LJMU Research Online

Ortega-Martorell, S, Bellfield, RAA, Harrison, S, Drewery, D, Williams, N and Olier-Caparroso, I

 Mapping the global free expression landscape using machine learning

http://researchonline.ljmu.ac.uk/id/eprint/21768/

Article

Research

Mapping the global free expression landscape using machine learning

Sandra Ortega-Martorell[1] · **Ryan A. A. Bellfield**[1] · **Steve Harrison**[2] · **Drewery Dyke**[3] · **Nik Williams**[4] · **Ivan Olier**[1]

Abstract

Freedom of expression is a core human right, yet the forces that seek to suppress it have intensified, increasing the need to develop tools that can measure the rates of freedom globally. In this study, we propose a novel freedom of expression index to gain a nuanced and data-led understanding of the level of censorship across the globe. For this, we used an unsupervised, probabilistic machine learning method, to model the status of the free expression landscape. This index seeks to provide legislators and other policymakers, activists and governments, and non-governmental and intergovernmental organisations, with tools to better inform policy or action decisions. The global nature of the proposed index also means it can become a vital resource/tool for engagement with international and supranational bodies.

Article highlights

- We propose a novel methodology using machine learning to model freedom of expression on a global scale.
- The proposed approach is in nature less prone to subjective interpretation and possibly more rigorous than previous rankings.
- The resulting freedom of expression indices can be used as a powerful tool to better inform policy or action decisions.

1 Introduction

In an increasingly atomised, polarised world, the free expression of ideas is more important than ever. But while the need for free expression has increased, so have the forces which seek to suppress it and the technologies which enable its suppression. Freedom of expression is among the core human rights set out in the United Nations (UN) Universal Declaration of Human Rights [1], the International Covenant on Civil and Political Rights

✉ Ivan Olier, I.A.OlierCaparroso@ljmu.ac.uk; Sandra Ortega-Martorell, S.OrtegaMartorell@ljmu.ac.uk; Ryan A. A. Bellfield, R.A.Bellfield@2019.ljmu.ac.uk; Steve Harrison, S.Harrison1@ljmu.ac.uk; Drewery Dyke, drewery@indexoncensorship.org; Nik Williams, nik@indexoncensorship.org | [1]Data Science Research Centre, Liverpool John Moores University, Liverpool L3 3AF, UK. [2]Liverpool Screen School, Liverpool John Moores University, Liverpool L3 5RF, UK. [3]Rights Realization Centre, London N16 7NL, UK. [4]Index on Censorship, London EC2A 3BA, UK.

(1966; entered into force in 1976) and subsequent treaties, including those in Europe, the Americas and Africa, for example, the European Convention on Human Rights [2], entry into force in 1953; the American Convention on Human Rights [3], entry into force in 1978; and the African Charter on Human and Peoples' Rights [4]; entry into force in 1986.

These global mechanisms uphold the principle that "everyone has the right to freedom of expression." Unfortunately, in today's world, this right is facing numerous challenges. Rapid advancements in technology have provided new avenues for those who wish to suppress freedom of expression. Censorship and surveillance tools are becoming more sophisticated and readily available, enabling governments and other entities to monitor and control the flow of information.

Censorship continues to operate across the globe, using several diverse tactics and drivers, including state laws or practices that restrict expression beyond what is included in international instruments [5]. Examples of this include the mixture of technological and legislative mechanisms deployed by the Chinese state to block access to online resources (colloquially called the Great Firewall of China - see for example [6]), the reduction of civil space for protests and other acts of civic participation, and the use of strategic lawsuits against public participation (SLAPPs)[1] [7] to prevent journalists and other public watchdogs from being able to report in the public interest.

With the entry into force of such standards, the UN and regional inter-governmental organisations established bodies or mechanisms to assess state adherence to the standards. This required techniques of assessment and measurement which had been developed by scholars starting in the 1930s with Greer's study into the Reign of Terror in revolutionary France [8] and which have become increasingly sophisticated in terms of data sources and statistical techniques – see, for example, [9–11]. The purpose of measuring human rights is to assess the extent to which these rights are upheld in theory, manifested in reality, and advanced through effective policies [12]. By conducting such measurements, we aim to identify areas where human rights are being violated or neglected so that appropriate solutions to address these challenges can be developed.

This research introduces the Index Index, an innovative analysis of global censorship practices, and proposes a novel methodological approach to calculate it. Specifically, the Index Index focuses on academic, digital, and media/

press freedom. It uses Generative Topographic Mapping (GTM, [13, 14]), an unsupervised Machine Learning algorithm, to cluster and visualise countries in terms of their levels of freedom of expression. By utilizing established and robust indices and metrics, this research offers a comprehensive and nuanced assessment of the international landscape of free expression. It sheds light on the various threats that impede, curtail, suppress, or manipulate the public's right to access information, express themselves, and engage with others[2]. Unlike recent studies that solely rely on data related to internet accessibility, such as [15–17], the Index Index integrates a wide range of existing analyses and expertise to provide a comprehensive ranking of the free expression environment in all countries or nations where sufficient data is available.

2 Materials and methods

2.1 Data and resources that informed the development of the Index Index

As this is an index of indices, the raw data comprises existing indices and metrics developed by a range of different national and international bodies such as research institutes, as well as international non-governmental organisations. Each pre-existing index has been selected based on several criteria, including its usage and reference by the wider community of practitioners, the robustness of its methodology, and its geographic scope. Individually, they are the product of internal testing and iterative development and as a result are used in a range of public advocacy and campaigning initiatives, including being referenced by international bodies, such as European institutions and UN bodies. For instance, V-Dem is funded by, among others, the European Commission, the Swedish Ministry of Foreign Affairs and the World Bank [18]; the World Press Freedom Index is cited by the European Parliament in its Normandy Index 2023 [19]; and the Committee to Project Journalists has submitted evidence to the UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression [20].

We selected these indices on account of their robustness and completeness. The datasets were collated after in-depth conversations between the project team. Several other sources were explored and ultimately discounted. Further details about the selected and discounted data

---

[1] SLAPPs are vexatious lawsuits targeting journalists and other whistleblowers whereby powerful individuals and institutions use civil lawsuits to intimidate and financially threaten critics [7].

[2] For the purpose of measurement, the term 'country' refers to a state or political entity, including Kosovo, Palestine, and Taiwan, which are not recognized as states by the UN.

sources are provided below and in the Supplementary Materials.

