https://doi.org/10.5194/egusphere-2022-810
Preprint. Discussion started: 4 October 2022

Quantifying gender gaps in seismology authorship

Laura Ermert[1,*], Maria Koroni[1,*], and Naiara Korta Martiartu[2,*]

[1]Swiss Seismological Service, ETH Zürich, Zürich, Switzerland
[2]Institute for Applied Physics, University of Bern, Bern, Switzerland
[*]These authors contributed equally to this work and are ordered alphabetically.

Correspondence: Laura Ermert (laura.ermert@sed.ethz.ch), Maria Koroni (maria.koroni@sed.ethz.ch), and Naiara Korta Martiartu (naiara.korta@unibe.ch)

Abstract. According to 2018 demographic data of the American Geophysical Union Fall Meeting, seismology is among the Geoscience fields with the lowest female representation. To understand whether this reflects seismology more generally, we investigate female authorship of peer-reviewed publications, a key factor in career advancement. Building upon open-source tools for web-scraping, we create a database of bibliographic information for seismological articles published in 14 international journals from 2010 to 2020. We use the probabilities of author names being either male or female-gendered to analyse the representation of female authors in terms of author position and subsequently per journal, year, and publication productivity. The results indicate that: 1) The overall probability of the first (last) author being female is 0.28 (0.19); 2) With the calculated rate of increase from 2010 to 2020, equal probabilities of female and male authorship would be reached towards the end of the century; 3) Compared to the overall probability of male authorship (0.76), single-authored papers in our database are disproportionately published by male authors (with probability 0.83); 4) Female representation decreases among highly productive authors; 5) Rather than being random, the composition of authorship appears to be influenced by gender: Firstly, all-male author teams are more common than what would be expected if teams were composed randomly. Secondly, the probability that first or co-authors are female increases when the last author is female, but first female authors have a low probability of working with female co-authors.

1 Introduction

In seismology, as in many fields of research, peer-reviewed articles are one of the most important ways to disseminate new scientific findings. They are also increasingly used as a metric of performance and productivity of individual researchers, and constitute a critical factor of career advancement, along with citation scores and the impact factor of journals where researchers publish (West et al., 2013). Gender inequality, especially in higher-level academic positions, is a persistent problem in the fields of science, technology, engineering, and mathematics (STEM). The attrition of female graduates from STEM fields has been described as a leaky pipeline, with females dropping out at higher rates than males at various career stages (Resmini, 2016), an effect which was also documented in the geosciences (Ranganathan et al., 2021; Pico et al., 2020, and references therein). There is no consensus as to why this problem exists and persists; studies have investigated the effects of social and cultural norms, implicit bias (Dutt et al., 2016), role models and science pedagogy, hostile workplace climates for female researchers

https://doi.org/10.5194/egusphere-2022-810
Preprint. Discussion started: 4 October 2022

(e.g. Marín-Spiotta et al., 2020; Casad et al., 2021), choosing to take on larger burdens of care work outside and service work inside academia (Agnini et al., 2020; Ceci and Williams, 2011; Canetto et al., 2012), and even the conception of science itself as a male endeavour due to its historic development (Keller, 2003). A review of various of these arguments can be found in Blickenstaff (2005).

Lerchenmueller and Sorenson (2018) point out that the loss of female researchers in the life sciences does not occur as a steady drip but rather as a heavy spill at critical career junctures, such as the postdoctoral to junior faculty transition. Publication productivity is an important predictor of success during these transitions. Under-representation of female authors with respect to the presence of female researchers in a research field, as observed by Pico et al. (2020), may consequently be one cause of the continued under-representation of female researchers in that field. According to demographic data from the American Geophysical Union (2018)[1], seismology is among the geoscience disciplines with the lowest female representation. In the present study, we therefore analyse bibliometric data from 14 peer-reviewed journals that are commonly chosen for publishing seismological research in a period of eleven years (2010 – 2020). We build upon the open-source toolkit that Pico et al. (2020) developed to analyse gender in geoscience authorship. This allows us to automatically scrape bibliometric information from journal websites, extract author names, and then obtain their likely gender from requests to web databases relating names to gender. We furthermore propose a method to account for the uncertainty in the author name–gender association by not setting a fixed threshold. For example, Züleyha is considered a female name with 98 % probability, while Hongbo is considered male with 91 % probability, and Andrea is commonly used for both genders (62 % male). By analysing various aspects of the authorship statistics with regard to gender, we aim to document the problem of female under-representation in seismology publications, its recent and possible future development, and point to several consequences that the status quo has both for female seismologists and the field itself. We consider both under-representation with respect to the general population (assuming a 1:1 gender ratio) and under-representation of female authors with respect to the presence of female researchers in the field. To the best of our knowledge, such a detailed study of the authorship gender demographics in seismology has not been undertaken to date, leaving a knowledge gap that needs to be closed in order to support the effort of diversifying all fields of Earth sciences. Geoscience graduation rates in the US have been near gender parity for decades (Holmes et al., 2008), so any under-representation of female seismologists would indicate that structural barriers exist. Moreover, research on diversity suggests that diverse teams are more innovative (Nielsen et al., 2017; Woolley et al., 2010) and more successful at publishing (Nelson, 2014; Lerback et al., 2020), providing additional motivation to work towards diversity.

In our analysis, we specifically focus on the following aspects: i) The overall representation of female authors in seismology; ii) The composition of female/male authors in publication teams; this was examined in terms of author position (first, intermediate, last position in the authors list); iii) The change of female author representation during the past decade; iv) Female author representation per journal; and v) Publication productivity according to the frequency of occurrence of an author in the articles database. Below we briefly introduce our methods and results with regard to these questions, followed by a discussion and conclusions which include our perspective on female authorship in seismology.

[1] Retrieved from: https://honors.agu.org/files/2018/09/2018-section-membership-by-gender-and-career-stage_Sept12.pdf, last accessed 18.08.2022

https://doi.org/10.5194/egusphere-2022-810
Preprint. Discussion started: 4 October 2022

2 Method

2.1 Collecting bibliometric data

We analyse the representation of female authors in peer-reviewed research articles published in seismology from January 2010 to December 2020. We consider 14 international journals representing a relevant, broad spectrum of sub-disciplines within the field and a range of impact factors (see Table 1). We collected bibliometric data from the online search masks of the journals, modifying the web-scraping Python code developed by Pico et al. (2020) available online for this purpose. This tool uses the Python package Selenium for opening and downloading search results and the Python package BeautifulSoup to parse the resulting HTML files (Muthukadan, 2022 [last update]; Richardson, 2022 [latest release]). We targeted articles that broadly fall into the field of seismology by selecting all articles with the keyword fragment 'seism' (matching for example seismic, seismological, seismicity) and the keyword 'earthquake' in the abstract. In this way, we obtained entries for 20'108 articles. We extracted the full names of all authors in each article, yielding a list of 88'331 authors. In approximately 20 % of cases, authors chose to use initials rather than first names, and we followed the strategy of Pico et al. (2020) to cross-reference initials and last names with full names in the list of all authors. In addition, we obtained bibliometric information from the SAO/NASA Astrophysics Data System digital library portal (https://ui.adsabs.harvard.edu/) of conference abstracts and used author information of 17'452 abstracts presented at the European Geosciences Union General Assembly during the period 2010 – 2020 to cross-reference initials and last names with full names. In this way, we could identify full names from initials in the majority of cases ($>$ 80 %). Publications with unidentified initialed names were omitted from the database.

