THE EUROPEAN
PHYSICAL JOURNAL
SPECIAL TOPICS

Regular Article

Global Earthquake Forecasting System (GEFS): The challenges ahead

A. Mignan[1,2,a], G. Ouillon[3,4], D. Sornette[1,4,5], and F. Freund[6]

[1] Institute of Risk Analysis, Prediction and Management, Academy for Advanced Interdisciplinary Studies, Southern University of Science and Technology (SUSTech), Shenzhen, P.R. China
[2] Department of Earth and Space Sciences, Southern University of Science and Technology (SUSTech), Shenzhen, P.R. China
[3] Lithophyse, Nice, France
[4] ETH Zurich, Department of Management, Technology and Economics (D-MTEC), Zurich, Switzerland
[5] Institute of Innovative Research, Tokyo Institute of Technology, Tokyo, Japan
[6] GeoCosmo Science Center, NASA Ames Research Park, Moffett Field, USA

Abstract. We conclude this special issue on the Global Earthquake Forecasting System (GEFS) by briefly reviewing and analyzing the claims of non-seismic precursors made in the present volume, and by reflecting on the current limitations and future directions to take. We find that most studies presented in this special volume, taken individually, do not provide strong enough evidence of non-seismic precursors to large earthquakes. The majority of the presented results are hampered by the fact that the task at hand is susceptible to potential biases in data selection and possible overfitting. The most encouraging results are obtained for ground-based geoelectric signals, although the probability gain is likely small compared to an earthquake clustering baseline. The only systematic search on satellite data available so far, those of the DEMETER mission, did not find a robust precursory pattern. The conclusion that we can draw is that the overall absence of convincing evidence is likely due to a deficit in systematically applying robust statistical methods and in integrating scientific knowledge of different fields. Most authors are specialists of their field while the study of earthquake precursors requires a system approach combined with the knowledge of many specific characteristics of seismicity. Relating non-seismic precursors to earthquakes remains a challenging multidisciplinary field of investigation. The plausibility of these precursors predicted by models of lithosphere-atmosphere-ionosphere coupling, together with the suggestive evidence collected here, call for further investigations. The primary goal of the GEFS is thus to build a global database of candidate signals, which could potentially improve earthquake predictability (if the weak signals observed are real and false positives sufficiently uncorrelated between different data sources). Such a stacking of disparate and voluminous data will require big data storage and machine learning pipelines, which has become feasible only

recently. This special issue compiled an eclectic list of non-seismic precursor candidates, which is in itself a valuable source of information for seismologists, geophysicists and other scientists who may not be familiar with such types of investigations. It also forms the foundation for a coherent, multi-disciplinary collaboration on earthquake prediction.

1 Introduction

Earthquake prediction has a long history of ups and downs, any past phase of renewed optimism [1–5] having been tempered by countervailing arguments pointing at a lack of generalization, at conflicting results, and/or at disputed uses of statistics [6–12].

The list of prior attempts to address earthquake precursors is long and includes, in chronological order, special issues:

- Earthquake prediction: an international review [13],
- Earthquake prediction [14],
- Evaluation of proposed earthquake precursors [15],
- Earthquake prediction: the scientific challenge [16],
- What comes next in the dynamics of lithosphere and earthquake prediction? [17],
- Is the reliable prediction of individual earthquakes a realistic scientific goal? [18],
- The 2004 Parkfield earthquake and the Parkfield earthquake prediction experiment [19],
- Critical point theory and space-time pattern formation in precursory seismicity [20],
- Overview of the Working Group for the development of Regional Earthquake Likelihood Models (RELM) [21],
- Collaboratory for the Study of Earthquake Predictability (CSEP): Italy [22],
- Overview of the first earthquake forecast testing experiment in Japan [23] (i.e., CSEP Japan),
- Biological anomalies prior to earthquakes [24],
- Current challenges in statistical seismology [25],
- Collaboratory for the Study of Earthquake Predictability (CSEP): New results and future directions [26];

Plus, thematic reviews on specific precursors, precursory processes and methods, such as:

- Chinese earthquake prediction efforts [27–32],
- Various seismic precursors [33,34],
- Unusual animal behavior [35],
- Precursory seismic quiescence [36],
- Precursory accelerating seismicity and short-term foreshocks [37–39],
- Mechanical processes [40–42],
- Ionospheric precursors [43],
- Various precursors including many non-seismic [44,45],
- Operational earthquake forecasting [46],
- Neural network applications in earthquake prediction [47].

By now, all low-hanging fruits, namely the obvious earthquake precursor candidates, should have already been identified. This means that earthquake prediction, if physically feasible, will require considerable efforts to advance from simple (but still challenging) forecasting of aftershocks [48] to the forecasting of more spontaneous events

ᵃ e-mail: mignana@sustech.edu.cn

that nearly always take civil protection agencies and populations by surprise. Notable exceptions were the 1975 Haicheng earthquake and some other major seismic events in China between 1966 and 1976 [32,49]. The Global Earthquake Forecasting System (GEFS) discussed here addresses the challenge to improve on this large existing knowledge base.

The GEFS special issue is reminiscent of the 1991 evaluation of proposed earthquake precursors [15], but it is also a reboot in the new era of big data analytics and machine learning and is powered by an improved understanding of the underlying physical processes. Despite the limits of each of the precursors proposed in this special issue (Sects. 2 and 3 below), it is commendable to revitalize research on non-seismic precursor candidates in view of the seismic-precursor centric view that currently dominates the research landscape. Importantly, the GEFS is initiating a new and much needed phase of cross-disciplinary collaboration [8,50]. The signals addressed in the GEFS issue include: ground-based geoelectric and magnetic anomalies [51–55], radon activity [56], thermal infrared anomalies [57–60] and longwave radiation [57], combined thermal and gas anomalies [61], (sub)ionospheric signals [53,62,63], total electron content [57], groundwater changes [56,64], precipitation [65], low-frequency "stress waves" [66], solar activity coupling [67], as well as combinations of various signals due to lithosphere-atmosphere-ionosphere coupling [56,57] and comparisons of non-seismic signals to seismic precursors [54,68]. The introductory article of this volume [69] provides a review of the solid-state processes that are proposed to form the basis of the non-seismic precursors and others (in the same volume [70]). While seismic precursors per se have not been considered in this special issue except for their comparison with non-seismic precursors by [54] and an earthquake clustering null hypothesis [68,71], they should of course be a crucial part of any all-encompassing strategy in GEFS. The next step, hints of which are given in the preface to the special issue [72], will be to assemble all types of signals – geodesic, geoelectric, geomagnetic, seismic, etc. – into a unified workflow. This will require big data storage and management capabilities as well as machine learning savviness and acumen combined with a deep physical understanding of the underlying mechanisms and their observable consequences (Sect. 4).