2.1.1 V-Dem (varieties of Democracy)

The Varieties of Democracy (V-Dem) Research Project [18] offers a nuanced and extensive analysis of democratisation, examining various dimensions and subcomponents. The data forming the foundation of V-Dem's component variables are collected through surveys administered to a network of over 3,500 Country Experts. The project aims to ensure a minimum of five experts for each indicator per country, facilitating a robust and diverse perspective. By employing a wide range of indicators and involving a substantial number of experts, V-Dem strives to provide a comprehensive understanding of democracy's complexities and variations across countries.

The V-Dem database offers a comprehensive range of democratic measures, surpassing the scope of the Index Index. Recognising this, the research team carefully extracted and isolated 171 variables from the extensive dataset that held significance for the model. These variables encompassed not only the three freedoms emphasised in the Index Index (academic, digital, and media/press freedom) but also encompassed broader contextual concerns, such as corruption and accountability measures, alongside various civil liberties.

2.1.2 World press freedom index

The World Press Freedom Index, compiled by Reporters Without Borders (RSF), serves the purpose of comparing the level of press freedom across 180 countries and territories [21]. It provides a snapshot of the press freedom situation in these locations during the preceding calendar year prior to its publication. The Index utilises a scoring system ranging from 0 to 100 to rank each country or territory. This score is derived from two key components: a quantitative assessment of abuses against journalists and media outlets, and a qualitative analysis of the overall situation within each country or territory.

To obtain the qualitative analysis, RSF distributes a questionnaire in 23 languages to press freedom specialists, including journalists, researchers, academics, and human rights defenders. Following the calculation of scores, the countries and territories are arranged in an ordinal list from 1 to 180, with 1 indicating the highest level of press freedom. It is this raw score calculated for each country that we have utilised as a variable in our model's development.

2.1.3 Committee to protect journalists (CPJ)

The Committee to Protect Journalists (CPJ) collects comprehensive data [22] on the imprisonment, killing, and disappearance of journalists. The CPJ's annual imprisonment census provides a snapshot of incarcerated journalists each year. However, this census does not account for the numerous journalists who are imprisoned and released throughout the year. Additionally, journalists who go missing or are abducted by non-state entities such as criminal gangs or militant groups are not included in the prison census.

Since 1992, the CPJ has maintained detailed records of journalist fatalities. Their researchers independently investigate and verify the circumstances surrounding each death. The CPJ's database encompasses both "confirmed" cases, where it is evident that a journalist was murdered as a direct reprisal for their work, during combat or crossfire, or while undertaking a hazardous assignment, as well as "unconfirmed" cases that involve unclear motives but may have a potential link to journalism. Ongoing research allows for the reclassification of cases. It is important to note that while both "confirmed" and "unconfirmed" cases are included in the CPJ's database, targeted statistical analyses only include the "confirmed" cases.

For the development of our model, we extracted the following information from the CPJ database for each country: the number of journalists and media workers killed, the number of journalists imprisoned, and the number of missing journalists. These variables serve as valuable inputs in our model development process.

2.1.4 UNESCO observatory of killed journalists

The Observatory of Killed Journalists, managed by UNESCO [23], serves as a visual representation of the institution's strategic commitment to combating impunity and addressing crimes against journalists. This initiative aligns with the General Conference 36 C/Resolution 53 (2011), which urges UNESCO to collaborate with other United Nations bodies in monitoring the state of press freedom and the safety of journalists. In order to provide comprehensive insights, the study analyses information supplied by UN Member States, which is then categorised as either Resolved or Ongoing/Unresolved, shedding light on the progress of investigations into journalist deaths. To conduct this analysis, we extracted data from the Observatory, specifically the number of journalists killed in each country, which was used as a variable in our model.

2.1.5 Cost of shutdown tool (COST)

COST [24], developed by NetBlocks, is an invaluable data-driven online service that empowers a wide range of users, including journalists, researchers, advocates, policymakers, businesses, and others, to swiftly and effortlessly generate approximate assessments of the economic impact caused by Internet disruptions. By leveraging established methodologies pioneered by esteemed institutions such as the Brookings Institution and the Collaboration on International ICT Policy for East and Southern Africa (CIPESA), COST accurately gauges the potential economic consequences of internet shutdowns, mobile data blackouts, and social media restrictions. This powerful tool utilises publicly available economic indicators that pertain to the global digital economy. We utilised the COST platform to construct an additional variable for model development, specifically capturing the hourly cost of shutdown in each country, expressed in USD.

2.1.6 Global cybersecurity index

The Global Cybersecurity Index (GCI) [25] is a reputable source that evaluates countries' dedication to cybersecurity on a global scale, with the aim of raising awareness about the significance and diverse aspects of the issue. Given that cybersecurity encompasses a wide range of applications spanning multiple industries and sectors, each country's level of development and engagement is assessed across five pillars: Legal Measures, Technical Measures, Organisational Measures, Capacity Development, and Cooperation. These pillars are then combined to form an overall score.

The GCI adopts a multi-stakeholder approach and relies on the expertise and capabilities of various organisations. Its objectives include enhancing the survey's quality, fostering international cooperation, and promoting knowledge exchange in the field of cybersecurity. The initiative is built upon the foundation and framework provided by the ITU Global Cybersecurity Agenda (GCA). To develop the model, the GCI score for each country were utilised as a variable.

2.2 Data not included in the development of the Index Index

The model does not include metrics which have no immediate bearing on, or a proxy indication of, issues relating to free expression. We nevertheless provide socio-economic data and broader contextual information that can be viewed when viewing data from a specific country on the online map that accompanies this project, in a hover-over box that appears while viewing specific country data.

The interactive map is included in the Supplementary Materials.

We included this information to provide broadly corollary metrics that immediately show texture and depth to the metrics featured. This first revived iteration of the Index Index is provided alongside contextual data on the UN Human Development Index (HDI), the Gross Domestic Product (GDP) per capita as compiled by the UN, and the Population data as compiled by the United Nations Population Fund (UNFPA), enabling the reader to explore links—if any—between this data and the core metrics.

2.3 A note on the political entities included

Our modelling and visualization are influenced by the indices comprising the dataset. This influence becomes evident through the inclusion and exclusion of various countries and political entities in the Index Index. The Index Index incorporates both UN and non-UN member states, countries with observer status, and other nations or regions that may be autonomous parts of other states. For example, Kosovo and Taiwan are included in the Index Index despite not being recognized as UN member states, while Greenland, an autonomous part of Denmark, lacks available data.