2.2 Relating first names to gender

We use several web databases to infer author gender from the name. We first submitted all first names in the database to the genderize.io API (https://genderize.io/) used by Pico et al. (2020). For each name, we stored the likely gender of the name ("female", "male" or None) and the probability of the gender returned by the API. Using genderize.io, we identified 73 % likely male, 20 % likely female, and 7 % not classifiable names. For names that could not be classified (None result), we repeated the process with the NamSor API (https://github.com/namsor/namsor-python-sdk2), which requires first and last name as input, and uses public "labelled" data such as voter registration lists, but also linguistic cues such as name endings. This resulted in a combined identification (from both genderize.io and NamSor) of 76 % likely male, 20 % likely female, and less than 4 % unclassified author names. After removing articles with any unclassified author names, our final dataset contains 18'150 articles (Table 1).

2.3 Representing gender through probabilities

The online databases to determine the gender of author names return a probability that the name in question is male or female. As is common practice, Pico et al. (2020) set a threshold at 0.5 to distinguish male and female names. In contrast, we retain

https://doi.org/10.5194/egusphere-2022-810
Preprint. Discussion started: 4 October 2022

Journal	Impact Factor	Number of Articles
Nature	42.779	59
Science	41.845	78
Nature Geoscience	14.480	169
EPSL	4.823	1239
GRL	4.50	2022
JGR: Solid Earth	3.64	3027
JGR: G3	3.28	736
SRL	3.131	1452
Tectonophysics	3.048	1606
Solid Earth	2.921	219
GEOPHYSICS	2.793	1753
GJI	2.574	3308
BSSA	2.274	2024
PEPI	2.237	458

Table 1. Number of articles per journal analysed in this study. We also indicate the 2-year impact factor reported by each journal in 2021. EPSL: Earth and Planetary Science Letters; GRL: Geophysical Research Letters; JGR: Journal of Geophysical Research; G3: Geochemistry, Geophysics, Geosystems; SRL: Seismological Research Letters; GJI: Geophysical Journal International; BSSA: Bulletin of the Seismological Society of America; PEPI: Physics of the Earth and Planetary Interiors.

the probability returned by the online tools and base our analysis on it. Using a fixed threshold can distort the results because not all first names are unambiguously gendered. As an example, consider the name Ashley, which is classified as female by genderize.io, but with a probability of only about 0.6. If our dataset contained 10 authors named Ashley, and we used a cut-off to assign them a binary gender, 10 out of 10 would be considered female. By continuing to work with the probability instead, and interpreting probability in terms of frequency, 6 out of 10 would be considered female, and 4 male, which provides a more accurate picture of the demographics.

2.3.1 Computing probabilities

Here, we summarize the most relevant mathematical operations used for computing the probabilities in Sect. 3. We base our analysis on a binary notion of name gender so that the probability of a name to be either male or female is 1. Our database contains a total of n articles, and we denote an article as x_i with $i \in \{1, ..., n\}$. From the output of the gender determination tools, we obtain the conditional probability $p(F_k \mid x_i)$ of having a female-gendered author name at the authorship position k in the article x_i. For example, $p(F_1 \mid x_i)$ refers to the probability of having a female-gendered first-author name in the article x_i.

https://doi.org/10.5194/egusphere-2022-810
Preprint. Discussion started: 4 October 2022

Then, the overall probability of having a female-gendered first-author name in our database can be computed as

$$p(F_1) = \sum_{i=1}^{n} p(F_1 \mid x_i) p(x_i). \tag{1}$$

We consider all articles equally likely; thus, $p(x_i) = 1/n$ for all i, so that Eq. (1) reduces to the arithmetic mean of $p(F_1 \mid x_i)$. Similarly, the overall probability of having a male-gendered first-author name is given by

$$p(M_1) = \sum_{i=1}^{n} (1 - p(F_1 \mid x_i)) p(x_i) = 1 - \frac{1}{n} \sum_{i=1}^{n} p(F_1 \mid x_i) = 1 - p(F_1). \tag{2}$$

We use Eqs. (1) and (2) indistinctly for any authorship position by replacing $p(F_1 \mid x_i)$ with the appropriate probability. For last-authors this becomes $p(F_{\text{last}} \mid x_i)$, and for co-authors we use

$$p(F_{\text{co}} \mid x_i) = \frac{1}{m-2} \sum_{k=2}^{m-1} p(F_k \mid x_i), \tag{3}$$

where m is the total number of authors of the article, and we assume uniform distribution for co-authorship positions. For the sake of this analysis, we define co-author as any author that is neither first nor last.

To analyse the gender composition of author teams, we start by computing the probability of an author list with only same-gender author names. For instance, the probability of all author names being female in the article x_i is computed as

$$p_i(F_{\text{all}}) = \prod_{k=1}^{m} p_i(F_k), \tag{4}$$

where $p_i(\cdot) \coloneqq p(\cdot \mid x_i)$, and we assume that the genders of the authors in an article are independent. We follow the same procedure to compute the probability of all author names being male $p_i(M_{\text{all}})$. Finally, the probability of having a mixed-gender author list can be derived as

$$p_i(\text{mix}) = 1 - p_i(F_{\text{all}}) - p_i(M_{\text{all}}). \tag{5}$$

We could also estimate the probability of having at least one female-gendered author name in the author list by computing $p_i(F_{\text{at least one}}) = 1 - p_i(M_{\text{all}})$. Similar to Eqs. (1) and (2), we use the arithmetic mean to estimate the overall probabilities of $p_i(\text{mix})$, $p_i(F_{\text{all}})$, and $p_i(M_{\text{all}})$.

To investigate gender dynamics in the composition of author teams, we derive conditional probabilities of the first-author gender given the last-author gender and vice versa. For instance, we compute the probability of having a female-gendered first-author name given that the last author name is also female-gendered as

$$p(F_1 \mid F_{\text{last}}) = \frac{\sum_{i=1}^{n} p_i(F_1) p_i(F_{\text{last}})}{\sum_{i=1}^{n} p_i(F_{\text{last}})}. \tag{6}$$

Similar equations can be defined for different combinations of first- and last-author genders or for different probabilities of interest (e.g., probability of having at least one female co-author given that the last author is female).

https://doi.org/10.5194/egusphere-2022-810
Preprint. Discussion started: 4 October 2022

2.4 Statistical analysis

Statistical analysis and visualization were performed using Python (version 3.8.10) with the SciPy (1.7.1), Pandas (1.3.2), and Seaborn (0.11.2) libraries. We used Pearson's r coefficient to calculate correlations of probabilities with publication years and Spearman's ρ coefficient for correlations with journal impact factors and the number of authors. Results were considered significant for p-values lower than 0.05. We used the slope of linear regressions to analyse the increase rate per year of the probabilities and forecast when the parity is reached. Additionally, we provide the average annual growth rates (AAGR) of the probabilities for direct comparisons with similar studies.

3 Results

3.1 Overall representation of female authors

The overall representation of female-gendered author names in our database is 23.6 %, approximately one-fourth of the total number of authors of the publications analysed, and more than the frequency of names classified as female when using a threshold of 0.5 for determining the gender (20.6 %). As shown in Fig. 1a, they appear most likely in the first authorship position (with a probability of 0.28), followed by the co-authorship (0.23) and the last authorship position (0.19). Seven (eight) out of ten articles therefore have a male-gendered first-author (last-author) name.

3.2 Gender composition of author teams

The percentage of articles with all female-gendered author names is only 2.7 % (Fig. 1b). That is, male-gendered author names appear in 97.3 % of publications. In contrast, 41.3 % of articles do not contain any female-gendered author name. Male authors are therefore 15 times more likely to work in same-gender teams than the female authors. Publications with mixed-gender author lists are most common in seismology, but not yet the norm (56.0 %).

