Shifting to broad patterns of interests in conservation culturomics with Bayesian dynamic factor analysis

by Jacopo Cerri[1], Lucilla Carnevali[2], Alessandro Piazzi[3], Piero Genovesi[2], Andrea Monaco[2], Sandro Bertolino[4]

1. *Faculty of Mathematics, Natural Sciences and Information Technologies, University of Primorska, Glagoljaška 8, 6000 Koper, Slovenia* jacopo.cerri@famnit.upr.si
2. *Institute for Environmental Protection and Research (ISPRA), Via Bracati 48, 00144, Rome, Italy*
3. *Legambiente Onls, Via Salaria 13, 00199 Rome, Italy*
4. *Department of Life Sciences and Systems Biology, Università degli Studi di Torino, Via Accademia Albertina 13, 10123, Torino, Italy*

Abstract Conservation culturomics aims to explore changes in public interest towards biodiversity and conservation, often by analyzing the temporal evolution of Internet searches. However, studies failed to embed the analysis of multiple time series within a formal modeling framework, where these are indicators of interest, an unobserved latent variable. In this study we adopted Bayesian dynamic factor analysis to see whether searches on Google and Wikipedia, in Italy, about queries associated to invasive alien species, could be considered indicators on increased interest about the topic. We downloaded the monthly Google Trends index (2010-2019) for the equivalent Italian words "*Alien species*" (singular), "*Alien species*" (plural), "*Invasive species*" (singular), "*Allochthonous species*" (singular) and the number of visits to the three pages from the Italian Wikipedia about this topic. Google searches and visits to Wikipedia reflected a common latent factor, therefore a common interest of Internet users about biological invasions. Factor scores increased for Google but not for Wikipedia. Interest towards biological invasions increased over the last 10 years in Italy, with a growing number of persons searching for related queries on Google. Visits to Wikipedia however did not increase, maybe because people did not document themselves in-depth. Conservation culturomics can therefore benefit by the use of multivariate time series analysis, to consider multiple latent trends as indicators of unobserved processes, like increased awareness about conservation topics. In turn this would make researchers better understanding trends in interests towards different environmental topics, and to evaluate policies and outreach initiatives.

Warning

This is a ***preprint***, not a peer-reviewed study. If you do not know what a preprint is, we encourage you to read more about this type of documents (https://en.wikipedia.org/wiki/Preprint), before evaluating and citing the study. The current version is an update of the preprint initially known as "*A nation-wide analysis of Wikipedia and Google Searches in Italy reveals a growing interest towards biological invasions*". The title was changed, because the study adopted an entirely new approach, which focused on multivariate time series analysis. However, as the data basis is the same, and conclusions are very similar to previous versions, we maintained the same doi for the study. Previous versions can be downloaded from EcoEvoRxiv.

https://doi.org/10.32942/osf.io/afrpk

Introduction

Conservation culturomics is a field of research attempting to capture changing interactions between people and nature through the analysis of big digital data, retrieved from the Internet[1]. This approach, while not without limitations[2], is extremely powerful in a world pervaded by the Internet, where people regularly access the web to search for information about the most diverse topics[[3][4].

The time-series analysis of volumes of searches on Google and Wikipedia became particularly prominent in conservation culturomics[5]. Nowadays, Google dominates by far the global market of search engines (https://www.webfx.com/blog/seo/2019-search-market-share/) and Wikipedia is the largest online encyclopedia, with 24 billion views in 2020 (https://stats.wikimedia.org//all-projects). AAvailable evidence, from diverse disciplines, shows that the time-series analysis of searches on these two websites is capable of revealing both short-term fluctuations (e.g. insurances)[6] and long-term trends (e.g. sexual dysfunctions; immigration)[7] [8] in public attention about different issues, as well as the power of traditional media to leverage public attention and drive searches on the Internet[9].

While some studies focused on cyclical patterns of internet searches[10], or to short-term responses[11] a significant portion of conservation culturomics explored long-term trends in Internet searches. These can reflect stable changes in values, attitudes and awareness, which would ultimately lead to increased support to environmental and conservation policies, in turn affecting their weight in the political agenda[12]. Nevertheless, existing studies did not explicitly conceive multiple time-series as indicators for broader latent proceses. While researchers often considered multiple time-series at a time, and interpreted their temporal evolution as a barometer for changes in public attention, they did not explicitly embedded them into an analytical framework accounting for this process-indicator structure. Even studies developing specific metrics for temporal changes, designed them to be used over single time series[13].

This approach has two limitations, when adopted to identify long-term trends in public attention about conservation. First it can suffer from species-specific fluctuations, whose drivers might be unknown and misleading. When searches about many species are combined together, each species can experience its own temporal fluctuations and long-term trajectories, depending upon factors other than increased interest in conservation. For example, species can peak in their searches, not because people are interested in their conservation, but because media covered their role in zoonoses[14][15] or because they became viral and widespread as food or pets.

This approach is also intrinsically limiting for the study of topics whose definition is complex and cannot be reduced to a single term, like invasive alien species (IAS)[16]. IAS are a major driver of global change, often with negative consequences for biodiversity and human well-being[17][18], and are increasingly targeted by public policies aimed at preventing their spread, mitigating their impacts and communicating their importance to stakeholders and the general public[19]. However, while IAS is an unambiguous term, many synonyms are often used in its place during communication with laypersons (e.g. “allocthonous species”, “exotic species”), as they sound more familiar. Therefore the general public is exposed to many synonyms, when instructed about biological invasions and IAS. To quantify any increased interest into this topic, researchers need to go beyond single time-series and consider the conjoint trajectory of searches about these synonyms, over time.