Moreover, the rankings of the British Overseas Territories, which are autonomous parts of the UK, and the overseas parts of France and the Netherlands, are attributed to their respective states. However, it is important to note that the nature of these overseas territories varies significantly.

Unfortunately, due to gaps in the datasets, the Index Index lacks data for several countries, including (but not limited to) Liberia, Papua New Guinea, Federated States of Micronesia, Kiribati, Palau, Tonga, Tuvalu, Samoa, Dominica, Saint Kitts and Nevis, Saint Lucia, Saint Vincent and the Grenadines, Grenada, Andorra, Liechtenstein, San Marino, and the Holy See.

2.4 Machine learning method used

We modelled the data using an unsupervised, probabilistic machine learning method, namely Generative Topographic Mapping (GTM) [13, 14]. The GTM is a machine learning algorithm designed for clustering, data stratification and visualisation, which has sound foundations in probability theory and provides a principled alternative to the Self-Organising Map (SOM) algorithm [26]. Rather than predicting whether two countries should be allocated the same cluster, the GTM predicts the probability of belonging to the same cluster. With this method, we created data clusters, where each of them represents a group of one or more countries that share similar characteristics. The GTM

performs a soft assignment of countries to clusters. This is a robust approach that considerably reduces the risk of countries being assigned to the wrong clusters.

The GTM assumes that the observed data is generated through a nonlinear and topology-preserving mapping from a low-dimensional latent space in $\Re^{\mathcal{L}}$ onto a manifold embedded in the high-dimensional space, $\Re^{\mathcal{D}}$, where the observed data reside. The function used to generate this embedding takes the form:

$$\mathbf{y} = \mathbf{W}\Phi(\mathbf{u}) \tag{1}$$

where $\mathbf{u}$ is a point in the L-dimensional latent space, $\mathbf{W}$ is a matrix containing parameters that govern the mapping, and Φ consists of S basis functions Φ_S, which for the standard GTM are radially symmetric Gaussians. If a prior probability distribution of $p(u)$ is defined for the latent space, then the distribution of data $\mathbf{x}$, for a given $\mathbf{u}$ and $\mathbf{W}$, is chosen to be a radially-symmetric Gaussian centred on $\mathbf{y} = \mathbf{W}\Phi(\mathbf{u})$ having a variance of β^{-1} so that:

$$p(\mathbf{x}|\mathbf{u}, \mathbf{W}, \beta) = \left(\frac{\beta}{2\pi}\right)^{\frac{D}{2}} \exp\left\{-\frac{\beta}{2}||\mathbf{y} - \mathbf{x}||^2\right\} \tag{2}$$

where $\mathbf{y}$ is as defined in (1). The GTM latent space is constrained to form a uniform discrete grid of M centres, analogous to the distribution of SOM units, in the form:

$$p(u) = \frac{1}{M} \sum_{i=1}^{M} \delta(u - u_i) \tag{3}$$

Each of these centres is responsible for generating a spherical Gaussian density function in the D-dimensional data space. In this sense, the GTM can be understood as a special case of a Gaussian mixture model in which each component in the mixture defines the probability of an observable data point (e.g., a country) given a latent centre. Therefore, assuming the observed data points x_n are independent and identically distributed (i.i.d.), the parameter matrix $\mathbf{W}$ and the inverse variance β can be determined by maximising the log-likelihood given by:

$$\mathbf{L}(\mathbf{W}, \beta|\mathbf{X}) = \sum_{n=1}^{N} \ln p(x_n|\mathbf{W}, \beta) = \sum_{n=1}^{N} \ln\left\{\frac{1}{M}\sum_{i=1}^{M} p(x_n|u_i, \mathbf{W}, \beta)\right\} \tag{4}$$

where.

$$p(x_n|u_i, \mathbf{W}, \beta) = \left(\frac{\beta}{2\pi}\right)^{\frac{D}{2}} \exp\left\{-\frac{\beta}{2}||y_i - x_i||^2\right\} \tag{5}$$

In Eq. (5), y_i is defined using Eq. (1) and is a D-dimensional point the manifold embedded in the data space for the point u_i in the latent space. The adaptive parameters of the model are optimised using the

expectation-maximisation (EM) algorithm. Matrix $\mathbf{W}$ is updated as the solution to the following system of equations:

$$\Phi^T G_{old} \Phi W_{new}^T - \Phi^T R_{old} X = 0 \tag{6}$$

where Φ is a $M \times S$ matrix with elements $\varphi_S(u_i)$; X is the observed data matrix $N \times D$ matrix with elements x_{nm}; $\mathbf{R}$ is the matrix of responsibilities that define the probability of the data point x_n being generated by the latent point u_i defined as $R_{in} = p(u_i|x_n, W_{old}, \beta_{old})$; and G is a diagonal matrix with elements $\sum_{n=1} N R_{in}$. Finally, the β parameter is updated according to the following:

$$\left(\beta^{new}\right)^{-1} = \frac{1}{ND} \sum_{n=1}^{N} \sum_{i=1}^{M} R_{in} ||y_i - x_n||^2 \tag{7}$$

Note that the observed data X requires to be normalised before training (e.g. by centring the data around zero and scaling the data so that the new standard deviation becomes 1). For the full details on the calculations, please refer to the original publication [13].

The GTM can not only assign data points to clusters but also can visualise them in a cluster membership map by projecting the latent centres. The GTM latent space can serve for visualisation purposes if its number of dimensions is 1 or 2, to which the mode probability (i.e. the highest cluster probability) is used to decide the country's cluster membership.

For the trained GTM, each cluster centre y_i, henceforth named as a reference vector, is a prototype of the data. Reference maps associated with each of the variables were generated based on the reference vector components. These reference maps can be visualised in the form of heatmaps and the high and low values can be used to interpret the relationship between each variable and each country cluster. This can provide further information/interpretation about the role of each variable used in the model.