If authors in a team are chosen at random, larger author teams are expected to be more diverse in terms of gender. We indeed find that the probability of having at least one female-gendered author name is significantly and positively correlated with the number of authors of a publication ($\rho = 0.52, p < 0.0001$). On average, it increases from 0.17 for single-authored articles to 0.93 for those with twelve authors (Fig. 2a). However, when comparing these observations with the expected values computed from the overall female author representation (0.24) assuming randomly composed teams, we find a negative bias regardless of the team size (Fig. 2b). Single-authored publications show the strongest under-representation of female-gendered author names, with a negative bias of 6.5 %. Interestingly, the expected probabilities also reveal that a female-gendered author name has a 95 % probability of appearing in a publication when the number of authors is twelve, whereas for male-gendered author names the number of authors required to reach the same probability is three. The median number of authors in our database is four, corresponding to a 0.66 expected probability of at least one female-gendered author name appearing in any position, while the observed value for this probability is 0.59 (Fig. 1b).

https://doi.org/10.5194/egusphere-2022-810
Preprint. Discussion started: 4 October 2022

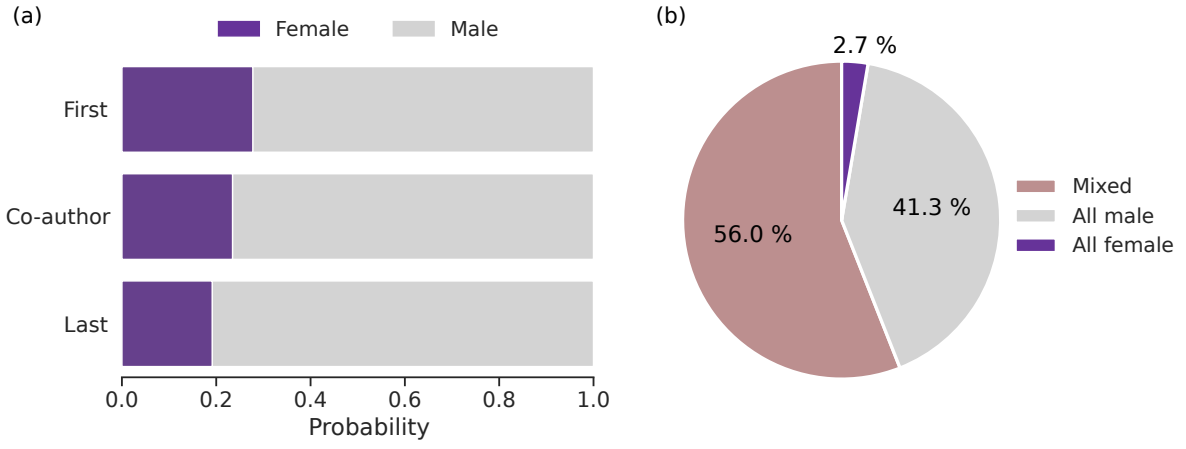

Figure 1. Gender distribution in the authorship list of peer-reviewed publications in seismology. (a) The probability of having a female- or male-gendered first-author, co-author, and last-author name in a publication. (b) Percentage of publications with an authorship list with all-female, all-male, or mixed-gender author names.

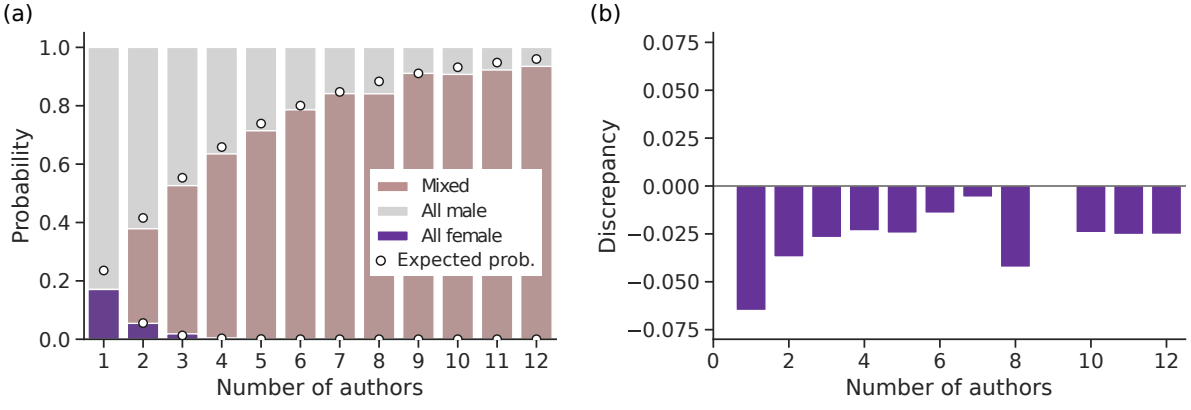

Figure 2. Probabilities with respect to the number of authors of a publication. (a) Expected and observed probability of an all-female, all-male, or mixed-gender author list. (b) Discrepancy between the observed probability of having at least one female-gendered author name and the expected probability for randomly composed teams.

3.3 Composition of author teams conditioned on first- and last-author gender

As Fig. 1b shows, working in mixed-gender author teams is the norm for female authors, while there is a substantial probability for male authors to work in all-male author teams. To investigate further whether the author team composition can be considered random with respect to gender, we evaluate the probabilities of the first- and co-author gender conditioned on the last-author gender and probabilities of the last- and co-author gender conditioned on the first-author gender. The results in Fig. 3a suggest

https://doi.org/10.5194/egusphere-2022-810
Preprint. Discussion started: 4 October 2022

that female-gendered first (last) author names are 4.4 % (3.4 %) more likely when the last (first) author name is also female-gendered, displaying a slight "gender unmixing" effect in teams consistent with the bias observed in the previous section. We observe a similar effect in the probability of having at least a female-gendered co-author name, which is 6.4 % larger for female last authors than male last authors. Interestingly, we find that female first authors are 16.8 % less likely than male first authors to work with female co-authors.

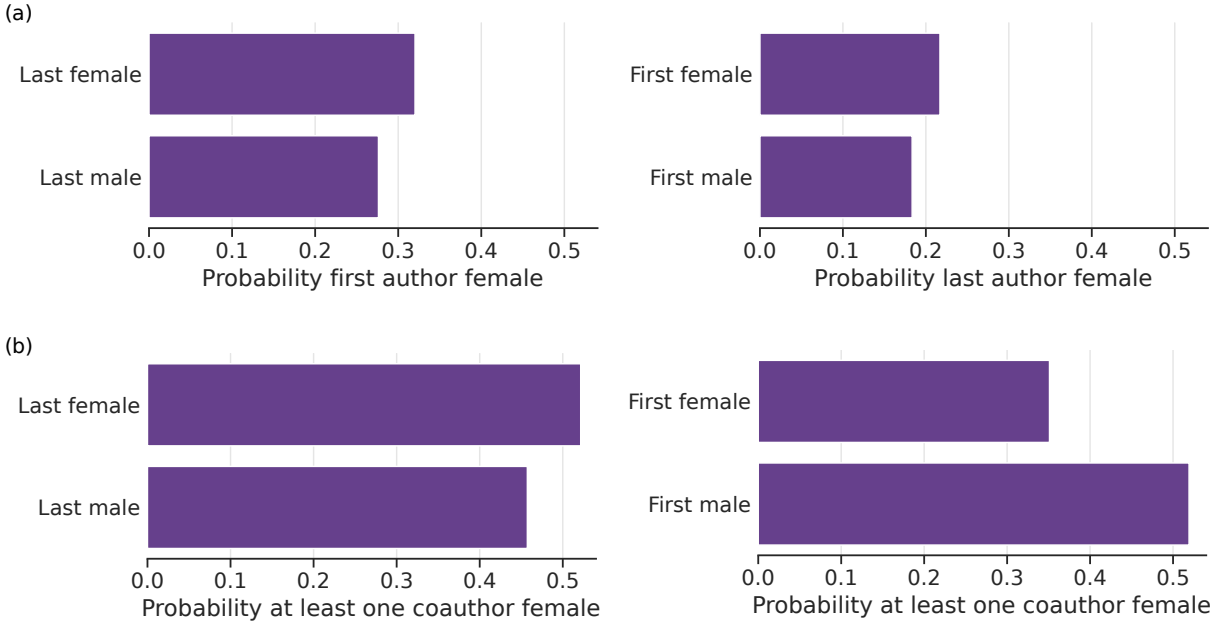

Figure 3. Conditional probabilities analysing author team composition. (a) Probability of having female-gendered first-author names (left) and last-author names (right), given the gender of the last and first author, respectively. (b) Probability of having at least one female-gendered co-author name given the gender of the last (left) and first (right) author.