From a practical point of view, given the challenges in data acquisition, analysis, and modelling, as well as in organization and operations, it is unlikely that predictions as successful as the one of the 1975 Haichang earthquake can be achieved in the very near future. For now, the best mitigation measures remain improved seismic design (including retrofitting) of buildings as well as improvements of the early warning systems [73]. The aim of earthquake prediction research should be to improve, in the short-to-medium term, the science of earthquakes, while safekeeping the long-term motivation to protect property and save lives. Keilis-Borok [34] considered earthquake prediction the "*current frontier of the solid Earth sciences*" that is often rife with disappointments. Quoting Jackson [10], let us not forget that "*false optimism, unrealized hopes and even outright failures are no reason to surrender. Every science has its failures, and many seemingly impossible challenges have been overcome*". If past failures were the justification to abandon attempts to crack a problem or discover new things, science would not progress at all.

While a number of theoretical frameworks exist to explain different types of seismic precursors, from bottom-up holism [34,74,75] to top-down reductionism [76–78], it is theories that connect crustal deformation to gas release, such as in the Lithosphere-Atmosphere-Ionosphere Coupling (LAIC) model [79] or in the Peroxy Defects model [69], which are so far the most promising approaches to explain a great variety of non-seismic precursors, from electromagnetic signals to animal behavior. This special issue, for most articles, revolves around this wide range of topics.

2 Critical review of proposed non-seismic precursors

This section provides a short summary and review for each of the 20 papers that compose this volume (numbered (2) to (21) in Section 2.2 – (1) being the introductory article [69]). We exert a critical analysis that emphasizes the importance of starting with the null hypothesis that earthquakes are not predictable, which counterbalances the relative optimism transpiring from the majority of the studies included here. Thus, before we start with individual reviews per se, we shall first recall some of the basic rules that constitute what is considered "good practice" in the field of statistical testing. This should provide the statistical foundation on which the "seismic" and "non-seismic" communities should agree when assessing the performance of their claimed precursors and models. We hope that these guidelines will also help non-statistician readers to better evaluate the quality of the papers published, as well as our own evaluation.

2.1 Guidelines for rigorous analysis of earthquake precursors

First of all, any empirical study should as a matter of fact provide sufficient details about the data that are used. Such descriptions should include: the time range and spatial zone covered, the type of signal, its frequency content, the sampling rates in space and time – as well as a discussion about the accuracy, uncertainties and other possible limitations of the raw data. In case where preprocessing was necessary, the output uncertainties should also be discussed.

When using earthquake data, one should also select a well-defined spatial zone (latitude, longitude, depth range) and time interval, as well as the minimum magnitude above which all events are selected systematically [80]. Just picking some specific events and not others without justification risks disqualifying any serious scientific conclusion. In that respect, while so-called case studies are useful in a first approach to collect examples that can help formulate hypotheses, they are not useful for a systematic statistical assessment of their existence and reliability. As it may often be advisable to deal only with a subset of independent events (in order to avoid biased statistics when comparing a geophysical signal with seismicity), one may disregard aftershocks (but this is not always necessary – see below). However, since the so-called foreshocks, mainshocks and aftershocks may be physically undiscernible [81,82], declustering becomes a non-deterministic process. We thus advise using several techniques of declustering in order to check the dependence of the final results on the details of that delicate empirical first step [83]. In case of a purely stochastic declustering scheme, where each event is assigned a probability of independence and a set of probabilities of having been triggered by predecessors [84], the best route is to adopt a Monte Carlo sampling strategy, thus to use many possible realizations of a catalog of independent events.

All methods of data processing should be detailed or referenced very clearly, so that it should be possible for any reader to reproduce all results without any ambiguity. The detailed algorithm has to be provided, and one should spell out what are the various parameters that have to be set manually, and which ones have not been specified for some physical reason. If certain parameters have to be set arbitrarily (e.g. thresholds or space and time windows), it is mandatory to check the sensitivity of the output results by varying these parameters, just as for the declustering step mentioned above.

Prediction and predictability experiments should also be conducted properly, and one should test candidate precursory anomalies as objectively as possible; thus, detection of anomalies should not depend on the prior knowledge of the earthquake catalog.

The performance of a proposed prediction method can be assessed using a variety of tools, like featuring the ratios of true positive, true negative, false positive and false negative predictions, which are inputs of the confusion (or error) matrix.

The use of Molchan diagrams [85] is also recommended, where the horizontal axis stands for the proportion of space-time covered by alarms, and the vertical axis for the ratio of events missed by those alarms. Notice that this definition of Molchan diagrams assumes as a null hypothesis that earthquakes occur uniformly and randomly in space and time, unrelated to the alarms. In cases where the null hypothesis is more complex, for instance with respect to the clustering of earthquakes, the horizontal axis stands for the prior probability of occurrence of the events in the chosen alarmed space-time domain. Using such a systematic approach, one can determine which sets of parameters allow for the best correlations between the reported anomalies and earthquakes. In case where the data come in sufficiently large volume, the best practice consists in training (i.e. optimizing) the model on some part of it, and then in selecting the best model. Thus, this model is applied to an independent validation dataset to check its score in the error matrix or the Molchan diagram.

When one cannot define a proper validation dataset (because, for instance, it would feature less data), the best solution is to compare the earthquakes/anomalies correlations with results obtained using surrogate earthquake catalogs: the latter consist in various randomizations of the original dataset. For instance, one may keep the location of all earthquakes the same, but randomly switch their occurrence times (and/or magnitudes if possible or useful). This process will destroy the joint space/time/magnitude earthquake structures, while keeping each of the marginal ones separately. Another possibility consists in generating purely random locations, occurrence times, and magnitudes (the latter still respecting the Gutenberg-Richter law). The geophysical signal remains untouched, so that the set of anomalies is still the same, but their correlation with surrogate seismicity should be removed. The model is then optimized on each random realization of the earthquake dataset, and the collection of best scores can be compared to the one obtained using the true catalog. Such tests have to be performed over a large number of different surrogate catalogs. The model is validated if the score obtained on the true catalog is significantly larger than the scores obtained using random datasets. In order to test more thoroughly for false negatives, one should define a sub-dataset revealing the behavior of the signal when there is no seismicity at all.