2.5 Index Index ranking

A ranking was then generated by leveraging aggregated, normalised information from the reference maps that represent the relevant extracted variables. In this sense, a country will be given a score, which is calculated as follows:

$$Score_n = \sum_{i=1}^{M} R_{in} \tilde{y}_i \tag{8}$$

where $\tilde{y}_i$ is the normalised reference vector or centre y_i. Countries are ranked according to their calculated
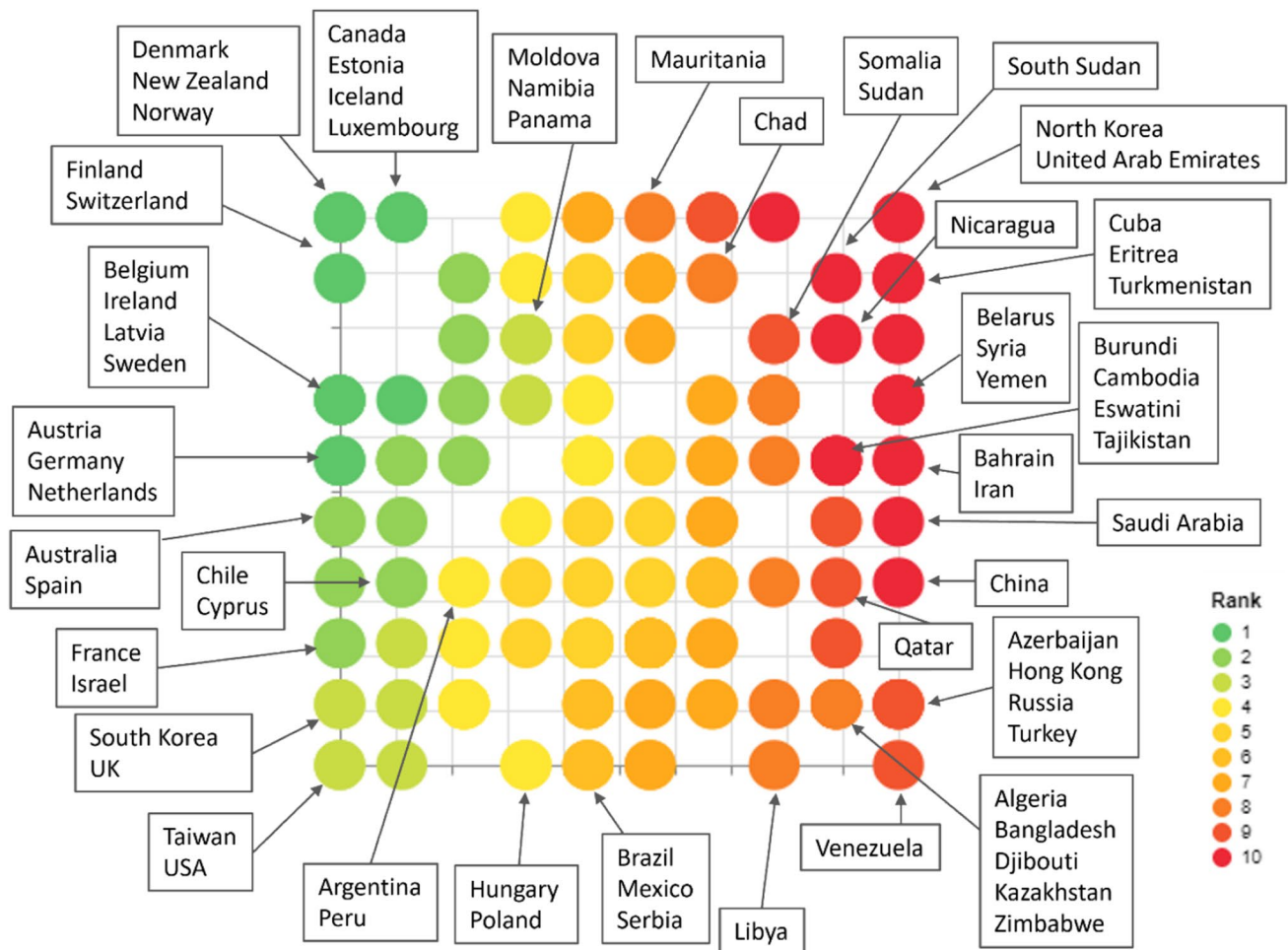
Fig. 1 Country clustering visualisation (cluster membership map) colour-coded by the cluster ranking. The countries allocated to a selection of clusters are displayed. Cluster separation indicates similarity (i.e. closer clusters are more similar than further clusters)

score. This ranking is not a direct ranking of countries, but instead, it is a ranking of the different country clusters that were automatically identified from the data using GTM. This means that in a single position of the ranking, we could have more than one country sharing such a position. The developed ranking was then divided into 10 groups according to its distribution of scores to form the 10 deciles of the scale of free expression, where lower deciles represent higher levels of free expression and higher deciles represent lower levels.

## 3 Results

3.1 Country clustering visualisation

The visualisation in Fig. 1 (representing the cluster membership map in the GTM latent space of the developed model) shows a representation of a different kind