3.4 Changes in the last decade

As shown in Fig. 4a, probabilities of having female-gendered names in first-, co-, and last-authorship positions have increased over the last decade (first author: $r = 0.81, p = 0.002$; co-author: $r = 0.95, p = 1 \cdot 10^{-5}$; last-author: $r = 0.85, p = 0.001$). The increase rate per year is fastest for co-authors (0.6 ± 0.1 % ; AAGR: 2.4 %), twice that of the first and last authors (0.3 ± 0.1 %; AAGR: 1.5 % and 1.7 %, respectively). Assuming these annual rates are constant in the following years, a probability of 0.5 representing parity with respect to the general population will be reached in 42 ± 5, 72 ± 17, and 93 ± 20 years, respectively, in the co-, first, and last author position.

The composition of author teams is rapidly becoming more mixed in gender with time ($r = 0.97, p = 5 \cdot 10^{-7}$). In 2020, publications with mixed-gender authors were approximately 13 % more likely than in 2010 (Fig. 4b). Consequently, there has been a substantial drop in all-male-authored publications, from 48.2 % in 2010 to 36.0 % in 2020. The mean number of authors

https://doi.org/10.5194/egusphere-2022-810
Preprint. Discussion started: 4 October 2022

of a publication has also increased by almost one author in the last decade (Supplementary Fig. 1), a trend that is generally observed in scientific publishing (e.g. Kuld and O'Hagan, 2018, in economics). This may have contributed to diversifying author teams, as already discussed in Fig. 2. With the observed annual rate, approximately all publications will be authored by
180 mixed-gender teams in 26 ± 1 years.

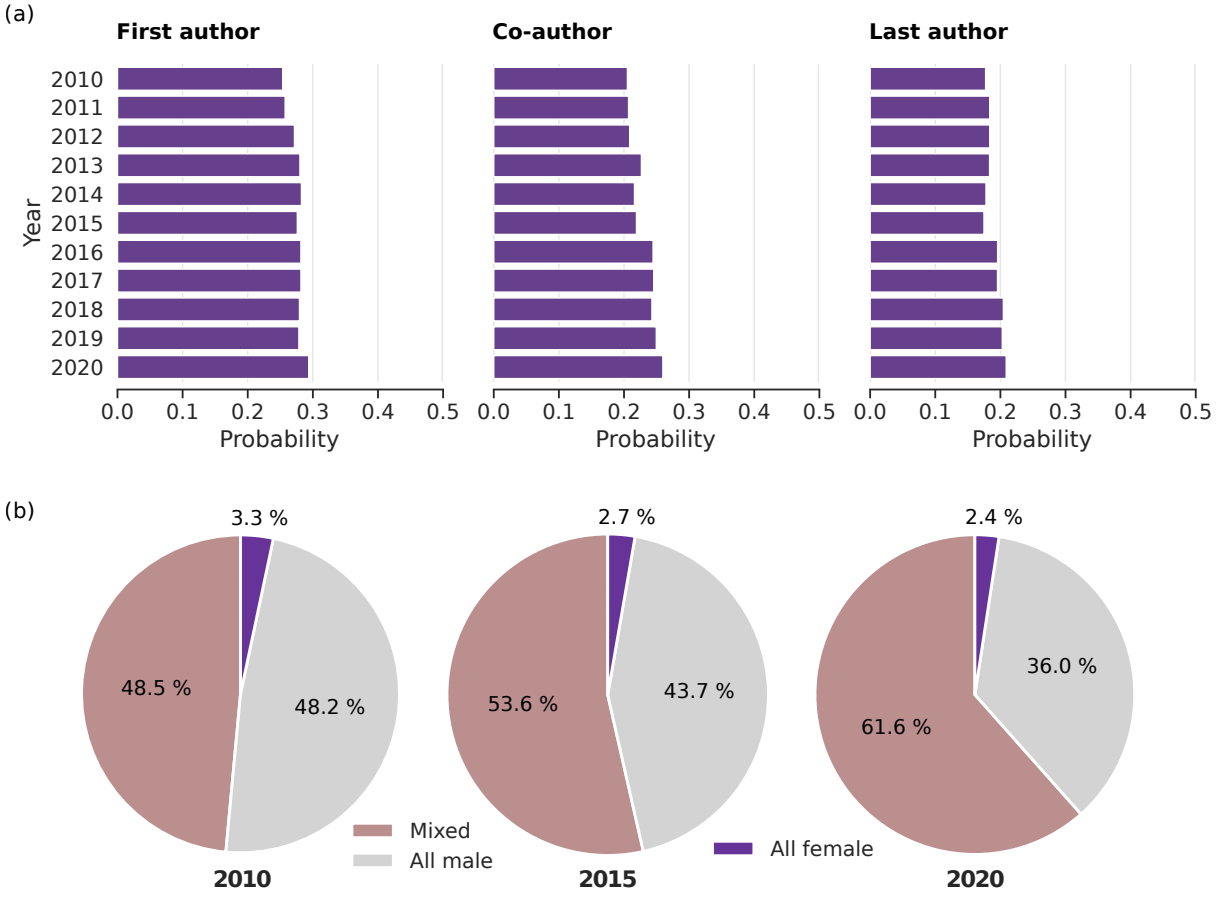

Figure 4. Changes in female representation and gender composition of author teams during the last eleven years. (a) Probabilities of first-author (left), co-author (center), and last-author (right) names being female-gendered per year in articles published from 2010 to 2020. (b) Percentage of publications with an authorship list with all-female, all-male, and mixed-gender author names in 2010 (left), 2015 (center), and 2020 (right).

3.5 Representation by journal

Figure 5 shows the probabilities of having female-gendered names in first-, co-, and last-authorship positions for each journal. We group the journals Nature, Science, and Nature Geoscience (Nat/Sci) to increase robustness, as our dataset contains a relatively small number of publications in these journals ($n = 306$). In general, by-journal probabilities are close to the

https://doi.org/10.5194/egusphere-2022-810
Preprint. Discussion started: 4 October 2022

overall probabilities shown in Fig. 1a. However, three bins stand out: (i) The journal GEOPHYSICS has the lowest female representation in all authorship positions. Compared to the overall probabilities, the underrepresentation is most substantial for first authors (negative bias of 7 %), followed by last authors (5 %) and co-authors (3 %). This journal also shows the smallest mean number of authors per publication (3.3 ± 1.5, Supplementary Fig. 2). (ii) The journal Geophysics, Geochemistry, Geosystems (*G3*) has an elevated probability of female-gendered first author names (positive bias of 6 %). (iii) The journals Nat/Sci, which have the highest impact factors, have a lower than overall female representation among first authors (negative bias of 5 %) and last authors (3 %). Female-gendered author names are therefore more likely to appear in the co-authorship position, unlike the general trend observed in Fig. 1a. These journals show the largest mean number of authors per publication (6.6 ± 7.4, Supplementary Fig. 2). Despite high-impact journals having a significantly larger number of authors per publication ($\rho = 0.71, p = 0.004$), we do not observe any correlation between the probability of having at least one female-gendered author and the journal impact factor ($p = 0.3$).