2.2 Review and critique of each contribution

(2) Freund et al. [55] offer insight from a physics perspective into some fundamental solid-state processes, by which stresses acting on rocks activate electronic charge carriers (specifically defect electrons, also known as holes or "positive holes") that can produce potentially powerful electric currents in the Earth's crust, where none existed without stress. These currents are thought to be responsible for bursts of unipolar electromagnetic pulses generated near hypocenters of impending seismic events. Basing their approach on the peroxy theory, they propose a model in which pressure-induced volume oscillations occur due to peroxy bond formation and dissociation associated with volume changes. This process is expected to be accompanied by the emission of (*i*) electromagnetic (EM) waves in the form of unipolar pulses and (*ii*) singular seismic compressional (P) waves. The authors claim that it should be possible to triangulate the location of the pulses using more than one station and point to the successful triangulation of unipolar pulses emitted from the subduction zone of the Nazca Plate off the coast of Lima, Peru, using stations onshore and on the San Lorenzo Island. No forecast is provided in that study; the theory should

be tested and validated by a systematic survey using multiple sets of antennas and seismometers to record the EM and expected P waves from unipolar pulse events together with their delay times due to the different speeds of propagation through the rock column.

(3) Chen et al. [51] offer a generalization of the Burridge-Knopoff block-spring toy model of earthquake dynamics that encompasses the mechano-electric coupling described in [69], providing a coarse-grained representation of the coupled earthquake-fault-electric precursors. Specifically, Chen et al.'s model uses a set of masses driven by and interacting through springs, sliding on an interface with Mohr-Coulomb friction. Each mass is modelled as an RLC circuit featuring a generator whose voltage depends on the local stress to mimic electric charges production predicted by the peroxy defects theory. Each time a block slides because it reached its frictional threshold, it can load adjacent blocks and trigger their slip too. Charges are allowed to flow from block to block, and stress fluctuations thus induce macroscopic electric current fluctuations at the scale of the whole fault. The model reproduces some features observed in laboratory experiments, and also some statistics of voltage fluctuations observed in the field. Notice that, in this model, electric charges do not influence the state of stress, which is very different from what is proposed by [69]. Any potential predictability gain offered by the model has yet to be tested.

(4) Scoville and Freund [59] present various scenarios commonly proposed in the literature to account for and explain thermal infrared (TIR) anomalies, but they call out these scenarios as unlikely. Instead they present/document a peroxy-defect based model, noticing that TIR anomalies are generally short-lived and most often observed coming from mountain tops. They stress that, contrarily to what many authors claim, the TIR anomalies do not stem from changes in the genuine body temperature, but rather from excess infrared radiation at the same wavelength range as the more common temperature effects. Thus, even if expressed in degrees, earthquake-related TIR anomalies do not correspond to actual temperature changes, but to the emission of IR photons from peroxy bonds that are forming at the Earth's surface. The authors then present a series of laboratory experiments, subjecting rocks of various sizes to localized stress and recording the TIR emission intensity and spectra from an unstressed part of the rocks as well as their variation with time. They show an association of spectral TIR emission and loading applying a new fluctuation spectroscopy approach. In particular, they show that the spectrum switches from white noise to pink noise when loading begins until the mechanical failure of the rock. Very sharp spectral bands are observed, which allows the authors to invert for the dissociation energy of peroxy bonds as previously observed or theoretically predicted. According to these authors, this seals the fate of the alternate theory that the anomalous TIR emissions are due to the emanation of warm gases from the Earth's ground. These experiments implicitly assume that laboratory observations extrapolate to the scale of crustal processes and the realm of large earthquakes, which remains to be fully verified.

(5) Piroddi [58] describes how precursory signals of seismic activity can be identified in thermal infrared data from geostationary satellites. Methodological in nature, the article presents the Night Thermal Gradient (NTG) algorithm, which reduces signal artefacts and produces more detailed maps, hence improving the recognition of potential pre-seismic anomalies. Results are presented for Central Italy before the 2009 L'Aquila earthquake and for an aseismic region in Sardinia. The generated NTG maps suggest excess TIR emission from the mountains on both sides of the L'Aquila valley but not from the valley itself, where the main shock occurred. This observation is in agreement with the fact argued by [59] that the electronic charge carriers responsible for the IR emission tend to accumulate at topographic highs. While other reported TIR anomalies mentioned by the author are related to several similar events,

the lack of systematic analysis and statistical testing suggests a potential selection bias. Even if those TIR patterns were associated to small seismic events, a large anomaly observed in 2008 is considered a false alarm but is suggested as potentially linked to a statistical change in minor seismic activity by the author. Taking only the L'Aquila earthquake as target event, this would indicate a 50% accuracy (this is not necessarily meaning a random skill as the selection of location enters into the relevance of the forecast). Interestingly, a long NTG anomaly lineament matches the location of the L'Aquila aftershock sequence. This may indicate some post-mainshock anomalies, although it may be random or correlated to the topography, as investigated by the author.

(6) Zhang et al. [60] reproduce the results of a previously published study claiming a close association between TIR anomalies and earthquakes in Greece [86]. However, using a Molchan diagram taking account of the prior spatial distribution of earthquakes, the systematic study of Zhang et al. shows that the proposed TIR-based precursory alarms do not perform better than a random choice. This leads them to claim that apparent high scores of predictability are indeed due to oversized space-time alarm windows. It is also important to stress that the use of a "naïve" Molchan diagram, without weighting it by its prior probabilities, wrongly indicates that the prediction was successful. This over-optimistic conclusion was indeed due to the spatial clustering of earthquakes, which distorted the obtained scores. To be clear, the Molchan diagram does not apply directly when seismicity is not homogenous and suitable adaptation needs to be applied (Sect. 2.1). These results thus call for a more systematic testing of various space-time alarm windows in order to estimate their optimal size, if any.