of world map, where every circle represents a cluster, and each cluster is representing one or more countries. In accordance with the original GTM publication [13], we set the number of clusters to 100 (arranged in a grid of 10 × 10) and the number of basis functions to 16 (arranged in a grid of 4 × 4). The GTM regularisation term was optimised, and the one resulting in the lowest error (negative log-likelihood) was selected (Table S1, Supplementary Material). As discussed earlier, the GTM predicted the probability of countries belonging to the same clusters in the below visualisation. The top-left-hand side of the visualisation represents the highest deciles of free expression, while the top right represents the lowest. This visualisation of the data is intended to help identify commonalities or differences and related factors to better understand the changing free expression landscape. Figure 1 shows the countries allocated to a selection of clusters. The full allocation of countries per cluster can be found in the Supplementary Materials.
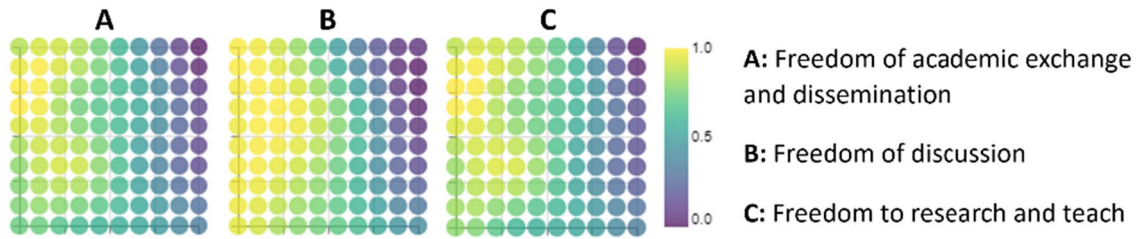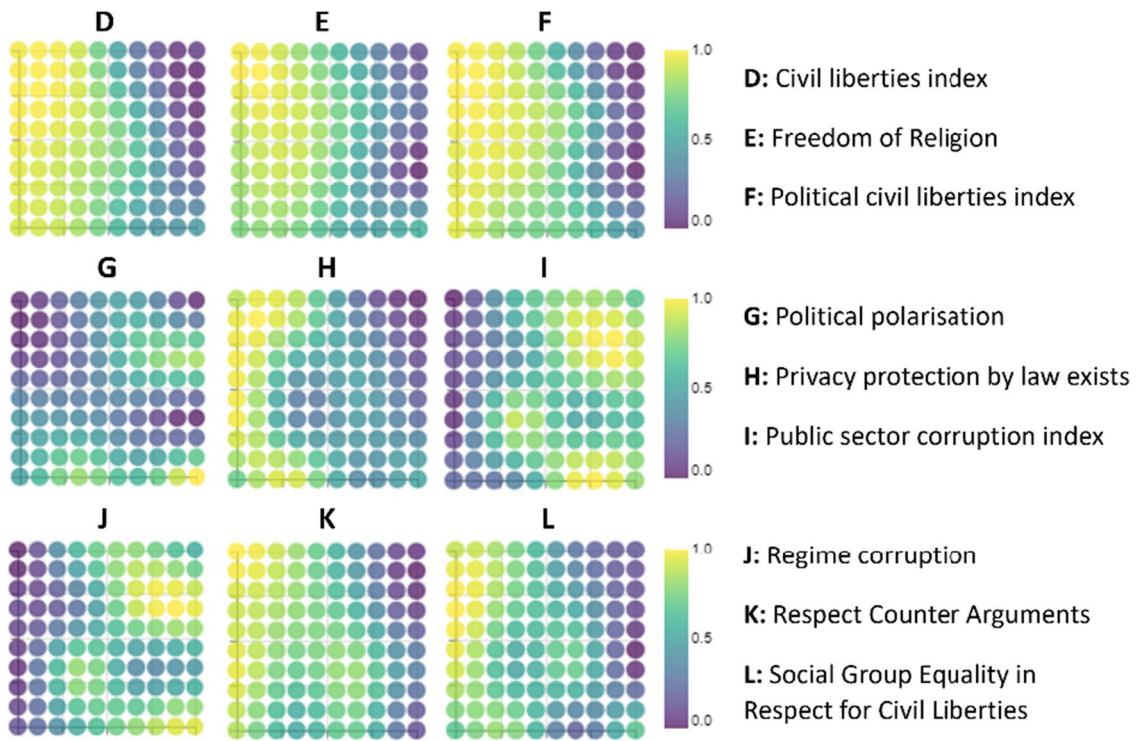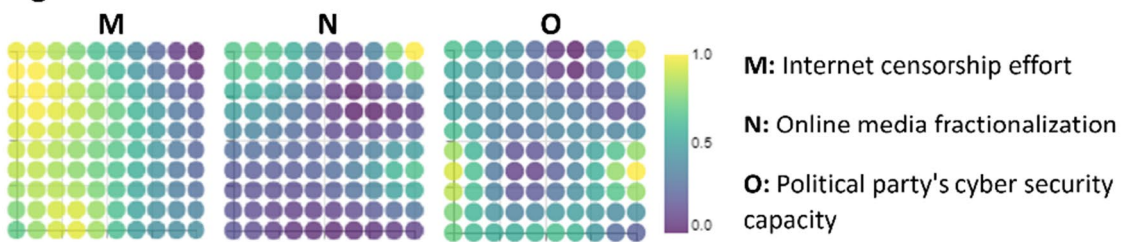
Fig. 2 Selected reference maps for 15 of the variables used to produce the GTM model

3.2 Visualisation of the reference maps

A selection of reference maps is presented in Fig. 2, showing the distribution of the clusters (and therefore countries) against the selected variables. They are organised by IoC freedom index areas: academic, media and digital freedom. The reference maps corresponding to all the variables used can be found in the Supplementary Materials.

3.3 Global ranking of countries/nations—deciles

The Index Index groups states' free expression ranking into ten categories - deciles - intended to convey the complexity and nuance of the global practice of

Table 1 Global ranking of free expression by deciles. Lower ranks represent higher levels of free expression while higher ranks represent lower levels of freedom

Countries and nations	Global rank
Austria, Belgium, Canada, Denmark, Estonia, Finland, Germany, Iceland, Ireland, Latvia, Lithuania, Luxembourg, Netherlands, New Zealand, Norway, Sweden, Switzerland	1
Australia, Barbados, Cape Verde, Chile, Costa Rica, Cyprus, Dominican Republic, France, Israel, Italy, Jamaica, Japan, Malta, Portugal, Slovakia, Spain, Trinidad and Tobago, Uruguay	2
Czechia, Greece, Moldova, Namibia, Panama, Romania, South Africa, South Korea, Suriname, Taiwan, Tunisia, United Kingdom, United States of America, Vanuatu	3
Argentina, Armenia, Benin, Botswana, Bulgaria, Croatia, Georgia, Ghana, Guyana, Hungary, Kosovo, Mongolia, Montenegro, Peru, Poland, Sao Tome and Principe, Senegal, Seychelles, Slovenia, Solomon Islands, Timor-Leste	4
Albania, Ecuador, Guatemala, Guinea-Bissau, Honduras, Madagascar, Malawi, Maldives, Mauritius, Mozambique, Niger, Nigeria, Paraguay, Sierra Leone, The Gambia	5
Angola, Bhutan, Bolivia, Bosnia and Herzegovina, Brazil, Indonesia, Ivory Coast, Jordan, Kenya, Kyrgyzstan, Lesotho, Mexico, Nepal, North Macedonia, Philippines, Serbia, Singapore	6
Burkina Faso, Central African Republic, Colombia, Comoros, Democratic Republic of the Congo, El Salvador, Fiji, Gabon, Haiti, India, Kuwait, Lebanon, Malaysia, Mali, Morocco, Pakistan, Sri Lanka, Tanzania, Togo, Ukraine, Zambia	7
Algeria, Bangladesh, Cameroon, Chad, Djibouti, Ethiopia, Guinea, Iraq, Kazakhstan, Libya, Mauritania, Rwanda, Thailand, Uganda, Zimbabwe	8
Afghanistan, Azerbaijan, Egypt, Hong Kong, Oman, Palestine, Qatar, Republic of the Congo, Russia, Somalia, Sudan, Türkiye, Uzbekistan, Venezuela, Vietnam	9
Bahrain, Belarus, Burma/Myanmar, Burundi, Cambodia, China, Cuba, Equatorial Guinea, Eritrea, Eswatini, Iran, Laos, Nicaragua, North Korea, Saudi Arabia, South Sudan, Syria, Tajikistan, Turkmenistan, United Arab Emirates, Yemen	10