In this study we aim to provide a first example of how this can be achieved by identifying long-term trends in Internet searches about biological invasions and IAS in Italy, through Bayesian Dynamic Factor Analysis, an approach which conceptualizes single time-series as indicators of latent-unobserved processes.

Methods

Study area

In this study, we restricted the analysis of on-line searches on Google and Wikipedia to Italy. Italy is one of the member states of the European Union (EU), where biological invasions raise the most serious concerns, due to the abundance of native biodiversity in Mediterranean ecosystems and the pressures these are already facing due to global change[20]. Italy is an invasion hotspot[21][22], many Italian freshwaters are highly invaded[23][24] and Italian seas are facing the colonization of Lessepsian species[25]. Overall, with more than 3000 IAS that established populations on its territory, Italy hosts more than 20% of European IAS(https://easin.jrc.ec.europa.eu/easin).

The first dedicated policies about IAS date back to the 1990s[26] and over the last two decades more than thirty projects were implemented, almost entirely within the LIFE+ (https://ec.europa.eu/easme/en/life) and the Interreg (https://www.interregeurope.eu/) programs. AAll these project included dissemination and outreaching initiatives about IAS and biological invasions and in 2015 a dedicated LIFE+

project, named "ASAP – Alien Species Awareness Program" (https://www.lifeasap.eu/index.php/it/) was approved with the aim to increase the awareness and the active participation of citizens on topics related to biological invasions and to promote the correct and efficient management of IAS.

In 2014 the European Union produced its first Regulation about biological invasions (EU Regulation n. 1143/2014) on the prevention and management of IAS, which entered into force in January 2015[27][28]. This Regulation includes a list of IAS of Union Concern, firstly adopted in 2016 (hereinafter the Union list)[29] and updated twice since then; nowadays it includes 66 species. Transposing the EU Regulation, Italy adopted a national legislative decree (n. 230/2017), which identifies management actors, with various competences in terms of biosecurity, management of IAS and the estimation of their impacts. Italy was therefore a good case study, because it is a country were biological invasions are widespread and also where many policy actions and communication about the topic were implemented.

In terms of study design, restricting our analysis to Italy also provided us with some advantages, in terms of data interpretation. The Italian language does not have similar languages in Europe, as it happens with countries in the Balkan sprachbund or in Scandinavia, and there are no other countries where Italian is a first or a second language. This makes reasonable to assume almost all the searches on the Italian version of Google and Wikipedia come from Italian citizens. Moreover, Italy almost doubled its number of Internet users between 2010 and 2020, passing from 46.8% of residents in 2010 to 84.0% of households in 2018. As we will discuss in the next subsection, this increase strongly deflates the GoogleTrends metric, enabling us to really identify a growing number of searches about IAS-related terms, net the effect of changes in the number of Internet users. Italy is also a suitable spatial scale to analyze GoogleTrends data, because Google is the leading search engine at the national level, accounting for 89.9% of online searches from desktop and 98.7% of searches from mobile devices [30][31].

Study design and data collection

We analyzed the volume of online searches about invasive species in Italy, between January 2010 and December 2019. We compared the monthly values of the GoogleTrends index and the overall number of Wikipedia searches. The GoogleTrends index is obtained by: *i*) dividing the monthly number of searches for a certain keyword (e.g. "Alien species"), for the total volume of Google searches in the same timespan, then by *ii*) dividing this ratio for the maximum value of the time series and *iii*) multiplying by 100. It is therefore a relative index, which can be informative about long-term trends in Google searches, discounting all those that do not remain popular through time. Considered that access to the Internet steadily increased over the last 10 years in Italy [30], a stable or increasing GoogleTrends index indicates that a certain keyword was searched by an increasing number of people, as the total increase in the number of searches was greater than the total increase in Google usage.

Visits to Wikipedia pages can be expressed as raw counts showing the total number of accesses. Wikipedia searches can complement GoogleTrends, by showing the real magnitude of change in information-searching behavior through time: while GoogleTrends might fail to show changes in online searches, when performed by a minority of Internet users, the number of Wikipedia searches can trace even minor changes. Also, information-searching on the Internet can be regarded as a two-stage process: most people first search for a certain keyword on Google, then some of them move to the dedicated Wikipedia page, to obtain more in-depth information about the topic. Given this mechanism, GoogleTrends would measure the real interest of the general public, while Wikipedia traffic would measure trends in motivated users of the Internet which want to gain a deeper level of knowledge over a certain topic. Therefore, we modeled Google and Wikipedia data separately, with two different models.