(7) Potirakis et al. [63] review the literature of potential electromagnetic precursors to large earthquakes (fracto-electromagnetic emissions, ultra-low frequency ULF, and sub-ionospheric VLF) and provide a survey of the methods employed as well as of the underlying theoretical frameworks. Results are given in the form of study cases, which may be susceptible to some selection bias. The analysis however provides a good overview of this field of research.

(8) Yan et al. [53] compare DC-ULF electric field data recorded by ground stations and at the low-altitude DEMETER satellite as a sanity check for future earthquake precursor analyses. The authors searched for signal synchronicity for all of China and found similar annual variation trends of both ionospheric and geoelectric fields, in the same directions and frequency. This method should help increase the robustness of the processed multi-data signals to be considered in the GEFS to evidence anomalies possibly related to seismic activity and better understand the lithosphere-atmosphere-ionosphere coupling mechanism. Such study is complementary to forecasting analyses, such as done by [62] on the same DEMETER data (see below).

(9) Parrot et al. [57] aim at identifying atmospheric and ionospheric anomalies prior to four large earthquakes in China and Sumatra. Ionospheric density, total electron content, thermal infrared, outgoing longwave radiation, and atmospheric chemical potential are considered as (short-term) precursors. The authors find various precursors, backed by results published in the literature for the same earthquakes. Of all precursors presented, the outgoing longwave wavelength radiation anomalies are the clearest in both space and time, requiring further investigations. In general, a larger number of cases together with the determination of the rate of false positive (namely the study of long time series of these indicators when no earthquake occurs) must be considered in order to validate the claimed precursory signals.

(10) Nemec et al. [62] investigate the possible link between about 1000 $M5+$ earthquakes and Very Low Frequency (VLF) transmitter signals observed by the DEMETER satellite. The authors look at possible VLF signals three days before

large earthquakes and one day afterward. Emphasizing the intrinsic signal processing limitations related to DEMETER, the authors note some weak signals (a decrease in intensity observed on two transmitters with two-standard deviation confidence), but only within three hours following an earthquake. No robust precursory signal is found despite a slight increase in intensity observed on one transmitter (but not the other) recorded 50–57 hours before an earthquake. The intensity decrease observed immediately after earthquakes is considered consistent with acoustic-gravity waves propagating from the earthquake region and influencing the bottom of the ionosphere.

(11) Meng and Zhang [61] explore the potential relationship between thermal and methane anomalies associated to the 12, May 2008 M8.0 Wenchuan earthquake, China, using data from the National Center for Environmental Prediction and the AQUA satellite. The authors identify anomaly durations of c. 10–15 days for each signal, with thermal anomalies (5–14 May) occurring before the methane anomalies (12–28 May). They explain this delay as due to crustal stresses causing the thermal anomalies as well as escaping channels for underground gases. However, all methane anomalies are post-mainshock and may be related to this major rupture. Hence, only the thermal anomalies could be precursory. Use of the Granger causality test to determine whether one signal time series is useful in forecasting the other time series does not provide information on the strength of thermal anomalies as a precursor, unfortunately, with the proposed causality chain remaining conjectural. Rarely used in earthquake predictability research, the Granger causality test, which was initially introduced in econometrics, is shown by Meng and Zhang to be a potentially useful approach in the present scientific field.

(12) Qin et al. [56] investigate the association of two signals (groundwater temperature and radon content) prior to the 2008 M8.0 Wenchuan and 2010 M7.1 Yushu earthquakes, using hourly groundwater temperature data at one station and daily radon measures at another station. Positive anomalies in temperature and radon were observed before the two mainshocks, suggesting quasi-synchronous anomalies. The groundwater temperature time series shows indeed clear anomalies compared to a relatively stable background. One additional sudden increase in temperature apparently not associated to an earthquake is reported. Regarding the radon anomalies, no continuous time series is provided to estimate the baseline rate of such radon activity burst (i.e. in absence of earthquakes). Although suggestive of possible multi-signal precursors, a systematic analysis to estimate the power of such multi-data approach is lacking so far. Qin et al. additionally provide a thorough review of past claims of similar precursors and others before the same events. Their survey of seventeen quasi-synchronous precursors claimed in the literature to be observed prior to the Wenchuan earthquake (Fig. 4) is promising but would require a more rigorous statistical analysis, which is one of the goals of the GEFS. In this preliminary work, the authors leave aside the proper quantification of the gain that could be expected of such multi-data tracking.

(13) Lai et al. [64] examine changes in groundwater temperature and level at one well in Yunnan, China, over a period of ten years. Although limited to one site, the search is done systematically in time with a one-day sampling rate. Since groundwater changes have been claimed in the literature to be potentially related to far away earthquakes [29,87], the spatial extent of the analysis is large enough (500 km radius) to encompass 23 M5+ earthquakes. Their method consists in comparing the correlation between groundwater temperature and groundwater level with the occurrence of large earthquakes. They find, with 1% chance of it being random (based on Molchan's error diagram and simulations), that "*(i) when the time window is 90 days and the correlation coefficient is greater than 0.10, a sizable earthquake ($M \geq$ 5.0) is more likely to happen within 200 km during the following 45 days; (ii) when the time window is 90 days and the correlation coefficient is greater than 0.65, a sizable*

earthquake (M ≥ 5.0) is more likely to happen within 150 km during the following 45 days." The condition (*i*) is suspicious as it essentially means that the condition to declare an alarm amounts essentially to the absence of correlation, an occurrence hardly susceptible to provide information. This would suggest, combined with the second condition, the presence of some overfitting. Also, application of the same rules on a test set should allow verifying or rejecting the hypothesis. Note also that the large size of the spatiotemporal prediction window would hamper, taken alone, any practical implementation for risk mitigation.