The countries within each grouping are ranked alphabetically and do not present a ranking within the groupings
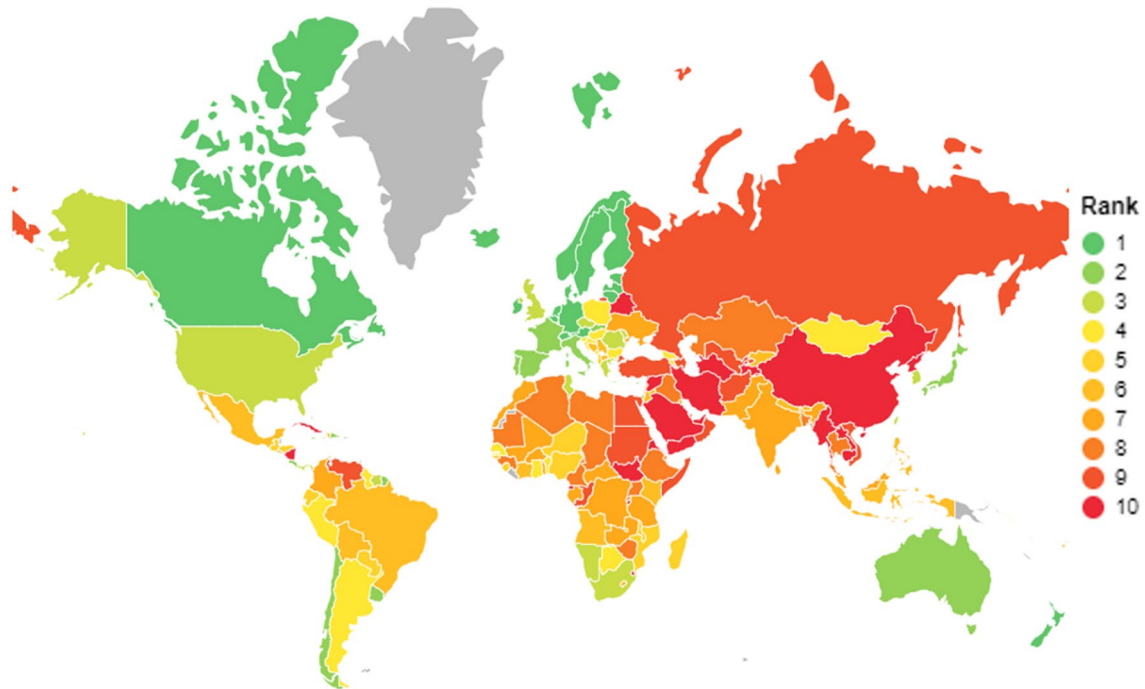

Fig. 3 World map showing the global free expression ranking

censorship, see Table 1. The deciles ensure the eventual ranking does not erase distinctions between countries/ nations, but also presents a clear picture of the global free expression environment. A world map representation showing the global ranking of censorship by deciles is shown in Fig. 3, with the highest deciles of free expression represented in green (lowest values), and the lowest levels in red (highest values). The rankings per area of freedom (academic, digital and media) can be found in the Supplementary Materials.

4 Discussion

4.1 Creating meaningful representations using GTM

Due to the challenges of data collection and data representation, there exist high levels of uncertainty that could potentially have a negative impact on the modelling process. GTM, being a robust probabilistic algorithm, calculates the probability of a cluster being responsible for a country while accounting for this uncertainty. In this analysis, the GTM cluster centres or prototypes serve as representations of freedom of expression, effectively stratifying the landscape of freedom of expression. A crucial property of GTM is the preservation of data topology, signifying that similar clusters will be positioned closer together in the latent space. Consequently, even if the most probable cluster assigned to a country does not precisely correspond to the actual one, it is expected to be closer to the correct one. In contrast, popular clustering techniques such as k-means, lacking probabilistic foundations, are not specifically designed to handle such levels of uncertainty.

In addition, GTM is particularly useful for crafting meaningful data representations by transforming high-dimensional information into a lower-dimensional space while retaining the intrinsic structure of the data. Alternative visualisation algorithms such as t-SNE [27] and UMAP [28] have gained popularity for data visualisation through dimensionality reduction. However, these techniques do not possess the capability to extract data prototypes in the manner that GTM does, which poses a challenge when it comes to stratifying countries based on freedom of expression. In contrast, GTM creates a visualisation (the cluster membership map) that captures the underlying patterns, relationships, and clusters within the data by mapping data points to these prototypes. This process allows for a more comprehensible and interpretable depiction of complex data, aiding in knowledge extraction and facilitating insights that might otherwise remain hidden in the original high-dimensional space. GTM has found applications in various real-world scenarios across different domains. In bioinformatics, it has been used to model protein structures and understand their conformational spaces, providing insights into protein folding and function, which is crucial for drug design [29], disease understanding [30], and other biomedical applications [31–33]. It has also been used to model species distributions and understand ecological patterns, e.g., to understand species composition of a forest to assess biodiversity [34], and to study the ecological status of streams [35]. It has also been used in the financial sector, e.g., for early identification of business opportunities [36]. These examples highlight the versatility of the GTM in addressing real-world challenges across diverse fields. However, to the best of our knowledge, GTM has not been used before to study censorship or freedom of expression, hence making this a positional article in the application of GTM within this field.

4.2 Interpreting the visualisations (membership and reference maps)

The country clustering visualisation (cluster membership map) presented in Fig. 1 provides another way to examine the data. It can then be used to show: (i) the details of the individual countries within each cluster, indicating that they share very similar characteristics; (ii) the location of the countries across all clusters, allowing for the representation of a certain degree of similarity if they are allocated to neighbouring clusters; and (iii) the assigned colour-coded ranking to each of the clusters, and therefore to the countries that these clusters represent.

The reference maps provide further information/interpretation about the role played by each variable in the development of the GTM model, with high values representing areas of the maps where the variables had a higher influence, and low values representing otherwise. When exploring the reference maps of the academic freedom variables from Fig. 2, which include freedom of academic exchange and dissemination (Fig. 2A), freedom of discussion (Fig. 2B), and freedom to research and teach (Fig. 2C); we can see that the higher values for those variables are on the left-hand side of the reference maps, which coincide with the areas with better rankings of freedom (see Fig. 1).