We collected GoogleTrends data with the 'gTrendsR' package in R [32], extracting values between January 2010 and September 2019. We queried for 10 common Italian terms describing IAS or biological invasions: "*specie aliena*", "*specie aliene*" (literally, "*alien species*", singular and plural), "*specie invasiva*", "*specie invasive*" ("*invasive species*", singular and plural), "*specie alloctona*", "*specie alloctone*" ("*allochthonous species*", singular and plural), "*specie esotica*", "*specie esotiche*" ("*exotic species*", singular and plural) and "*invasione biologica*", "*invasioni biologiche*" ("*biological invasions*", singular and plural). As GoogleTrends data were based on a relative index, we could not aggregate the five time series and we considered them separately. We also collected Wikipedia data through R, with the 'pageviews' package. We extracted the daily number of visits to Wikipedia pages about IAS in Italian, associated with the following terms: "*specie aliena*" or "*specie alloctona*" ("*alien*" and "*allochthonous species*", same page, created in December 2005), "*specie invasive in Italia*" (invasive species in Italy, created in May 2013), "*specie invasive in Europa*" (invasive species in Europe, created in March 2013). To compare Wikipedia searches with GoogleTrends, data were aggregated on a monthly basis. For Wikipedia, data were extracted from July 2015, as the platform does not allow to download previous information (https://pageviews.toolforge.org/).

https://doi.org/10.32942/osf.io/afrpk

Data analysis

To see if observed time-series of Internet searches reflected a unique latent trend of general interest about IAS, we adopted Bayesian dynamic factor analysis (hereinafter, BDFA). In classic factor analysis observed data (Y_i) are a linear combination of latent factors (F_i), their loadings (Z), an intercept (u_i) and an error term (ϵ_i):

$$Y_i = u_i + Z\mathbf{F}_i + \epsilon_i$$

Both loadings and latent factor scores can be estimated, based on various inferential approaches. For example, factor scores are often of interest for dimension-reduction problems, offering various advantages compared to principal component analysis, like greater flexibility in the distribution of observations and the estimation of error terms.

In dynamic factor analysis [33], the dimensionality of multiple time-series is reduced by estimating multiple shared factors, each one affecting observations. Latent processes are usually modeled as random walks, where the value of i_{th} trend at time $t + 1$ ($x_{i,t+1}$) depends on its past values ($x_{i,t}$) plus some random deviation ($w_{i,t}$), approximated as a multivariate Normal or Student distribution. Then latent trends ($x_{i,t}$) are linked to observations (y_t) through a matrix of factor loadings, whose values do not evolve in time ($\mathbf{Z}_{xt}$), and some additional parameters, like time-varying intercepts and (a) and effects (B) of time-varying covariates (d_t):

$$y_t = \mathbf{Z}x_t + a + Bd_t + \epsilon t$$

This formulation characterizes dynamic factor analysis as a state-space model, divided into an observed and an unobserved part. The unobserved part (latent trends, $x_{i,t}$), and its relationship with observations (factor loadings, $\mathbf{Z}_{xt}$) are estimated both with maximum likelihood or with Bayesian approaches to inference, like in BDFA.

In our analyses, we compared a BDFA, based on a Student's-t distribution of the process term of the random-walk model, to account for the skewed distribution of the data, with one based on a Gaussian distribution. Models were fitted with 5,000-10,000 MCMC iterations and a burn-in of 1,000-2,000 observations. We compared various models, with a different number of latent trends, to identify groups of time-series. Because time series had a similar variance, we opted for models with a fixed variance of the observed trends. Model selection was based on leave-one-out cross validation. Time-series data were centered to their mean, before model fitting.

Statistical analyses were carried out with the statistical software R [32], and BDFA were fitted with the package 'bdfa' [34], which estimated parameters through the software STAN [35].

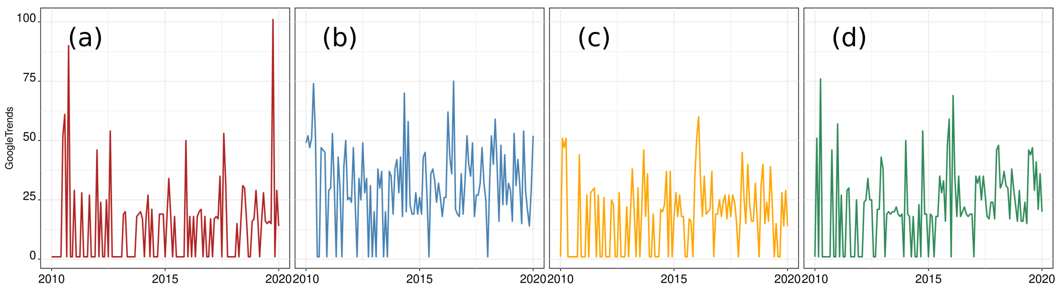

Figure 1: Google trends index for the four queries: "*Alien species*" (singular) (a), "*Alien species*" (plural) (b), "*Invasive species*" (singular) (c), "*Allochthonous species*" (singular) (d). The GoogleTrends index ranges between 0 and 100.

Results

Out of the 10 terms that we searched for, we were only able to download the GoogleTrends index for "*Alien species*" (singular), "*Alien species*" (plural), "*Invasive species*" (singular), "*Allochthonous species*" (singular) (Fig. 1). As the number of latent trends is equal to the number of observe trends, minus 1, we fit up to 3 latent trends for GoogleTrends data.