(14) Anagnostopoulos et al. [67] investigate the correlation between solar activity and large-to-great earthquake occurrences over several solar cycles. They test the hypothesis of the electromagnetic coupling of the solar wind with the earthquake occurrence and search for the preferential conditions of such a physical relation (see the short reviews by [12,67] for the various triggering mechanisms proposed for this coupling). Their analysis yields a negative correlation between the sunspot number and M7.8+ earthquakes, shown to be significant compared to a test on a randomized set of earthquakes, during solar cycles SC21–SC24 (1980–2018). The authors further investigate possible correlations with smaller earthquakes down to $M5.5$, and with great M8.5+ events, during the period 1900–2018 (SC14–SC24), with a similar trend observed (i.e. earthquakes more likely in the decay phase of solar cycles and negative correlation between earthquakes and sunspot number). The occurrence of earthquakes would be favored during the decay phase of solar cycles when corona-hole-driven high-speed streams are most prominent or during any other phase during which coronal hole areas would also extend (sunspot number decrease). Another important result is also a twenty seven-day periodic variation of global $M > 6$ earthquakes, which matches the mean rotational period of the Sun. Since no pseudo-prospective forecast on an independent set is performed while many variables are considered, a future study should repeat this work and include such a test set to validate the strong correlations observed by the authors and investigate whether there was any degree of overfitting in the original study. An alternative approach could consist in first applying the protocol of the null hypothesis of no interaction, as proposed by [12], and demonstrate from there how the use of different data and/or methods may reject the null hypothesis. Anagnostopoulos et al. [67] discuss in depth the potential physical mechanisms, which should be further explored in the future in parallel to new statistical tests.

(15) Daneshvar and Freund [65] investigate the possible relationship between precipitation and earthquakes. Using a small data set of five $M8+$ earthquakes in Chile, the authors find a correlation $R = 0.7$ between the two phenomena. In particular, they observe that the amount of precipitation is systematically higher prior to the great earthquakes than after them, when considering equal time windows of fifteen-day, two-month, six-month and one-year, with the largest difference observed for the two-month window. Due to the limited number of cases, 5, included in the study, the results might not be representative of the global data. For instance, the considered earthquakes are not uniformly distributed over the years (occurring between end of February and mid-September), which may lead to spurious correlations with the seasonal variations of the regional weather. Indeed, it is not sufficiently acknowledged (but is well-documented in Statistics) that spurious serial correlations often emerge from diverse neglected non-stationarities (e.g. the short review in Section 2 of [88]) Only a reproduction of the same study but on the full datasets (NASA portals for precipitation and the USGS archive for earthquakes, both global datasets) would allow one to validate or reject a claim of correlation between precipitation and earthquakes. Choice of the optimal time window should also be selected using training data (in-sample), and statistical results then given for an independent test set (out-of-sample). Noteworthily, the most credible claim of a correlation between seismicity

and precipitation is the study by [89] who only considered microseismicity and not great earthquakes.

(16) Kalenda and Neumann [66] use a network of subsurface inclinometers in Eastern Europe (primarily set up to investigate thermoelastic waves due to effects by the sun) to demonstrate that horizontal deformation can also be detected before large to great earthquakes. Based on several years of observations, the authors define a stress anomaly as a sudden change of the tilt in the same direction as the principal component of the local stress field, but also by the amplitude of the anomaly against long-term background. They detect positive anomalies before 95 earthquakes from a total of 182 above magnitude seven (i.e. for half of the events). The most notable anomalies are observed from Kamchatka and Sichuan earthquakes in 2008. The results remain difficult to analyze, however, as many earthquakes occurred, and signals are observed with various time delays. While the time delay is noted to be proportional to the magnitude, localization of the mainshock is hampered by limitations associated to the stress data. Anomalies are interpreted as indicating the propagation of low to very low-frequency stress waves but in order to ascertain this observation, and for the interpretation to become convincing, future related works will need to apply the Receiver Operating Characteristic (ROC) curve (e.g. [54]), or equivalently the Molchan error diagram, to investigate true positive rates versus false positive rates. The ROC makes it possible to scan the full range of parameters and thresholds (or more generally the range of criteria used to associate an earthquake to an anomaly), providing in this way a global test that is not sensitive to specific (always difficult to justify) parameter choices.

(17) Bhardwaj et al. [70] do not present new results but propose instead a concise and worthwhile review of possible strategies when using non-seismic precursors, and insist on the necessary balance between empirical observations, instrumentation, and theoretical concepts. The authors also point to the time scales of non-seismic precursors, which are most often of the order of days to weeks (but in reality also up to years) – with very strongly varying amplitudes, the latter being only vaguely related to the size of the following seismic events, rendering prediction questionable. In our view, this suggests that a gain might be obtained when coupling them with long-term seismic precursors, with possible time scales of months to years. All in all, the authors agree with our proposition to develop a global scale monitoring system, and to rigorously test all proposed precursors – an ingredient they consider as a necessity to convince funding agencies to change their mindset about the alleged impossibility to predict earthquakes and to better support this work.

(18) Chen et al. [52] investigate geoelectric signals as potential precursors to large earthquakes (for which Chen et al. [51] provided a physical framework). They improve the Geoelectric Monitoring System's Time of Increased Probability (GEM-STIP) algorithm of [90] and test it on $M5+$ earthquakes in Taiwan. They use as geoelectric parameters the skewness (its absolute value) and kurtosis of geoelectric field two-components time series at different stations and define a so-called Anomaly Index Number (a tuned threshold) derived from those parameters. Optimal index and space-time interval of precursors and target earthquakes are tested using a binary classification scheme. An optimization based on geoelectric signal frequency band selection is also performed. The authors use the Molchan score as the metric to optimize (train) their model parameterization. They find a significant correlation between geoelectric and seismic signals on independent validation sets although no probability gain is estimated that would quantify if a precursory signal is strong or weak. They also define a new method to convert results obtained using a multi-station approach into a probabilistic map of earthquake occurrence. This new feature is unfortunately not tested on an independent dataset. Future investigations could aim at combining

geoelectric data with other types of data such as seismicity, as also proposed by [54,56].

