, shown by the crosses in the upper part of the figure. This ratio remains consistently close to unity across all years examined.

A similar pattern is observed for the heliophysics mission Cluster (right-hand panel), where archival and non-archival publications receive comparable citation rates. These findings indicate that archival papers are not only numerous but also impactful, receiving the same level of scientific attention as those authored by original data producers. Given that archival papers now represent the majority of the scientific output from most ESA missions (see Figure 7), one can argue that the accessibility and usability of well-curated mission archives play a fundamental role in advancing space science. The promotion of open data policies and the continued development of robust archival systems are therefore essential components of maximising the long-term scientific return of ESA missions.

PoS(FRAPWS2024)004

6. Conclusions

This study has examined nearly fifty years of scientific output generated by missions within the ESA Science Programme. From fewer than ten publications annually in the late 1970s — when COS–B was the sole operational mission — the number of papers has grown steadily, exceeding 4,000 per year by 2021. This increase mirrors the expansion of the ESA Science Programme itself, which now comprises both active and legacy missions that continue to support new scientific discoveries.

While the overall growth in scientific publishing across astronomy, planetary science, and heliophysics has been substantial in recent decades, the proportion of publications based specifically on ESA mission data has grown at an even faster pace. ESA-led missions accounted for just over 4% of the field's publications in 2000; by 2021, that share had nearly tripled to over 11%, or over 15% when including ESA partner-led missions. This accelerating contribution underscores the increasing scientific relevance and global impact of ESA's missions.

Importantly, the data show that scientific productivity does not end with mission operations. Across all fields of space science, the majority of ESA-related publications are now archival, i.e. produced using data retrieved from mission archives without the involvement of the original investigators. These archival papers are generally cited just as frequently as those written by the original data producers, demonstrating their continued impact. This underscores the essential role of the ESA Space Science Archives in ensuring long-term access to mission data and enabling sustained scientific discovery.

The geographical distribution of publications largely reflects the size and financial contributions of ESA Member States, which are proportional to national GDP. However, several countries, both large and small, consistently exceed expectations in terms of scientific output, illustrating the effectiveness of their research communities.

In summary, ESA's Science Programme has not only grown in scale and scope over the decades, but has also increased its scientific influence, enabling a broad and sustained contribution to global space science. The findings reinforce the importance of long-term mission planning, open data policies, and strong archival infrastructure to maximise the scientific return on investment in space science and exploration.

References

[1] G. De Marchi, A. Parmar, ESA Mission Publications and Their Impact, in ESA Science Programme Missions: Contributions and Exploitation, ISSI Scienific Report series 18, ed. by A. Parmar, R. Bonnet, G. De Marchi, et al. Springer, 9-69 (2024)

[2] A. Parmar, R. Bonnet, G. De Marchi, et al., ESA Science Programme Missions: Contributions and Exploitation, ISSI Scienific Report series 18, Springer (2024)

[3] R. Reinhard, The Giotto mission to Halley's comet. Adv. Space Res. 2(12), 97–107 (1982)

[4] V. Domingo, B. Fleck, A.I. Poland, The SOHO Mission: An Overview. Solar Phys. 162(1), 1–37 (1995)

[5] R. Schulz, C. Alexander, H. Boehnhardt, K. Glassmeier, Rosetta: ESA's Mission to the Origin of the Solar System. Springer (2009)

[6] J.E. Hirsch, An index to quantify an individual's scientific research output. Proc. Natl. Acad. Sci. 102(46), 16569–16572 (2005)

[7] A. Accomazzi, M.J. Kurtz, E.A. Henneken, et al., ADS: the next generation search platform, in Open Science at the Frontiers of Librarianship, ed. by A. Holl, S. Lesteven, D. Dietrich, et al. Astronomical Society of the Pacific Conference Series, vol. 492 (2015), p. 189

[8] L. Scarsi, K. Bennett, G.F. Bignami, et al., The Cos-B experiment and mission, in Recent Advances in Gamma-Ray Astronomy, ed. by R.D. Wills, B. Battrick, vol. 124. ESA Special Publication (European Space Agency, Paris, 1977), p. 3

[9] P. Benvenuti, The IUE mission. Mem.. Ital. Astron. Soc. 54, 359 (1983)