Regarding the media freedom variables, we can also see high values on the left-hand side of Fig. 2D and F which represent civil liberties and political civil liberties, respectively. In the case of the public sector corruption index (Fig. 2I), we see high values in the top right quadrant where the clusters represent countries such as Nicaragua, Yemen, Somalia, and Eswatini, among others. Also in this

area, we see high values in the reference map of regime corruption (Fig. 2J).

In the case of the digital freedom variables, we can see high levels of online media fractionalization (Fig. 2N) in the cluster of Eritrea, North Korea, and the United Arab Emirates, and high levels of internet censorship effort (Fig. 2M, which higher values meaning that the governments allow generally unrestricted Internet access) in countries represented by a higher level of freedom (left-hand side of Fig. 1). These examples illustrate how the role of each of the variables used to produce the GTM model can be studied by visualising their respective reference maps.

4.3 Insights from the global ranking of countries/nations

A closer inspection of the global ranking in Table 1 and the rankings in the different areas of freedom (academic, digital, and media/press) show that Europe dominates the list of countries that were in the 1st decile (least censorship/greatest freedom) for all three freedoms. These include Austria, Belgium, Denmark, Estonia, Finland, Germany, Iceland, Ireland, Latvia, Lithuania, Luxembourg, Netherlands, Norway, Sweden and Switzerland. The G20 Member States are spread across the full Index Index. Using the global ranking, Australia, Canada and Germany are the highest place members (1st decile), with Saudi Arabia and China being the lowest (10th decile).

For the global ranking, G7 Member States are placed: Canada = 1st, France = 2nd, Germany = 1st, Italy = 2nd, Japan = 2nd, United Kingdom = 3rd and USA = 3rd decile.

Much like G20 Members, UN Security Council members, including both permanent and non-permanent members, are spread across the full Index. Using the global ranking, Ireland and Norway are the highest place members (1st decile) and China and the United Arab Emirates are the lowest (10th decile). Out of the Permanent members, France (2nd decile) is the highest-ranking member, with Russia (9th decile) the lowest. Across the three freedoms, the United Kingdom is consistently found in the 3rd decile. This is similar to the United States of America. However, the latter is in the 4th decile for academic freedom.

The countries that were in the 10th decile for all three freedoms are Bahrain, Belarus, Burma/Myanmar, Cuba, Equatorial Guinea, Eritrea, Iran, Laos, North Korea, Syria, Turkmenistan, United Arab Emirates and Yemen.

4.4 Use and potential impact of the Index Index

By making available indices that provide objectively verifiable, clearly ranked data about rates of freedom of expression, in contrast to or perhaps as linked to academic freedom, the Index Index seeks to provide legislators and other policymakers, activists and governments, and non-governmental and intergovernmental organisations, with tools to better inform policy or action decisions. Developing a wide range of campaigning and advocacy tools that can benefit from emergent and innovative technologies and research approaches to synthesize and present compelling and data-rich information is vital to ensure rights advocacy is underpinned by all available expertise that can be accessed easily and clearly. As seen in previous metrics, including those that are incorporated into the dataset for the Index Index, empirical data generated by this pilot project can be highly effective when communicated with policymakers to encourage more affirmative action when it relates to free expression, including more robust protection for journalists [37, 38], the formulation of rights policies for educational institutions and ensuring all surveillance policies deployed for policing or national security purposes are rights-respecting. These are a few examples of how the Index Index can be used but should not be assumed to limit how it can be used by a wide range of stakeholders.

While the Index Index abstracts from the particular experiences of writers, journalists and academics facing daily repression across the globe, the overall ranking hints at what is at stake. It constitutes a call, directing the attention of those with a voice to denounce it, to where free expression is at greatest risk and providing insights into the granular policy areas needing attention. The global nature of the proposed index also means it can become a vital resource and tool for engagement with international and supranational bodies such as the United Nations, as well as other regional mechanisms such as the European Union, Council of Europe, African Union and the Inter-American Commission on Human Rights, whose work requires country-by-country, regional and global data sources.

As the Index Index is an index of existing respected and trusted indices and metrics it depends on robust and accurate data produced by the wider community of experts. The process of compiling and producing the Index has demonstrated its own use-case as it has identified the need for increased monitoring, verification and sharing of granular country-by-country level data on a wide range of markers against free expression more broadly, as well as academic, artistic, digital and media/press freedom. While also strengthening further iterations of this pilot project, this will also strengthen the global movement to protect free expression.

In this, too, the study provides the basis for developing insights into the political economy of censorship and freedom which shine a light - not always flattering - on human conduct towards others in our midst. Objective data and analysis provided by the Index Index encourage us to ask,

simply, what will it take for us to live less censored lives and what must we do to achieve greater respect towards human dignity.

5 Conclusions

This project collected and collated pre-existing, robust data on the status of the free expression landscape on a global scale. We modelled the data using the GTM, an unsupervised, probabilistic machine learning method, to explore whether the model produced new insights into state conduct, human rights, and governance. The use of such a model removes an element of subjective interpretation from the modelling process and provides the resulting Index Index with a greater degree of rigour than previous rankings.

On close examination, the reader can be expected to find unexpected outcomes that call into question, correlation or causality. The Index Index provides a powerful policy tool for all those seeking a clear picture of the health of the free expression environment, as well as what needs to happen to change the rankings.

Author contributions Conceptualization, I.O. and S.O.M.; methodology, I.O., S.O.M. and R.A.A.B; software, R.A.A.B, I.O. and S.O.M.; validation, N.W., D.D., and S.H.; formal analysis, I.O., S.O.M. and R.A.A.B; resources, N.W., D.D; data curation, R.A.A.B and D.D.; writing—original draft preparation, all authors; writing—review and editing, all authors; visualization, R.A.A.B, I.O. and S.O.M.; supervision, I.O. and S.O.M.; project administration, I.O.; funding acquisition, I.O. and N.W. All authors have read and agreed to the published version of the manuscript.

Data availability The data will be made available with the paper.