(19) Zhuang et al. [54] discuss the critical point theory [91,92] in which large earthquakes may be forecasted in probabilistic terms. Using the branching crack model, their main claim is that the final size of an earthquake is not predictable, even at the time when it starts to propagate. They thus suggest that the only predictable feature might be the maximum size an earthquake could reach, the latter one being directly related to the size of the domain which is critically loaded. They stipulate that the critical area prior to the mainshock can be inferred from various seismicity patterns, such as the increase in the corner magnitude and the related decrease in b-value (i.e. the slope of the Gutenberg-Richter law) or the accelerating moment release [38]. They also describe the Load/Unload Response Ratio (LURR) algorithm [93] and the Epidemic-Type Aftershock Sequence (ETAS) model [68,94–96] as methods to identify criticality. They also mention gravity and electromagnetic field perturbations being possible in the critical region as a consequence of changes in the stress and related strain fields. They review the probability gains obtained in previous studies, suggesting that the ETAS probability gain is much larger than the one obtained by non-seismic precursors such as electric and magnetic signals. Moreover, beyond this review, the authors also propose a way to mix the high gains obtained from ETAS-based modelling with lower gains obtained from possibly significant precursors. For this, they propose to add some extra kernel terms that would serve as explanatory variables. In the original ETAS approach, triggering kernels are introduced to model the causative relationship between some events and those occurring later. Alternatively, one can forget the causative assumption and rather consider those kernels as correlation functions. It is then natural to introduce some other kernels that just express the correlations between some observed non-seismic anomalies with some later events as well. We would like to stress here that the ETAS model has been considerably improved over recent years by modeling subtle features such as the spatial variation of some triggering parameters or their magnitude dependence. It follows that triggered sequences are now so well described that alternative methods are under great challenge to provide better forecasting power.

(20) Nandan et al. [68] present rigorous tests of state-of-the-art statistical seismicity models, based on the Epidemic-Type Aftershock Sequence (ETAS) model. The best model is derived from the nonlinear and physics-based multifractal stress activation (MSA) model [97,98]. The tests consist in a series of pseudo-prospective forecasts over 30 days long time windows but can be easily tuned to different time horizons (from hours to years). Though there is, as the authors point out, certainly room for improvement, this is the first forecasting model of this kind at the global scale. In this sense, it defines the best null hypothesis that one should use when comparing its results with those obtained using non-seismic (or other seismic) precursors.

(21) Kamer et al. [71] propose a concrete bridge among all communities interested in earthquake prediction, in the form of a publicly accessible web-based platform (called Richter-X) that allows one to test prospectively any seismic forecasting statement. The platform is flexible enough to allow the users to define their own space, time and magnitude windows, as well as the number of predicted events. The platform allows one to compare on the same footing forecasts based on probabilistic approaches and predictions based on alarms. Together with the upstream building of a global database, this downstream feature defines the ultimate tool to rigorously monitor the set of precursors in order to distinguish those that are reliable and those that should be set aside.

Table 1. Claimed precursors and statistical robustness.

Ref.	Claimed precursor(s)*	Cases count†	Metric	Test set‡
Anagnostopoulos et al. [67]	✓(Solar activity)	Systematic	Correlation	✗
Chen et al. [52]	✓(Geoelectric)	Systematic	Molchan score	✓
Daneshvar and Freund [65]	✓(Rain)	5	Correlation	✗
Kalenda and Neumann [66]	✓(Stress)	Systematic	N/A	✗
Lai et al. [64]	✓(Groundwater)	Systematic	Molchan score	✗
Meng and Zhang [61]	✓(Thermal)	1	N/A	✗
Nemec et al. [62]	✗(Ionospheric electromagnetic)	Systematic	p-value like	✗
Parrot et al. [57]	✓(Various)	4	N/A	✗
Piroddi [58]	✓(Thermal)	1	N/A	✗
Potirakis et al. [63]	✓(Electromagnetic)	7	N/A	✗
Qin et al. [56]	✓(Groundwater, radon)	2	N/A	✗
Zhang et al. [60]	✗(Thermal)	Systematic	Weighted Molchan diagram	✗

*A cross indicates that the claimed precursor, after re-analysis, is found to be non-significant;
†A small number of cases means potential under-sampling and possibly selection bias;
‡No independent test set means potential overfitting on training set (note that overfitting depends on the variance of the proposed model, limited in the case of strongly constrained models with minimal number of free parameters).

3 Common pitfalls of pattern recognition in earthquake prediction research

We can conclude from the review of Section 2.2 that most of the reported non-seismic precursors have yet to undergo strict statistical tests (Tab. 1). This will require a re-analysis of the various datasets and re-investigation of the proposed methods, which could be realized in the GEFS. Of the most encouraging results, we note ground-based geoelectric signals. Chen et al. [52] did a systematic search, used reasonable statistical tests, and independent test sets. Zhuang et al. [54] reviewed studies based on such robust statistical tests but noticed the weakness of non-seismic precursors compared to seismic ones and it is here not clear what the probability gain of the Chen et al. [52] model is compared to an ETAS-clustering baseline for instance [68]. Both Chen et al. [52] and Zhuang et al. [54] however, emphasized the importance of probabilistic forecasts (instead of deterministic predictions) due to the stochasticity of the earthquake generating process. Regarding satellite-based analyses, Nemec et al. [62], who explored the DEMETER data set in depth (although limited in space and time and not covering the globe due to intrinsic difficulties in signal processing), found no significant precursory signal. Similarly, Zhang et al. [60] found no TIR anomaly in their systematic search, after spatial corrections. Yan et al. [53] compared the DEMETER data to ground station data and observed some synchronicity, which could help in the future to combine multi-source data to potentially obtain more robust signals, as recommended by Qin et al. [56] and Zhuang et al. [54], in the GEFS spirit. In other ground-based and satellite-based cases [56–58,61,63–67], since no systematic exploration was done and/or no test on independent data was performed, one cannot infer whether non-seismic precursors exist or are just random fluctuations that seem to be significant as a result of some data selection bias and some kind of over-fitting. Moreover, despite the many graphs of data provided, quantitative analyses remain rare. The plotting of threshold lines, windows, ellipses and other arrows pointing to claimed anomalies in some articles remains mostly based on visual inspection. Although some patterns are visually striking, e.g. outgoing long-wave radiation anomalies in Parrot et al. [57], quantitative tests remain indispensable. Immediate limitations are therefore methodological, but they are correctable.