The monthly number of Wikipedia searches remained relatively stable, between 2015 and 2020, across the three pages that we queried: "*Alien species*" (Mean ± sd = 1657.5 ± 257.0), "*Invasive species in Italy*" (Mean ± sd = 196.7 ± 75.3), "*Invasive species in Europe*" (Mean ± sd = 863.9 ± 155.12; Fig. 2). We fit up to 2 latent trends for Wikipedia data.

https://doi.org/10.32942/osf.io/afrpk

Leave-one-out cross validation did not detect any difference between models with a different number of latent factors. Therefore, we retained a model with one latent factor, both for Google and Wikipedia trends. This highlighted that our time-series of Google searches and visits to Wikipedia reflected a common latent trend.

While the volume of the four Google searches decreased between 2010 and 2012, we noticed a moderate long-term increase in the latent factor scores and also a jump between 2014 and late 2015, when the EU regulation about IAS (n.1143/2014) was approved and entered into force. On the other hand, we did not detect any long-term increase in the latent factor scores of the Wikipedia model, which remained stable between 2015 and December 2019 (Fig. 3).

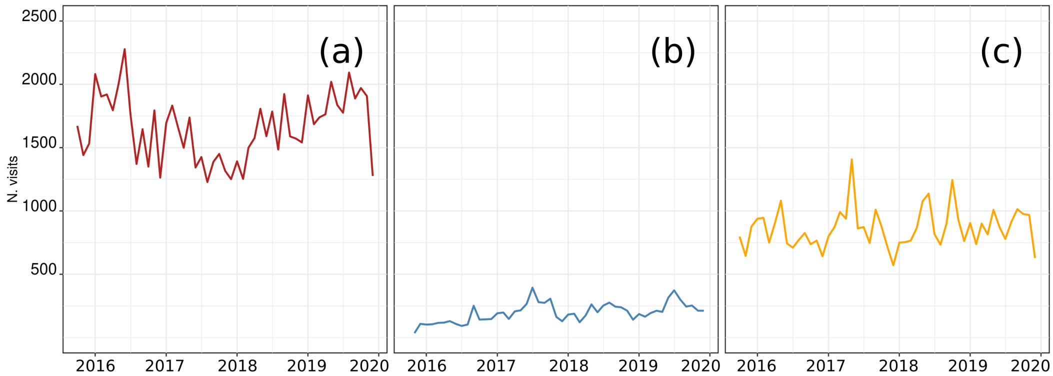

Figure 2: Monthly visits to the three Wikipedia pages: "*Alien species*" (a), "*Invasive species in Italy*" (b), "*Invasive species in Europe*" (c).

Discussion

Over the last decade, conservation culturomics blossomed, providing researchers with the chance to map the interest of billions of people towards environmental issues, by exploring the temporal evolution of their Internet searches. While many important conservation-related behaviors are still carried out by small segments of the society (e.g. poaching), and therefore cannot be detected by the analysis of Internet searches, we believe that culturomics can really answer many fundamental questions which still puzzle conservation biologists. For example, large-scale dynamics like the "extinction of experience"[36] or changes in wildlife value orientations[37], that usually require the implementation of cross-cultural surveys to be understood, can be explored across large spatial scales with minimum effort. The analysis of Internet searches can also be a barometer for public awareness about major drivers of change, like biological invasions, and so a tool for evaluating the environmental policies and their outreach.

However, to fully achieve these goals, we believe that culturomics studies should shift to a more holistic approach in how they analyze multiple time series. Multivariate time series analysis, like Bayesian dynamic factor analysis (BDFA) can represent a major advance in this direction. In facts, BDFA treats multiple observed time series as indicators reflecting one, or more, unobserved latent trend. This is a big advantage to explore the long-term interest about those environmental topics for which citizens can search across a wide range of keywords on Google and Wikipedia: rather than decomposing disjoint time-series, putting trends together and graphically interpreting their trajectories, BDFA can easily estimate the standardized scores of unobserved factors, representing the interest of the general public for a certain environmental topic, which subsequently made people searching for related queries on Google, generating GoogleTrends time-series. From Fig. 3 it is clear that standardized factor scores can provide a rapid overview of overall change, summarizing the temporal variation shared by multiple observed trends. Moreover, BDFA can highlight groups of queries with different temporal trends, with a higher degree of flexibility than longitudinal k-means clustering.

Concerning our case study, BDFA indicates that a model with a single latent factor was the best one at modelling Google searches about queries associated with IAS and biological invasions in Italy. Therefore, the progressive implementation of management projects targeting IAS, and their associated outreach, were probably effective at raising a general public interest in Italy, from which Google searches were generated. BDFA summarized this general interest into factor scores that steadily increased since 2012 (Fig. 3), in turn also increasing searches on Google (Fig. 1).

While the magnitude of this increase could seem modest, we have to keep in mind that the Google-

Trends metric is designed to deflate for those queries that people stop searching on Google. In a decade when Internet connection, and Google searches, almost doubled in Italy, our queries still increased their GoogleTrends index. The only way for this to happen is to have an increase in the number of searches about the queries, greater than that in the overall number of searches on Google. Thus, a growing number of people in Italy, between 2010 and 2020, searched for these 4 terms on Google, because of their increased interest, counteracting the deflation that the GoogleTrends would have naturally experimented, absent any increase in searches.