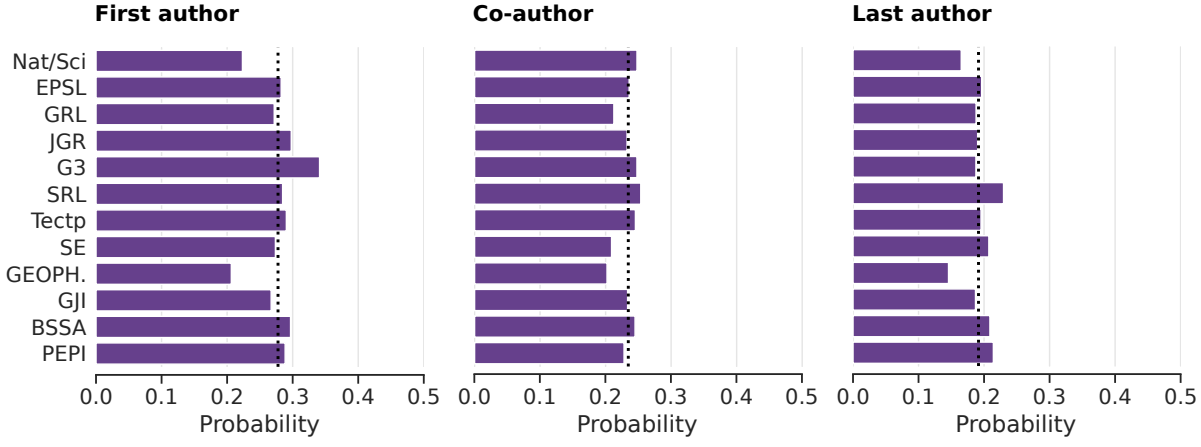

Figure 5. Probabilities of first-author (left), co-author (center), and last-author (right) names being female-gendered per journal. Dashed lines indicate the overall probabilities from Fig. 1a in each case. Nat/Sci: Nature, Science and Nature Geoscience; EPSL: Earth and Planetary Science Letters; GRL: Geophysical Research Letters; JGR: Journal of Geophysical Research: Solid Earth; G3: Geochemistry, Geophysics, Geosystems; SRL: Seismological Research Letters; Tectp: Tectonophysics; SE: Solid Earth; GEOPH.: GEOPHYSICS; GJI: Geophysical Journal International; BSSA: Bulletin of the Seismological Society of America; PEPI: Physics of the Earth and Planetary Interiors.

3.6 Publication productivity

We compare authorship probabilities to seismologists' participation at the European Geosciences Union (EGU) General Assembly between 2016 and 2020 in Fig. 6 (Data obtained by request to EGU). Since 2016, EGU systematically collects self-declared demographic information from participants upon registration. In 2016–2017, response rates to the question about gender were low, around 50 % for overall seismology section attendees and around 40 % for early career scientists (ECS). The response rates increased in 2018–2019 to around 60 % (ECS: around 50 %) before reaching close to 100 % in 2020. In 2021

https://doi.org/10.5194/egusphere-2022-810
Preprint. Discussion started: 4 October 2022

(not included in our dataset), EGU reports higher female participation than in 2020, with 33 % overall and 38 % ECS. We thus note that for most years, particularly those with higher response rates, and for both levels of seniority, conference participants have a larger probability of being female than manuscript authors. This could point to a gender gap in authorship.

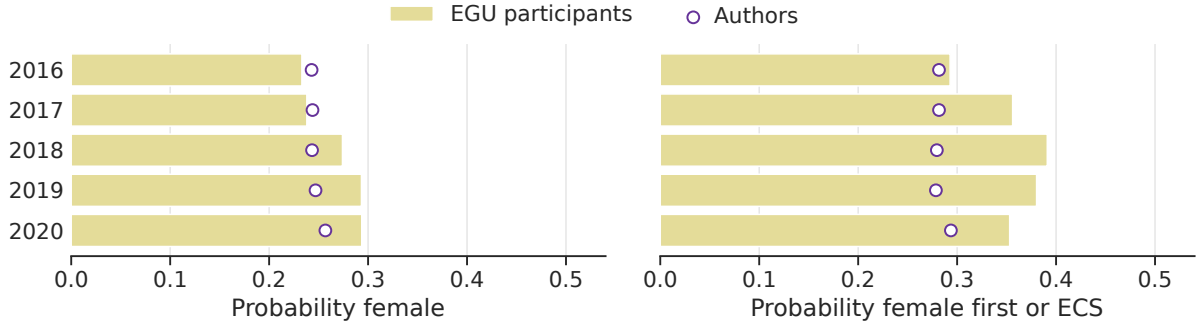

Figure 6. Comparison of probability of female authorship in our article database to demographic data of conference participation at the European Geosciences Union General Assembly. Left panel: All authors. Right panel: First authors of articles and early career scientist participants at EGU.

205 However, it is difficult to compare the populations of conference participants (with possible geographic preference) to article authors. Therefore, we take a second approach to analyse the effect of gender on productivity using only our data set. We identify unique full names of first and last authors in the database and rank them by the number of their publications. Fig. 7 (purple curve) shows how the probability of female-gendered first and last author names evolves with the ranking by productivity. Considering all 9305 unique first authors in the first ten years of the database, the probability of unique female-
210 gendered author names among them is close to 0.3. However, when we consider only the most productive 10 % of authors, this drops to 0.25. Among the most productive few percent, probability appears to drop further; however, the absolute number of female-gendered first-author names is too low in this range to draw statistically meaningful conclusions. Last author names ranked by productivity show the same pattern. Hypothesizing that this lack of highly productive female authors is an effect of women entering the field later, we separate the data for the years 2010-2014 and 2015-2019 (blue and pink curves in Fig. 7,
215 respectively). We can make several observations based on these results: (i) The number of unique authors has increased by 21 % in the span of five years. (ii) Authors with female-gendered names are overall more strongly represented in 2015-2019. (iii) This increase in overall representation includes some of the most productive authors ranked between 10 % and 20 %, but a gender gap among highly productive authors still persists.

4 Discussion

220 We present the first study analysing the representation of female researchers in peer-reviewed articles in the field of seismology. In geosciences, of which seismology is a sub-discipline, the gender distribution of first-author names has already been analysed by Pico et al. (2020). Here, we focus on a subfield and extend the scope to analyse female representation in first-, co-, and

https://doi.org/10.5194/egusphere-2022-810
Preprint. Discussion started: 4 October 2022

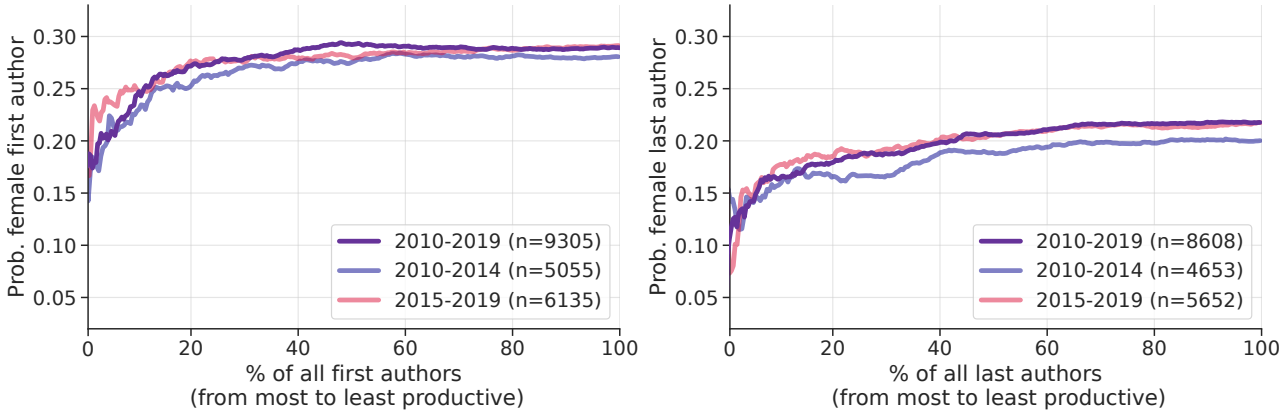

Figure 7. Female representation by ranking in terms of output. We show the probability of female first and last authors as a function of productivity quantile from most to least productive, where productivity is measured as the number of articles published. We do not interpret the most productive 10 % (small number of authors).