[10] G.L. Pilbratt, J.R. Riedinger, T. Passvogel, et al., Herschel space observatory. An ESA facility for far-infrared and submillimetre astronomy. Astron. Astrophys. 518, L1 (2010)

[11] J.A. Tauber, N. Mandolesi, J.L. Puget, et al., Planck pre-launch status: the Planck mission. Astron. Astrophys. 520, A1 (2010)

[12] M.A.C. Perryman, Hipparcos: astrometry from space. Nature 340(6229), 111–116 (1989)

[13] M.F. Kessler, J.A. Steinz, M.E. Anderegg, et al., The infrared space observatory (ISO) mission. Astron. Astrophys. 315(2), L27 (1996)

[14] F. Jansen, D. Lumb, B. Altieri, et al., XMM–Newton observatory. I. The spacecraft and operations. Astron. Astrophys. 365, L1–L6 (2001)

[15] C. Winkler, T.J.L. Courvoisier, G. Di Cocco, et al., The INTEGRAL mission. Astron. Astrophys. 411, L1–L6 (2003)

[16] M.A.C. Perryman, The GAIA mission, in GAIA Spectroscopy: Science and Technology, ed. by U. Munari. Astronomical Society of the Pacific Conference Series, vol. 298 (Society of the Pacific Astronomical, San Fransisco, 2003), p. 3

[17] J.P. Lebreton, D.L. Matson, The huygens probe: science, payload and mission overview. Space Sci. Rev. 104(1), 59–100 (2002)

[18] J. Zarnecki, A. Parmar, Payload Provistion to the ESA Science Programme, in ESA Science Programme Missions: Contributions and Exploitation, ISSI Scienific Report series 18, ed. by A. Parmar, R. Bonnet, G. De Marchi, et al. Springer, 71-86 (2024)

[19] J. Peek, V. Desai, R.L. White, et al., Robust archives maximize scientific accessibility (2019). https://arxiv.org/abs/1907.06234

[20] C.P. Escoubet, M. Fehringer, M. Goldstein, Introduction: the cluster mission. Ann. Geophys. 19, 1197–1200 (2001)

PoS(FRAPWS2024)004

DISCUSSION

GENNADY BISNOVATYI–KOGAN: The budget is very high for big missions. Did you analyse the connection between the budget of a mission and the quality and quantity of scientific results obtained?

GUIDO DE MARCHI: Undoubtedly, large missions are more expensive to build and to operate, but they serve a wider community and result in a larger amount of data. This combination typically leads to a larger number of publications and results in a bigger impact. But this does not mean that smaller missions are less interesting or relevant. ESA missions are selected by the Member States through a bottom-up approach and are the result of years of collaboration within and across countries and within and across research areas. The results of the missions are very relevant to their respective communities, and this is shown clearly by the steady growth of the h-index for all of them (see Figure 2).

CLAUDIA MARASTON: Did you include the journal impact factor in your evaluation? And the number of active scientists over the number of citizens per country?

GUIDO DE MARCHI: We have considered the impact factor of the journals, but since the ESA Science Programme covers a broad field of research, it is not surprising that the publications are spread over a wide range of journals. Nevertheless, there does not appear to be a specific correlation between specific journals and countries, for instance. Your second point is excellent and we have not addressed it yet, but will do so in the future. However, [18] have already performed a study about the gender distribution of the authors (see Chapter 3 in [2]) and see a steady evolution since the 1970s with a growing female representation in the number of PIs of science instruments and experiments in the ESA Science Programme.

EZEQUIEL MARCHESINI: What is in your opinion the reason behind Gaia's success?

GUIDO DE MARCHI: Gaia builds on the strength and success of another astrometry mission, Hipparcos. The success of the Gaia mission can be attributed to a combination of unique scientific value, strong knowledge base in the community, advanced technological design, broad applicability of its data, and clearly an open data policy. When it comes to the number of publications, the immediate public availability of Gaia's data is the major factor of its success, in my opinion. Not only can researchers start working with the data as soon as they are released, but since no one "owns" the data it is easier for scientists to work on areas that they might normally avoid due to potential competition with the original PIs. I am convinced that making Gaia's data immediately public was not just a boost, but a transformative decision that turned Gaia from a successful mission into a global scientific engine.

PoS(FRAPWS2024)004