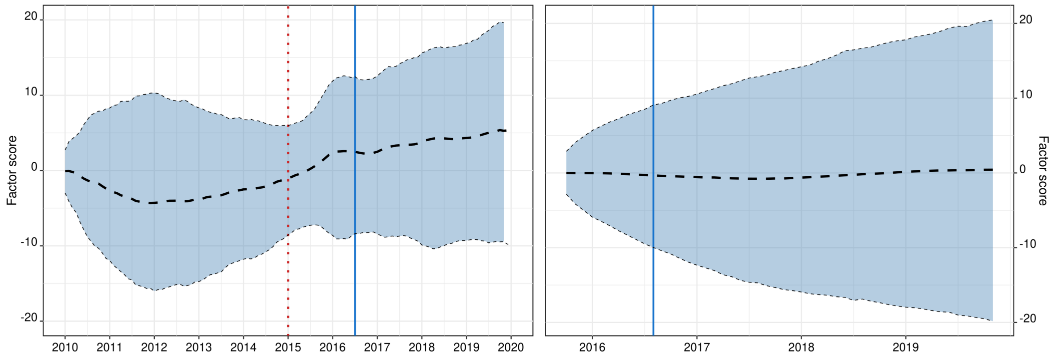

Figure 3: Long-term trend of factor scores from the BDFA models about Google and Wikipedia data. Data are centered, and should be interpreted as deviations from the mean of the two groups of time series. Solid vertical lines refer to January 2015, when the EU regulation (n.1143/2014) entered into force (dotted red line), and the publication of the first list of IAS of EU concern, in August 2016 (solid blue line).

On the other hand, despite we found that visits to Wikipedia pages covaried in time, we did not detect any long-term increase in their overall volume. This can indicate that, although more people searched about IAS and biological invasions on Google, they probably did not delve in Wikipedia for detailed information an used other web pages instead.

Despite no survey-based study measured detailed individual knowledge of IAS in Italy, or attitudes towards them, an independent survey[38] found that 67% of a representative sample from the Italian population were able to correctly identify the definition of IAS, over a set of different definitions. This information might be somehow complementary to our findings: on the one hand, after having heard about IAS, some people became interested towards them, on the other hand, people are able to attach meaning to this definition. Taken together, these two points corroborate the idea that IAS are penetrating the social debate in Italy, at least to a minimum extent. This can be an important precondition for the creation of stable attitudes towards IAS, and eventually their management, within some segments of the Italian society. Of course, this does not imply that managing IAS will become easier: people will probably still conflict because of their value orientations or their identity[37][39][40], and having more stable attitudes might further exacerbate these dynamics[41][42]. However, a greater interests towards IAS paves the way for citizen engagement[43][44] and, in general, a social debate about IAS might have a generative role, challenging the current top-down approach characterizing biological invasion policymaking and producing more shared and endorsed policies[41][45][46]. Given the importance of these changes for the long-term management of biological invasions, we absolutely encourage the replication of our approach to larger geographical scales and different countries, to see if similar patterns in online searches emerge. Moreover, we emphasize that it would be fundamental to complement the analysis of online searches with longitudinal surveys about attitudes and knowledge of IAS, from representative samples of the general population. Despite analyzing searching volumes can reveal temporal patterns in the social dimension of IAS, surveys are the only way to make rigorous inference about them and to better understand their occurrence within the various segments of the population[47].

Our findings also highlight that major policy initiatives can affect public interest towards biological invasions and IAS. The long-term trend of factor scores, and observed trends, from the model had a marked increase between 2014 and early 2016 (Fig. 1, 2). This jump occurred at a time when the European Commission was designing, approving, and implementing its first communication about the topic (EU regulation, n. 1143/2014). Probably this process attained media resonance, which boosted the interest of professionals, researchers, the media, and citizens about the topic. This in turn affected on-line searching behavior on Google. We acknowledge that including information regarding media coverage about policymaking initiatives can be extremely important to appreciate their importance. Unfortunately, no structured dataset was available for Italy at the time of the study, but culturomics studies carried out over larger spatial scales and in English-speaking countries can extract news from on-line archives like GDELT (https://www.gdeltproject.org/). BDFA can then be a valuable framework also to test for their effect over on-line searching behavior, because news volume can be included as a time-varying covariate in the model.

To the best of our knowledge this research was the first adopting multivariate time series analysis to explicitly treat time series about searches on Google and Wikipedia as indicators of a latent, unobserved, interest towards biological invasions and invasive alien species. We believe that conservation culturomics can greatly benefit from the adoption of this indicator-trend approach, which would enable researchers to really understand whether multiple volumes of Internet searches underlie a common change in public interest or not. This knowledge will be of the uttermost importance to answer questions about long-term changes in public attitudes and awareness about biodiversity and its drivers of pressure and change. Moreover, it will also enable practitioners and policymakers to assess the long-term effect of conservation outreach and communication campaigns.

N. factors	Process error distribution	ELPD
GoogleTrends		
1 factor	Gaussian	-2033.5 ± 24.1
1 factor	Student’s t	-2033.4 ± 24.1
2 factors	Gaussian	-2034.0 ± 24.0
2 factors	Student’s t	-2034.1 ± 24.0
3 factors	Gaussian	-2034.6 ± 24.1
3 factors	Student’s t	-2034.3 ± 24.1
Wikipedia visits		
1 factor	Gaussian	-991.8 ± 11.2
1 factor	Student’s t	-991.8 ± 11.2
2 factors	Gaussian	-991.9 ± 11.2
2 factors	Student’s t	-991.9 ± 11.2

Table 1: Model comparison based on leave-one-out cross validation: number of latent factors of each model, type of distribution of the process error term of the random-walk model and expected log-pointwise predictive density (ELPD).