last-authorship positions, as well as gender dynamics in the author team composition and gender differences in publication productivity. In addition, we use a new approach to estimate the probability of author genders. Rather than considering a threshold to assign a binary gender to each name (either female or male), we use the probability of names being female-gendered. In this way, we automatically account for uncertainties in name-gender association, with no need to disregard names that are not gender-specific (e.g., Bendels et al., 2018; Pico et al., 2020).

In this study, we found that female researchers in seismology account for 24 % of authorships, precisely the same rate reported in earth and environmental sciences by Bendels et al. (2018) and 5-6 % below the female representation in the whole area of natural sciences (30 %) (Larivière et al., 2013; European Commission, Directorate-General for Research and Innovation, 2021). For every ten publications in seismology, only three are authored by female first authors and two by female last authors. Thus, female researchers are most likely to publish in seismology as main contributors and least likely as senior authors or project leaders, who tend to be named last. This pattern of authorship corresponds reasonably well to the findings of Bendels et al. (2018): The overall female author odds ratio, which indicates how female authorship is split between first, co- and last author positions, is 1.4, 1.0, and 0.7 for first, co-, and last authors in our dataset, while they are approximately 1.5, 0.9, and 0.7 in Bendels et al. (2018). This difference in representation among authorship positions evidences the progressive absence of women in academic positions with higher responsibility, commonly described as the leaky pipeline (e.g., Blickenstaff, 2005).

Similar to female researchers in other disciplines, female seismologists are particularly under-represented in single-authored publications (West et al., 2013; Walker, 2019). Sarsons (2017) found that, in economics, this type of under-representation penalizes female researchers' odds of receiving tenure while playing no significant role in the promotion of male researchers. In seismology, solo articles are only a small fraction of all articles (approx. 6 %), with an overall declining trend observed in multiple fields of research (Kuld and O'Hagan, 2018; West et al., 2013). However, this gap may still affect the career

https://doi.org/10.5194/egusphere-2022-810
Preprint. Discussion started: 4 October 2022

advancement of female researchers, as single-authored papers are considered proof of the authors' scientific skills (Sarsons, 2017; Kwiek and Roszka, 2022).

The vast majority of publications authored by female researchers (95 %) are composed of mixed-gender author teams, meaning that women routinely collaborate with colleagues diverse in gender. In contrast, male researchers co-publish exclusively with other male-gendered colleagues once every three articles. This is not only caused by seismology being a male-dominated field but goes beyond what we would expect from a random team composition (Fig. 2). Similar to what has been reported in other STEM fields (Ghiasi et al., 2016; Holman and Morandin, 2019), researchers in seismology appear to work with same-gendered co-authors more often than expected. However, this effect is not necessarily mutual: As will be further discussed below, it is possible that only one gender displays a same-gender preference. When the majority group displays this preference, it constrains the possibilities for the minority group to participate in academic networking and access novel ideas, information, and research opportunities (e.g., McPherson et al., 2001).

We furthermore observe that the probability of female first authors increases in publications with a female last author, and vice versa (Fig. 3a). Studies on unconscious bias suggest that this is not driven by the hiring preferences of female faculty, which tend to be the same as those of male faculty (Moss-Racusin et al., 2012). We hypothesize that the "unmixing" may in part be a geographic effect, i.e., that female researchers are not randomly distributed across universities, but tend to be better represented in more equitable institutes and countries, and collaborate more frequently within their institute and within their country (Kwiek and Roszka, 2021; Hanson et al., 2020). Another contributing factor may also be the implicit gender bias of male researchers. Araújo et al. (2017) found that female researchers are equitable in their collaborations, with their collaborator pools representing the gender balance of their fields, but male researchers show a preference for working with other males. Similarly, Hanson et al. (2020) demonstrated that while female AGU abstract authors have a higher rate of connections to other women compared to male abstract authors, the female-female rate of connections reflects the actual share of women among AGU members, whereas the male-female rate of connections remains below the share of AGU women members. If we assume that EGU demographic data (Fig. 6) accurately represents the distribution of female researchers in the field, our results indicate that female under-representation in important authorship positions is indeed stronger with male first or last authors.

A curious finding in this context is related to the conditional probability that at least one co-author is female given a last or first female author (Fig. 3b). It is comparatively high when the last author is female but comparatively low when the first author is female. It shows that: (1) Female senior researchers have a positive effect on overall female participation in seismology authorship, similar to the productivity increase shown by female graduate students working with female advisors (Pezzoni et al., 2016). (2) Early-career female seismologists tend to be more isolated from female co-authors who might be either additional senior researchers or peers at a similar career stage. Although by the time of this work, we did not find studies supporting or contradicting this observation, it agrees with our own experience in the field.

Female authorship has been slightly increasing from 2010 to 2020 for all authorship positions, most strongly in the co-author position. Both our observations and the self-reported demographic data of the EGU seismology section members indicate that female participation does not increase at a steady rate, but fluctuates year by year. We have used a linear model to fit the increase in the probability of female authorship with time. Conversely, it is common practice to assume an exponential model and report

https://doi.org/10.5194/egusphere-2022-810
Preprint. Discussion started: 4 October 2022

growth in female researchers' participation in terms of compound or average annual growth rates (Bendels et al., 2018; Pico et al., 2020; European Commission, Directorate-General for Research and Innovation, 2021). Obviously, the outlook changes drastically when an exponential vs. a linear model is used. Besides the linear increase rates of female author probabilities (0.3, 0.6, and 0.3 % / year for first, co-, and last authors), we find compound annual growth rates of 1.5, 2.4, and 1.7 % for the probability of female first, co-, and last authors, respectively. The rates we find for first and last authorship are roughly in agreement with the overall rate in natural sciences in the European Union determined by the European Commission (1.35 %) for the years 2015 – 2019, while the co-author rate is well above it (European Commission, Directorate-General for Research and Innovation, 2021). Our rates fit less well with Bendels et al. (2018), who report corresponding average annual growth rates of approximately 3, 1.5, and 2 % in Earth & Environmental sciences (Fig. 2 in Bendels et al., 2018), which is likely caused by a different choice of analysed journals.

As to the future development, there is little indication of exponential growth of female authorship in our 11-year dataset. The compound annual growth rates determined from the data are low enough for the exponential model to behave almost linearly (Supplementary Fig. 3). A model determined by linear regression actually provides a better fit to the data in all cases, especially for first authorship. Ultimately, our observation period is too short to answer this question, particularly in light of the strong year-to-year fluctuations. Longer-range data are needed to study the development in more detail, make better-founded predictions, and investigate which factors affect temporal fluctuations in diversity.

Female-gendered first names are less likely to appear in the first and last authorship position in high-impact seismology articles than in most other journals, with the exception of GEOPHYSICS. First authorship in high-impact journals can be a career-advancing achievement: Despite efforts to abolish the journal impact factor in hiring and promotion decisions (e.g. Lariviere and Sugimoto, 2019), using it as a performance metric has been commonplace and continues at many institutions (McKiernan et al., 2019). Thus, publishing in high-impact journals at lower rates may put female seismologists at lower odds of being hired and promoted.