The problems encountered in earthquake prediction studies are common pitfalls observed in statistical inference [99] and more recently in machine learning. As potential issues to keep in mind, [100] mentioned splitting data incorrectly, hidden
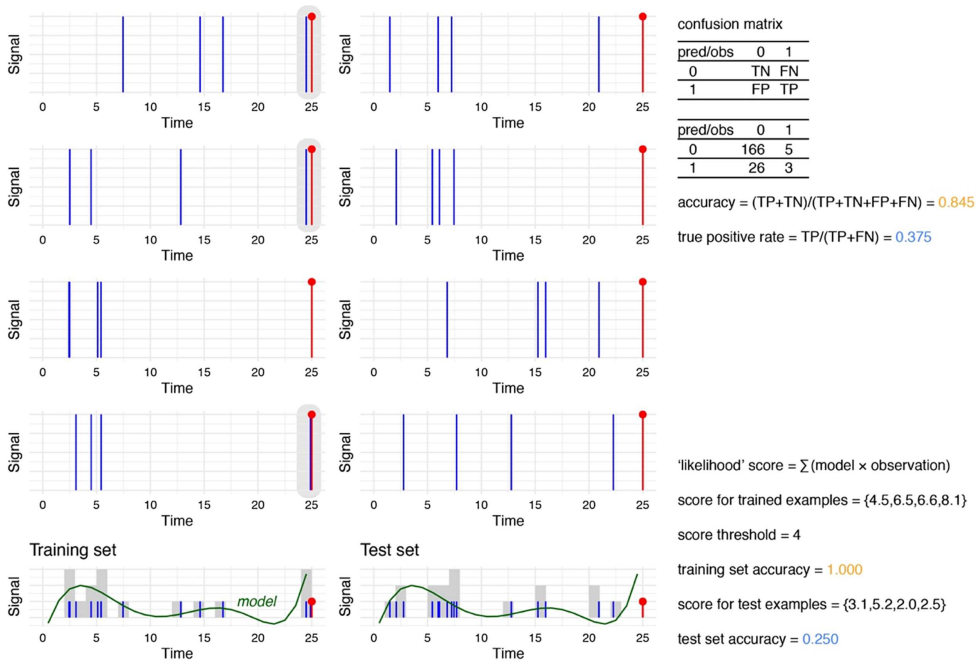
Fig. 1. Common pitfalls in retrospective investigations. In this scenario, all patterns (in blue) are random uniform in time prior to the occurrence of a mainshock (in red): (*i*) *selection bias*: despite the patterns being random, it is possible to find apparent anomalies prior to selected mainshocks (grey boxes). These events seem correlated compared to the other signals that occur much earlier, an example of finding a pattern where there is none; (ii) overfitting: fitting a flexible model (here a polynomial of degree six, in dark green) on some training set (signals on the left column) provides an accuracy of one, obtained by the definition of some scoring system and fixed threshold definition. However, application of the same model to an independent test set (right column) here yields an accuracy of 0.25 (would tend to 0.5 over larger test sets); (iii) mistaking the objective: if, for each time bin, one counts the true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN), one obtains a 0.85 accuracy due to the dataset being unbalanced. This misses the fact that we are interested in cases of mainshocks being preceded by some signal, not of no signal not followed by a mainshock. In comparison, the true positive rate (TPR) is here only 0.38.

variables, and mistaking the objective. In this respect, almost none of the reviewed studies split data between a training set and an independent test set (Tab. 1). Flexible, high-variance, machine learning models will tend to overfit the data without stringent constraints, out-of-sample tests and corrections for model flexibility. This is less problematic for high-bias models (i.e. with limited number of free parameters), although this cannot be easily assessed, hence the importance of the (out-of-sample) test set. High accuracy results on training sets are thus meaningless if not complemented by out-of-sample tests, as they do not usually generalize (Fig. 1). Beyond lack of data splitting in systematic searches, case studies of specific mainshocks, including manual selection, restricted spacetime windows, and other parameter choices may lead to biases (e.g. the claims of a universal distribution of inter-earthquake times explained simply with ETAS by [101], the claims of universal scaling laws in seismology debunked by [102], investigations of [11,103] on the role of clustering and windowing in generating spurious accelerating seismicity, of [39] on the role of the

foreshock/mainshock interval in foreshock prognostic value, and of [104] on possible spatial biases in Gutenberg-Richter b-value analyses). Retrospective selection bias is best addressed in [9] where it is described as being capable of falsely representing inherent statistical fluctuations as significant anomalies requiring fundamental explanations (Fig. 1). Once again, proper data splitting is one solution together with suitable corrections for multiple testing [105,106]. Mulargia [9] also suggested to define anomalies as patterns of extremely low probability of being random (e.g. six-sigma threshold) but recognized that real signals may be falsely rejected with such strong threshold. A balanced and well-informed statistical methodology is needed, which is often lacking in natural and geo-sciences [107,108].

As another pitfall, one should be aware of the presence of hidden variables. In other words, correlation does not imply causation. The problem of confounding variables is deep and difficult to track from data alone (e.g. thermal anomaly and topography, rain and seasonal variations, etc.), not to mention data leakage [109]. It must be considered on a case-to-case basis and under causality arguments [110]. It also depends on the experimental setting, such as the data selection discussed above. Note also that Pearson's correlation coefficient R, used by a number of studies, assumes that the data is normally distributed, which may not be the case in earthquake precursor analyses, potentially leading to further erroneous conclusions [111]. Better measures of dependence, which are not influenced by the non-Gaussian nature of the variables, include copula-based measures such as Spearman correlation, Kendall's tau and tail dependence parameters [112].

Finally, mistaking the objective can mean maximizing the wrong scoring metric. Again, one should not make inferences based on R alone. Accuracy is also not a sound metric when the data is unbalanced, i.e., when most samples are in the same category (e.g. trivial case of no signal followed by no mainshock, Fig. 1). Use of the ROC curve or Molchan diagram would already provide a comprehensive summary of the model performance (preferentially on an independent test set) although unbalanced data remains a problem in this case [60,113]. The objective should also be to obtain a probability gain compared to a sound null hypothesis, which so far is best characterized by the ETAS model [54,68,96,114]. Random simulations may be misleading when null hypotheses are over-simplified, i.e. not representative of the normal behavior of seismicity.

As a side note, since all the aforementioned issues may concern other peer-reviewed publications as well, the number of publications on a given precursory signal should not be taken as argument for the validity of that hypothetical precursor. In other words, the truth in Science is not (or should not be) decided by democratic-like majority considerations or as Mahatma Gandhi had said: "*An error does not become truth by reason of multiplied propagation, nor does truth become error because nobody sees it.*"

4 Concluding remarks: The challenges ahead and the GEFS mission

In view of the difficulties in convincingly finding causal relationships between earthquake precursors and earthquake occurrences, two alternative approaches are worth undertaking: (1) the physically based perspective of using a strong theoretical framework that explains and quantifies precursory patterns, and (2) the high-variance perspective of letting the data speak for themselves. These represent the two end members of the well-known bias-variance tradeoff in statistical inference and more recently rediscovered in machine learning. In the first case, no claim of predictability should be given if the proposed physical laws are not confirmed by data. Some of

the main investigators of the GEFS are proponents of various theories [41,50,69,76–78,91,92] which may need further investigations and formalization to validate or reject specific "laws" based on GEFS data. Although these theoretical frameworks and others will provide directions to follow in data exploration and model validation, they do not form the core of the GEFS, which is data-oriented and theoretically agnostic. In this second case, the main challenge is the creation of a large database for data analytics and forecasting.