Acknowledgments

This research was funded by a contribution from Legambiente to JC under Action C3 of the LIFE15 GIE/IT/001039 ASAP Project.

References

1. Troumbis, A. Y., & Iosifidis, S. (2020). A decade of Google Trends-based Conservation culturomics research: A critical evaluation of an evolving epistemology. *Biological Conservation*, 248, 108647. https://doi.org/10.1016/j.biocon.2020.108647
2. Correia, R. A., Di Minin, E., Jarić, I., Jepson, P., Ladle, R., Mittermeier, J., Roll, U., Soriano-Rendondo, A., & Veríssimo, D. (2019). Inferring public interest from search engine data requires caution. *Frontiers in Ecology and the Environment*, 17(5). https://doi.org/10.1002/fee.2048
3. Mellon, J. (2014). Internet search data and issue salience: The properties of Google Trends as a measure of issue salience. *Journal of Elections, Public Opinion Parties*, 24(1), 45-72. https://doi.org/10.1080/17457289.2013.846346
4. Ripberger, J. T. (2011). Capturing curiosity: Using internet search trends to measure public attentiveness. *Policy studies journal*, 39(2), 239-259. https://doi.org/10.1111/j.1541-0072.2011.00406.x
5. Correia, R. A., Ladle, R., Jarić, I., Malhado, A. C., Mittermeier, J. C., Roll, U., Sorian-Redondo, A., Veríssimo, D., Fink, C., Hausmann, A., Guedes-Santos, J., Vardi, R., & Di Minin, E. (2021). Digital data sources and methods for conservation culturomics. *Conservation Biology*, 35(2), 398-411. https://doi.org/10.1111/cobi.13706
6. Gizzi, F. T., Kam, J., & Porrini, D. (2020). Time windows of opportunities to fight earthquake under-insurance: evidence from Google Trends. *Humanities and Social Sciences Communications*, 7(1), 1-11 https://doi.org/10.1057/s41599-020-0532-2
7. Russo, G. I., di Mauro, M., Cocci, A., Cacciamani, G., Cimino, S., Serefoglu, E. C., Albersen, M., Capogrosso, P., Fode, M., & Verze, P. (2020). Consulting “Dr Google” for sexual dysfunction: a contemporary worldwide trend analysis. *International journal of impotence research*, 32(4), 455-461. https://doi.org/10.1038/s41443-019-0203-2

8. Wanner, P. (2020). How well can we estimate immigration trends using Google data?. *Quality Quantity*, 1-22. https://doi.org/10.1007/s11135-020-01047-w

9. Tizzoni, M., Panisson, A., Paolotti, D., & Cattuto, C. (2020). The impact of news exposure on collective attention in the United States during the 2016 Zika epidemic. *PLoS computational biology*, 16(3), e1007633 https://doi.org/10.1371/journal.pcbi.1007633

10. Mittermeier, J. C., Roll, U., Matthews, T. J., & Grenyer, R. (2019). A season for all things: Phenological imprints in Wikipedia usage and their relevance to conservation. *PLoS biology*, 17(3), e3000146. https://doi.org/10.1371/journal.pbio.3000146

11. Fernández-Bellon, D., & Kane, A. (2020). Natural history films raise species awareness—A big data approach. *Conservation letters*, 13(1), e12678. https://doi.org/10.1111/conl.12678

12. Heberlein, T. A. (2012). Navigating environmental attitudes. *Oxford University Press*, Oxford. https://doi.org/10.1093/acprof:oso/9780199773329.001.0001

13. Millard, J. W., Gregory, R. D., Jones, K. E., & Freeman, R. (2021). The species awareness index as a conservation culturomics metric for public biodiversity awareness. *Conservation Biology*. https://doi.org/10.1111/cobi.13701

14. Cerri, J., Mori, E., Ancillotto, L., Russo, D., & Bertolino, S. COVID-19, media coverage of bats and related Web searches: a turning point for bat conservation?. *Mammal Review*. https://doi.org/10.1111/mam.12261

15. Vijay, V., Field, C. R., Gollnow, F., & Jones, K. K. (2021). Using internet search data to understand information seeking behavior for health and conservation topics during the COVID-19 pandemic. *Biological Conservation*, 109078. https://doi.org/10.1016/j.biocon.2021.109078

16. Streicker, D. G. *et al.* (2012). Ecological and anthropogenic drivers of rabies exposure in vampire bats: implications for transmission and control. Proceedings of the *Royal Society B: Biological Sciences*, 279(1742), 3384-3392. https://doi.org/10.1098/rspb.2012.0538

17. Jarić, I., Bellard, C., Correia, R.A., Courchamp, F., Douda, K., Essl, F., Jeschke, J.M., Kalinkat, G., Kalous, L., Lennox, R.J., Novoa, A., Proulx, R., Pyšek, P., Soriano–Redondo, A., Souza, A.T., Vardi, R., Veríssimo, D., & Roll, U. (2021). Invasion Culturomics and iEcology. *Conservation Biology*. https://doi.org/10.1111/cobi.13707