The most highly productive authors are more likely to have male first names than authors of average productivity, both for first and last authors. This gap has been slightly narrowing over time, but still persists. A productivity gender gap has been documented in various areas of STEM (Larivière et al., 2013; Bravo-Hermsdorff et al., 2019; Lerchenmueller and Sorenson, 2018; Bendels et al., 2018; European Commission, Directorate-General for Research and Innovation, 2021, and references therein). A recent study suggests that female researchers are less likely to be offered authorship in collaborative projects (Ross et al., 2022), leading to an "attribution gap" that could explain parts of the productivity gap. Whatever its causes, it likely presents an obstacle to increasing the participation of women in seismology and reaching parity at the faculty level. Lerchenmueller and Sorenson (2018) found that the lower success rate of female life science researchers during the transition from postdoctoral to principal investigators (PI) is to 60 % explained by their lower overall productivity and that the remaining gap can be almost entirely accounted for by the effect that outstandingly productive authors are more highly rewarded (so-called Matthew effect).

The results of this study are limited to our small sample of selected journals. They have been chosen according to their impact factor and popularity based on our experience as seismologists in European institutions. Other journals may be more popular

https://doi.org/10.5194/egusphere-2022-810
Preprint. Discussion started: 4 October 2022

for different sub-communities in seismology. These may include regional journals or very specialized ones. Furthermore, we consider gender as binary even though we are aware of the complexity and multiplicity of gender identities. We make this simplification due to the current lack of data on other gender identities in geosciences. We determine genders mostly from authors' first names. This approach is reasonable for many names in western cultures, but can fail to generalize to other cultures where the names are not strongly gendered, or where gender is encoded in the full name. In this regard, we expect that considering gender probabilities rather than fixed thresholds improves our results compared to earlier studies. Finally, our study does not consider equally contributing authors and assumes that the seniority of a researcher is reflected by their position in the author list, with the last author being more senior than the first one. This is common in seismology, but may not be applicable in other scientific fields.

5 Conclusion

By analysing a large bibliographic database, we identified several gender gaps in seismology authorship. Firstly, the probability of female first authorship in recent years is below the probability of female Early Career Scientist conference attendance at the EGU general assembly. Secondly, the probability of having at least one female author team member is below what would be expected if teams were assembled randomly, which points to a gender bias in how authorship teams are composed. Thirdly, female-gendered first names are less represented among the most productive authors, the first and last authors in prestigious high-impact journals, and solo authors than they are overall in our dataset. Female author representation in seismology has improved from 2010 to 2020, but the improvement is fastest for the less prestigious co-author positions and is moving at a rather slow and stumbling pace. Given the observed rates,

- current early career researchers in seismology would experience that author teams become universally mixed-gender towards the end of their career;
- undergraduate students entering university in 2022 and going on to become seismologists would experience gender parity in seismology co-authorship;
- people born in the early 2020s who decide to become seismologists might experience parity in first authorship in seismology but would probably spend their entire careers still waiting for parity in last authorship.

A more optimistic scenario is possible if gender diversity increases exponentially as Pico et al. (2020) and Bendels et al. (2018) assume (instead of 42, 72 and 93, it would take approximately 30, 40 and 50 more years to reach gender parity in seismology co-, first and last authorship). Based on our findings, we can offer the following comments and recommendations to actively work towards gender diversity in seismology authorship:

- Our analysis profited from comparison to carefully collected and curated demographic data, e.g. by the European Geosciences Union. We encourage professional societies to collect or continue collecting demographic data where appropriate and where sufficient anonymity can be granted.

https://doi.org/10.5194/egusphere-2022-810
Preprint. Discussion started: 4 October 2022

- An effort in the form of continued studies should be made to understand why there is a high-productivity, solo-author, and high-impact gender gap in seismology publications.
- PIs could positively contribute by communicating and implementing transparent and realistic authorship criteria for their research groups. In addition, they should consider establishing criteria defining what output merits submission to a high-impact journal.
- PIs should pay particular attention to the opportunities for collaboration that they can offer to their female (and, by proxy, all minority) mentees. If the gender "unmixing" is not a geographic effect, it means that these persons have less choices for collaboration.
- For the case that the "unmixing" effect is geographic, researchers who cannot find female collaborators at their institute may support diversity by seeking them elsewhere and internationally.
- Those evaluating research performance should remain aware that there are, as of now, gender gaps in high-productivity, solo, and high-impact authorship in seismology.

Importantly, we have only considered the influence of gender on seismology authorship. Specific to context, researchers can hold various, sometimes intersecting minority identities who face serious obstacles in the geosciences (e.g. due to racism; Bernard and Cooperdock, 2018; Dutt, 2020; Dowey et al., 2021). We have focused on gender for two reasons: It is possible to infer it approximately from author names in many cases; and we are part of the minority in question. However, we sincerely hope that the seismological community will continue to uncover and remove obstacles for all minority researchers and work towards becoming more diverse and inclusive.

6 Code availability

All the tools we used in this research are openly available and can be obtained through github or the Python Package index as indicated by URLs and references.

7 Data availability

All data used in this study is publicly available. For the convenience of anyone who wishes to use it, the dataset of bibliographic information will be provided by the corresponding authors upon request.

8 Author contributions

All authors contributed equally in all steps involved in this study and manuscript writing.

https://doi.org/10.5194/egusphere-2022-810
Preprint. Discussion started: 4 October 2022

9 Competing interests

The authors declare that they have no conflict of interest.

https://doi.org/10.5194/egusphere-2022-810
Preprint. Discussion started: 4 October 2022

References

Agnini, C., Pamato, M., Salviulo, G., Barchi, K., and Nestola, F.: Women in geosciences within the Italian University system in the last 20 years, Adv. Geosci., 53, 155–167, https://doi.org/10.5194/adgeo-53-155-2020, 2020.

Araújo, E. B., Araújo, N. A., Moreira, A. A., Herrmann, H. J., and Andrade Jr, J. S.: Gender differences in scientific collaborations: Women are more egalitarian than men, PloS one, 12, e0176 791, 2017.

Bendels, M., Müller, R., Brueggmann, D., and Groneberg, D.: Gender disparities in high-quality research revealed by Nature Index journals, PLOS ONE, 13, https://doi.org/10.1371/journal.pone.0189136, 2018.

Bernard, R. E. and Cooperdock, E. H.: No progress on diversity in 40 years, Nature Geoscience, 11, 292–295, 2018.

Blickenstaff, J. C.: Women and science careers: leaky pipeline or gender filter?, Gender and education, 17, 369–386, 2005.

Bravo-Hermsdorff, G., Felso, V., Ray, E., Gunderson, L., Helander, M., Maria, J., and Niv, Y.: Gender and collaboration patterns in a temporal scientific authorship network, Appl Netw Sci, 112, https://doi.org/10.1007/s41109-019-0214-4, 2019.

Canetto, S. S., Trott, C. D., Thomas, J. J., and Wynstra, C. A.: Making sense of the atmospheric science gender gap: Do female and male graduate students have different career motives, goals, and challenges?, Journal of Geoscience Education, 60, 408–416, 2012.

Casad, B. J., Franks, J. E., Garasky, C. E., Kittleman, M. M., Roesler, A. C., Hall, D. Y., and Petzel, Z. W.: Gender inequality in academia: Problems and solutions for women faculty in STEM, Journal of Neuroscience Research, 99, 13–23, https://doi.org/https://doi.org/10.1002/jnr.24631, 2021.