Declarations

References

1. United Nations. Universal Declaration of Human Rights (1948) https://www.un.org/en/about-us/universal-declaration-of-human-rights. Accessed December 2, 2022
2. European Court of Human Rights (1950) European Convention on Human Rights.
3. Organization of American States (1969)
4. African States members of the Organisation of African Unity (1979) African Commission on Human and Peoples' Rights Legalinstruments.
5. Bollinger LC, Callamard A (2021) Regardless of frontiers: global freedom of expression in a troubled world.
6. Stevenson C (2007) Breaching the great firewall: China's internet censorship and the Quest for Freedom of Expression in a Connected World. Boston Coll Int Comp Law Rev 30:531–558
7. Pring GW, Canan P, SLAPPs (1996) Getting sued for speaking out. Temple University Press
8. Greer D (1935) The incidence of the terror during the French Revolution: a statistical interpretation. Harv Hist Monogr. ;196
9. Landman T, Schwarz K (2022) Human rights indicators and implementation. In: Murray R, Rachel H, Long D (eds) Human rights indicators and implementation. Edward Elgar Publishing, pp 309–326
10. Restrepo JA, Spagat M, Vargas JF (2016) J Peace Res 43:99–115. https://doi.org/10.1177/0022343306059924. Special Data Feature; The Severity of the Colombian Conflict: Cross-Country Datasets Versus New Micro-Data
11. Ball P, Asher J, Sulmont D, Manrique D, For DM (2003) How many peruvians have died? An estimate of the total number of victims killed or disappeared in the armed internal conflict between 1980 and 2000. American Association for the Advancement of Science
12. Todd L (2010) Carvalho Edzia. Measuring human rights. Routledge
13. Bishop C, Svensén M, Williams C (1998) GTM: the generative topographic mapping. Neural Comput 10:215–234
14. Olier I, Vellido A (2008) Advances in clustering and visualization of time series using GTM through time. Neural Netw 21:904–913. https://doi.org/10.1016/J.NEUNET.2008.05.013
15. Simurgh Aryan J (2013) Alex Halderman. Internet Censorship in Iran: A First Look. 3rd USENIX Workshop on Free and Open Communications on the Internet.
16. Dainotti A, Squarcella C, Aben E, Claffy KC, Chiesa M, Russo M et al (2014) Analysis of country-wide internet outages caused by censorship. IEEE/ACM Trans Netw (TON) 22:1964–1977. https://doi.org/10.1109/TNET.2013.2291244
17. Niaki AA, Cho S, Weinberg Z, Hoang NP, Razaghpanah A, Christin N et al (2020) ICLab: a global, longitudinal internet censorship measurement platform. Proc IEEE Symp Secur Priv 2020–May:135–151. https://doi.org/10.1109/SP40000.2020.00014
18. Coppedge M, Gerring J, Knutsen CH, Lindberg SI, Teorell J, Altman D et al (2022) V-Dem codebook v12 varieties of Democracy (V-Dem) project. https://www.v-dem.net/data/the-v-dem-dataset/
19. Lazarou E, Stanicek B (2023) Mapping threats to peace and democracy worldwide. https://www.europarl.europa.eu/RegData/etudes/STUD/2023/751422/EPRS_STU(2023)751422_EN.pdf
20. Radsch CC, Paterson K (2021) Submission from the Committee to Protect Journalists to the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression. https://www.ohchr.org/sites/default/files/Documents/Issues/Expression/disinformation/2-Civil-society-organisations/Committee-to-Protect-Journalists.pdf

21. Reporters Without Borders. World Press Freedom Index (2022) https://rsf.org/en/index?year=2022

22. Committee to Protect Journalists. CPJ's database of attacks on the press (2022) https://cpj.org/data/

23. UNESCO (2022) UNESCO observatory of killed journalists. https://en.unesco.org/themes/safety-journalists/observatory

24. NetBlocks COST (2022) : The NetBlocks Cost of Shutdown Tool. https://netblocks.org/projects/cost

25. International Telecommunication Union (ITU). Global Cybersecurity Index (2022) https://www.itu.int/en/ITU-D/Cybersecurity/Pages/global-cybersecurity-index.aspx

26. Kohonen T (2001) Self-Organizing maps. vol. 30. Third exte. Springer, New York

27. van der Maaten L (2008) Visualizing data using t-SNE. J Mach Learn Res. https://doi.org/10.1007/s10479-011-0841-3

28. McInnes L, Healy J, Saul N, Großberger L (2018) UMAP: Uniform Manifold approximation and projection. J Open Source Softw 3:861. https://doi.org/10.21105/JOSS.00861

29. Horvath D, Marcou G, Varnek A (2019) Generative topographic mapping in drug design. Drug Discov Today Technol. https://doi.org/10.1016/j.ddtec.2020.06.003

30. Orlov AA, Khvatov EV, Koruchekov AA, Nikitina AA, Zolotareva AD, Eletskaya AA et al (2019) Getting to know the Neighbours with GTM: the Case of Antiviral compounds. Mol Inf. https://doi.org/10.1002/minf.201800166

31. Bathen TF, Engan T, Krane J, Axelson D (2000) Analysis and classification of proton NMR spectra of lipoprotein fractions from healthy volunteers and patients with cancer or CHD. Anticancer Res. ;20

32. Gaspar HA, Hübel C, Breen G (2019) Biological pathways and Drug Gene-Sets: analysis and visualization. Eur Neuropsychopharmacol 29. https://doi.org/10.1016/j.euroneuro.2017.08.095

33. Olier I, Amengual J, Vellido A (2011) A variational bayesian approach for the robust analysis of the cortical silent period from EMG recordings of brain Stroke patients. Neurocomputing 74. https://doi.org/10.1016/j.neucom.2010.12.006

34. Polyakova A, Mukharamova S, Yermolaev O, Shaykhutdinova G (2023) Automated Recognition of Tree species Composition of Forest communities using Sentinel-2 Satellite Data. Remote Sens (Basel). https://doi.org/10.3390/rs15020329

35. Vellido A, Martí E, Comas J, Rodríguez-Roda I, Sabater F (2007) Exploring the ecological status of human altered streams through Generative Topographic Mapping. Environ Model Softw 22. https://doi.org/10.1016/j.envsoft.2006.06.005

36. Feng J, Liu Z, Feng L (2021) Identifying opportunities for sustainable business models in manufacturing: application of patent analysis and generative topographic mapping. Sustain Prod Consum 27. https://doi.org/10.1016/j.spc.2021.01.021

37. Irum SA, Laila AS (2015) Media censorship: Freedom versus responsibility. J Law Confl Resolution 7:21–24. https://doi.org/10.5897/JLCR2015.0207

38. Clark M, Grech A (2017) Council of Europe. Journalists under pressure: unwarranted interference, fear and self-censorship in Europe.