18. Mollot, G., Pantel, J. H., & Romanuk, T. N. (2017). The effects of invasive species on the decline in species richness: a global meta-analysis. *Advances in ecological research*, 56, 61-83. https://doi.org/10.1016/bs.aecr.2016.10.002

19. Simberloff, D. (2013). Invasive species: what everyone needs to know. *Oxford University Press*, Oxford. https://global.oup.com/academic/product/invasive-species-9780199922017?q=simberloff&lang=en&cc=it

20. Turbelin, A. J., Malamud, B. D., & Francis, R. A. (2017). Mapping the global state of invasive alien species: patterns of invasion and policy responses. *Global Ecology and Biogeography*, 26(1), 78-92. https://doi.org/10.1111/geb.12517

21. Cramer, W., Guiot, J., Fader, M., Garrabou, J., Gattuso, J.P., Iglesias, A., Lange, M.A., Lionello, P., Llasat, M.C., Paz, S., Peñuelas, J., Snoussi, M., Toreti, A., Tsimplis, M.N., & Xoplaki, E. (2018). Climate change and interconnected risks to sustainable development in the Mediterranean.. *Nature Climate Change*, 8(11), 972-980. https://doi.org/10.1038/s41558-018-0299-2

22. Tsiamis, K., Gervasini, E., Deriu, I., D'amico, F., Nunes, A., Addamo, A., & Cardoso, A. C. (2017). Baseline distribution of invasive alien species of Union concern. Ispra (Italy): *Publications Office of the European Union*, 1-96. https://core.ac.uk/download/pdf/84886646.pdf

23. Loy, A., Aloise, G., Ancillotto, L., Angelici, F.M., Bertolino, S., Capizzi, D., Castiglia, R., Colangelo, P., Contoli, L., Cozzi, B., Fontaneto, D., Lapini, L., Maio, N., Monaco, A., Mori, E., Nappi, A., Podestà, M., Russo, D., Sarà, M., Scandura, M., & Amori, G. (2019). Mammals of Italy: an annotated checklist. *Hystrix*, 30, 87–106. https://doi.org/10.4404/hystrix-00196-2019

24. Gherardi, F., Bertolino, S., Bodon, M., Casellato, S., Cianfanelli, S., Ferraguti, M., Lori, E., Mura, G., Nocita, A., Riccardi, N., Rossetti, G., Rota, E., Scalera, R., Zerunian, S., & Tricarico, E. (2008). Animal xenodiversity in Italian inland waters: distribution, modes of arrival, and pathways. *Biological Invasions*, 10(4), 435-454. https://doi.org/10.1007/s10530-007-9142-9

25. Nocita, A., Tricarico, E., & Bertolino, S. (2017). Fine-scale analysis of heavily invaded Italian freshwater fish assemblages. *Integrative zoology*, 12(6), 500-511. https://doi.org/10.1111/1749-4877.12267

26. D'Amen, M., & Azzurro, E. (2020). Lessepsian fish invasion in Mediterranean marine protected areas: a risk assessment under climate change scenarios. *ICES Journal of Marine Science*, 77(1), 388-397. https://doi.org/10.1093/icesjms/fsz207

27. Bertolino, S., & Genovesi, P. (2003). Spread and attempted eradication of the grey squirrel (*Sciurus carolinensis*) in Italy, and consequences for the red squirrel (*Sciurus vulgaris*) in Eurasia. *Biological Conservation*, 109(3), 351-358. https://doi.org/10.1016/S0006-3207(02)00161-1

28. Genovesi, P., Carboneras, C., Vila, M., & Walton, P. (2015). EU adopts innovative legislation on invasive species: a step towards a global response to biological invasions?. *Biological Invasions*, 17(5), 1307-1311. https://doi.org/10.1007/s10530-014-0817-8

29. Tollington, S., Turbé, A., Rabitsch, W., Groombridge, J. J., Scalera, R., Essl, F., & Shwartz, A. (2017). Making the EU legislation on invasive species a conservation success. *Conservation Letters*, 10(1), 112-120. https://doi.org/10.1111/conl.12214

30. Roy, H. E., Bacher, S., Essl, F., Adriaens, T., Aldridge, D. C., Bishop, J. D., Blackburn, T. M., Branquart, E., Brodie, J., Carboneras, C., Cottier-Cook, E. J., Copp, G. H., Dean, H. J., Eilenberg, J., Gallardo, B., Garcia, M., García-Berthou, E., Genovesi, P., Hulme, P. E., Kenis, M., Kerckhof, F., Kettunen, M., Minchin, D., Nentwig, W., Nieto, A., Pergl, J., Pescott, O. L., Peyton, J. M., Preda, C., Roques, A., Rorke, S. L., Scalera, R., Schindler, S., Schönrogge, K., Sewell, J., Solarz, W., Stewart, A. J. A., Tricarico, E., Vanderhoeven, S., van der Velde, G., Vilà, M., Wood, C. A., Zenetos, A. & Rabitsch, W. (2019). Developing a list of invasive alien species likely to threaten biodiversity and ecosystems in the European Union. *Global Change Biology*, 25(3), 1032-1048. https://doi.org/10.1111/gcb.14527