Ceci, S. J. and Williams, W. M.: Understanding current causes of women's underrepresentation in science, Proceedings of the National Academy of Sciences, 108, 3157–3162, 2011.

Dowey, N., Barclay, J., Fernando, B., Giles, S., Houghton, J., Jackson, C., Khatwa, A., Lawrence, A., Mills, K., Newton, A., et al.: A UK perspective on tackling the geoscience racial diversity crisis in the Global North, Nature Geoscience, 14, 256–259, 2021.

Dutt, K.: Race and racism in the geosciences, Nature Geoscience, 13, 2–3, 2020.

Dutt, K., Pfaff, D. L., Bernstein, A. F., Dillard, J. S., and Block, C. J.: Gender differences in recommendation letters for postdoctoral fellowships in geoscience, Nature Geoscience, 9, 805–808, 2016.

European Commission, Directorate-General for Research and Innovation: Report: She Figures 2021. Gender in Research and Innovation. Statistics and Indicators, 2021.

Ghiasi, G., Larivière, V., and Sugimoto, C. R.: On the Compliance of Women Engineers with a Gendered Scientific System, PLOS ONE, 10, 1–19, https://doi.org/10.1371/journal.pone.0145931, 2016.

Hanson, B., Wooden, P., and Lerback, J.: Age, Gender, and International Author Networks in the Earth and Space Sciences: Implications for Addressing Implicit Bias, Earth and Space Science, 7, e2019EA000 930, https://doi.org/https://doi.org/10.1029/2019EA000930, e2019EA000930 2019EA000930, 2020.

Holman, L. and Morandin, C.: Researchers collaborate with same-gendered colleagues more often than expected across the life sciences, PloS one, 14, e0216 128, 2019.

Holmes, M. A., O'connell, S., Frey, C., and Ongley, L.: Gender imbalance in US geoscience academia, Nature Geoscience, 1, 79–82, 2008.

Keller, E. F.: Gender and Science, pp. 187–205, Springer Netherlands, Dordrecht, https://doi.org/10.1007/978-94-010-0101-4_11, 2003.

Kuld, L. and O'Hagan, J.: Rise of multi-authored papers in economics: Demise of the 'lone star' and why?, Scientometrics, 114, 1207–1225, https://doi.org/10.1007/s11192-017-2588-3, 2018.

https://doi.org/10.5194/egusphere-2022-810
Preprint. Discussion started: 4 October 2022

Kwiek, M. and Roszka, W.: Gender disparities in international research collaboration: A study of 25,000 university professors, Journal of Economic Surveys, 35, 1344–1380, https://doi.org/https://doi.org/10.1111/joes.12395, 2021.

Kwiek, M. and Roszka, W.: Are female scientists less inclined to publish alone? The gender solo research gap, Scientometrics, 127, 1697–1735, https://doi.org/10.1007/s11192-022-04308-7, 2022.

Lariviere, V. and Sugimoto, C. R.: The journal impact factor: A brief history, critique, and discussion of adverse effects, in: Springer handbook of science and technology indicators, pp. 3–24, Springer, 2019.

Larivière, V., Ni, C., Gingras, Y., Cronin, B., and Sugimoto, C. R.: Bibliometrics: Global gender disparities in science, Nature, 504, 211–213, 2013.

Lerback, J. C., Hanson, B., and Wooden, P.: Association Between Author Diversity and Acceptance Rates and Citations in Peer-Reviewed Earth Science Manuscripts, Earth and Space Science, 7, e2019EA000 946, https://doi.org/https://doi.org/10.1029/2019EA000946, e2019EA000946 2019EA000946, 2020.

Lerchenmueller, M. J. and Sorenson, O.: The gender gap in early career transitions in the life sciences, Research Policy, 47, 1007–1017, 2018.

Marín-Spiotta, E., Barnes, R. T., Berhe, A. A., Hastings, M. G., Mattheis, A., Schneider, B., and Williams, B. M.: Hostile climates are barriers to diversifying the geosciences, Advances in Geosciences, 53, 117–127, 2020.

McKiernan, E. C., Schimanski, L. A., Muñoz Nieves, C., Matthias, L., Niles, M. T., and Alperin, J. P.: Meta-Research: Use of the Journal Impact Factor in academic review, promotion, and tenure evaluations, eLife, 8, e47 338, https://doi.org/10.7554/eLife.47338, 2019.

McPherson, M., Smith-Lovin, L., and Cook, J. M.: Birds of a Feather: Homophily in Social Networks, Annual Review of Sociology, 27, 415–444, http://www.jstor.org/stable/2678628, 2001.

Moss-Racusin, C. A., Dovidio, J. F., Brescoll, V. L., Graham, M. J., and Handelsman, J.: Science faculty's subtle gender biases favor male students, Proceedings of the national academy of sciences, 109, 16 474–16 479, 2012.

Muthukadan, B.: selenium-python [source code], https://github.com/baijum/selenium-python, 2022 [last update].

Nelson, B.: The Data on Diversity, Commun. ACM, 57, 86–95, https://doi.org/10.1145/2597886, 2014.

Nielsen, M. W., Alegria, S., Börjeson, L., Etzkowitz, H., Falk-Krzesinski, H. J., Joshi, A., Leahey, E., Smith-Doerr, L., Woolley, A. W., and Schiebinger, L.: Gender diversity leads to better science, Proceedings of the National Academy of Sciences, 114, 1740–1742, https://doi.org/10.1073/pnas.1700616114, 2017.

Pezzoni, M., Mairesse, J., Stephan, P., and Lane, J.: Gender and the Publication Output of Graduate Students: A Case Study, PLOS ONE, 11, 1–12, https://doi.org/10.1371/journal.pone.0145146, 2016.

Pico, T., Bierman, P., Doyle, K., and Richardson, S.: First authorship gender gap in the geosciences, Earth and Space Science, 7, e2020EA001 203, 2020.

Ranganathan, M., Lalk, E., Freese, L. M., Freilich, M. A., Wilcots, J., Duffy, M. L., and Shivamoggi, R.: Trends in the representation of women among US geoscience faculty from 1999 to 2020: The long road toward gender parity, AGU Advances, 2, e2021AV000 436, 2021.

Resmini, M.: The 'Leaky Pipeline', Chemistry – A European Journal, 22, 3533–3534, https://doi.org/https://doi.org/10.1002/chem.201600292, 2016.

Richardson, L.: beautifulsoup4 [source code], https://pypi.org/project/beautifulsoup4/, 2022 [latest release].

Ross, M. B., Glennon, B. M., Murciano-Goroff, R., Berkes, E. G., Weinberg, B. A., and Lane, J. I.: Women are Credited Less in Science than are Men, Nature, pp. 1–2, 2022.

https://doi.org/10.5194/egusphere-2022-810
Preprint. Discussion started: 4 October 2022

Sarsons, H.: Recognition for Group Work: Gender Differences in Academia, American Economic Review, 107, 141–45, https://doi.org/10.1257/aer.p20171126, 2017.

Walker, K. A.: Females Are First Authors, Sole Authors, and Reviewers of Entomology Publications Significantly Less Often Than Males, Annals of the Entomological Society of America, 113, 193–201, https://doi.org/10.1093/aesa/saz066, 2019.

West, J. D., Jacquet, J., King, M. M., Correll, S. J., and Bergstrom, C. T.: The Role of Gender in Scholarly Authorship, PLOS ONE, 8, 1–6, https://doi.org/10.1371/journal.pone.0066212, 2013.

Woolley, A. W., Chabris, C. F., Pentland, A., Hashmi, N., and Malone, T. W.: Evidence for a Collective Intelligence Factor in the Performance of Human Groups, Science, 330, 686–688, https://doi.org/10.1126/science.1193147, 2010.