31. Bologna, E., Fornari, R., Zannella, L., Matarazzo, G., & Dolente, C. (2018). Internet@ Italia 2018: Domanda e offerta di servizi online e scenari di digitalizzazione. https://www.istat.it/it/files/2018/06/Internet@Italia-2018.Pdf

32. Eurostat. (2019). Digital economy and society statistics – households and individuals. https://ec.europa.eu/eurostat/statistics-explained/index.php/Digital_economy_and_society_statistics_-_households_and_individuals#Internet_access

33. R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

34. Dorner, B., Catalano, M. J., & Peterman, R. M. (2018). Spatial and temporal patterns of covariation in productivity of Chinook salmon populations of the northeastern Pacific Ocean. *Canadian Journal of Fisheries and Aquatic Sciences*, 75(7), 1082-1095. https://doi.org/10.1139/cjfas-2017-0197

35. Ward, E. J., Anderson, S. C., Daminano, L. A., Hunsicker, M. E., & Litzow, M. A. (2019). Modeling regimes with extremes: the bayesdfa package for identifying and forecasting common trends and anomalies in multivariate time-series data. *The R Journal*, 11(2), 46-55. https://journal.r-project.org/archive/2019-2/

36. Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M. A., Guo, J. I., Li, P., & Riddell, A. (2017). Stan: A Probabilistic Programming Language. *Journal of Statistical Software*, 76(1), 1. http://dx.doi.org/10.18637/jss.v076.i01

37. Soga, M., & Gaston, K. J. (2016). Extinction of experience: the loss of human–nature interactions. *Frontiers in Ecology and the Environment*, 14(2), 94-101. https://doi.org/10.1002/fee.1225

38. Manfredo, M. J., Teel, T. L., Don Carlos, A. W., Sullivan, L., Bright, A. D., Dietsch, A. M., Bruskotter, J., & Fulton, D. (2020). The changing sociocultural context of wildlife conservation. *Conservation Biology*, 34(6), 1549-1559. https://doi.org/10.1111/cobi.13493

39. Carnevali, L., Alonzi, A., & Piazzi, A. (2018). LIFE ASAP (LIFE15GIE/IT/001039). Azione C.1. Report ex-ante indagine sul grande pubblico. https://www.lifeasap.eu/images/prodotti/6.1.14.1%20Survey%20on%20public%20awareness%20on%20IAS-%20ex%20ante%20Report.pdf

40. Bruskotter, J. T., Vucetich, J. A., Dietsch, A., Slagle, K. M., Brooks, J. S., & Nelson, M. P. (2019). Conservationists' moral obligations toward wildlife: values and identity promote conservation conflict. *Biological Conservation*, 240, 108296. https://doi.org/10.1016/j.biocon.2019.108296

41. van Eeden, L. M., Newsome, T. M., Crowther, M. S., Dickman, C. R., & Bruskotter, J. (2019). Social identity shapes support for management of wildlife and pests. *Biological conservation*, 231, 167-173. https://doi.org/10.1016/j.biocon.2019.01.012

42. Crowley, S. L., Hinchliffe, S., & McDonald, R. A. (2017). Conflict in invasive species management. *Frontiers in Ecology and the Environment*, 15(3), 133-141. https://doi.org/10.1002/fee.1471

43. Lioy, S., Marsan, A., Balduzzi, A., Wauters, L. A., Martinoli, A., & Bertolino, S. (2019). The management of the introduced grey squirrel seen through the eyes of the media. *Biological Invasions*, 21(12), 3723-3733. https://doi.org/10.1007/s10530-019-02084-9

44. Hester, S. M., & Cacho, O. J. (2017). The contribution of passive surveillance to invasive species management. *Biological Invasions*, 19(3), 737-748. https://doi.org/0.1007/s10530-016-1362-4

45. Bryce, R., Oliver, M. K., Davies, L., Gray, H., Urquhart, J., & Lambin, X. (2011). Turning back the tide of American mink invasion at an unprecedented scale through community participation and adaptive management. *Biological conservation*, 144(1), 575-583. https://doi.org/10.1016/j.biocon.2010.10.013

46. Courchamp, F., Fournier, A., Bellard, C., Bertelsmeier, C., Bonnaud, E., Jeschke, J. M., & Russell, J. C. (2017). Invasion biology: specific problems and possible solutions. *Trends in ecology evolution*, 32(1), 13-22. https://doi.org/10.1016/j.tree.2016.11.0017

47. Shackleton, R.T., Richardson, D.M., Shackleton, C.M., Bennett, B., Crowley, S.L., Dehnen-Schmutz, K., Estévez, R.A., Fischer, A., Kueffer, C., Kull, A.C., Elizabete, M., Novoa, A., Potgieter, L.J., Vaz, A.S., & Larson, B.M.H. (2019). Explaining people's perceptions of invasive alien species: a conceptual framework. *Journal of Environmental Management*, 229, 10-26. https://doi.org/10.1016/j.jenvman.2018.04.045

48. Callegaro, M., & Yang, Y. (2018). The role of surveys in the era of "Big Data". In: Vannette, D.L., Krosnick, J.A. (Eds.) *The Palgrave handbook of survey research*. Palgrave MacMillan, pp. 175-192. https://www.palgrave.com/gp/book/9783319543949

Supplementary Information

The data, altogether with the reproducible software code, are available on the Open Science Framework repository, at the following link: https://osf.io/ujkfz/