Data types include satellite images, in the range of tens to hundreds of Gb per source for a couple of years of recordings, navigation satellite data, in the range of several Tb per provider, other satellite data (e.g. DEMETER dataset is eight Tb in size), ground station data such as earthquake catalogs, geoelectric or air ionization time series, etc. (in Mb-Gb range per source and/or region). This means that the GEFS database will enter in the Big Data category, with everything that this entails in terms of data processing in volume, variety and speed of execution. Challenges exist at all levels, from data gathering to data visualization via data storage, querying and analysis [115]. The capacities only exist since recently [116] and have yet to be tapped for earthquake predictability research. None of the aforementioned challenges are trivial to solve, each requiring the development of new methods for data mining (e.g. mixture modelling for censored seismicity data [117]) and proper machine learning usage such as avoiding overcomplex models [118] and data leakage [109], on top of the technological requirements in storage, pipelining, etc.

Is it worthwhile to engage on such a complicated journey? Virtually all earthquake prediction projects have failed to achieve their aim (apart from the forgotten and abandoned Chinese prediction initiatives from 1966 to 1976 [32]), and the studies of the present special issue confirm the overall lack of robust precursory signals, despite some encouraging results observed. Yet, earthquake science is disappointing if not able to predict its sole object of study. Kagan [8] asked more than 20 years ago whether earthquake seismology was a hard, quantitative science. The answer remains unclear to this day. The introductory quote by one of us [50] remains as valid now as it was then: "*Notwithstanding past efforts in several countries in the last decades, I fail to see that the scientific community has used the full potential of artificial/computational intelligence, statistical physics, super-computer modelling, large scale monitoring of a full spectrum of physical measurements, coupled together with more traditional seismological and geological approaches to make a dent in the earthquake problem. What we have learned is that past failures in earthquake prediction reflect the biased view that it was a simple problem.*" Using a systematic, all-encompassing approach to the problem, using all data that is currently available, the GEFS has the aim to provide inspiring answers regarding the degree of predictability of earthquakes within a decade.

Publisher's Note The EPJ Publishers remain neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. C.H. Scholz, L.R. Sykes, Y.P. Aggarwal, Science **181**, 803 (1973)
2. P. Varotsos, K. Alexopoulos, K. Nomicos, M. Lazaridou, Nature **322**, 120 (1986)
3. M. Wyss, R.E. Habermann, Pure Appl. Geophys. **126**, 319 (1988)
4. D.D. Bowman, G. Ouillon, C.G. Sammis, A. Sornette, D. Sornette, J. Geophys. Res. **103**, 24359 (1998)
5. N.V. Sarlis et al., PNAS **110**, 13734 (2013)
6. F. Mulargia, P. Gasperini, Geophys. Res. Lett. **23**, 1327 (1996)
7. R.J. Geller, D.D. Jackson, Y.Y. Kagan, F. Mulargia, Science **275**, 1616 (1997)
8. Y.Y. Kagan, Pure Appl. Geophys. **155**, 233 (1999)
9. F. Mulargia, Geophys. J. Int. **146**, 489 (2001)
10. D.D. Jackson, State Planet Front. Challenges Geophys. **150**, 335 (2004)
11. J.L. Hardebeck, K.R. Felzer, A.J. Michael, J. Geophys. Res. **113**, B08310 (2008)
12. J.J. Love, J.N. Thomas, Geophys. Res. Lett. **40**, 1165 (2013)
13. L.R. Sykes, D.W. Simpson, P.G. Richards, Maurice Ewing Ser. **4**, vii (1981)
14. F.F. Evison, *Earthquake prediction* (Terra Scientific Publishing Company, Tokyo; UNESCO, Paris, 1984)
15. M. Wyss, Am. Geophys. Union **72**, 411 (1991)
16. L. Knopoff, PNAS **93**, 3719 (1996)
17. V.I. Keilis-Borok, Phys. Earth Planet Int. **111**, 179 (1999)
18. I. Main, *Moderator, Is the reliable prediction of individual earthquakes a realistic scientific goal?*, Nature online discussion (25th February – 8th April 1999), Available at https://www.nature.com/articles/nature28107.pdf
19. R.A. Harris, J.R. Arrowsmith, Bull. Seismol. Soc. Am. **96**, S1 (2006)
20. K.F. Tiampo, M. Anghel, Tectonophysics **413**, 1 (2006)
21. E.H. Field, Seismol. Res. Lett. **78**, 7 (2007)
22. W. Marzocchi, D. Schorlemmer, S. Wiemer, Ann. Geophys. **53**, i (2010)
23. K.Z. Nanjo, H. Tsuruoka, N. Hirata, T.H. Jordan, Earth Planets Space **63**, 159 (2011)
24. F.F. Freund, V. Stolc, Animals **3**, 513 (2013)
25. Q. Huang, M. Gerstenberger, S. Chen, Pure Appl. Geophys. **173**, 1 (2016)
26. A.J. Michael, M.J. Werner, Seismol. Res. Lett. **89**, 1226 (2018)
27. Q. Deng, J. Pu, L.M. Jones, P. Molnar, A preliminary analysis of reported changes in ground water and anomalous animal behavior before the 4 February 1975 Haicheng earthquake, in *Earthquake Prediction, An International Review*, edited by D.W. Simpson, P.G. Richards (American Geophysical Union, Washington, D.C., 1981), pp. 543–565
28. J.-Z. Zheng, J. Seismol. Soc. Jpn. **34**, 43 (1981)
29. Z. Ma, Z. Fu, Y. Zhang, C. Wang, G. Zhang, D. Liu, *Earthquake prediction (nine major earthquakes in China, 1966-1976)* (Seismological Press Beijing, Springer-Verlag, 1990)
30. K. Wang, Q.-F. Chen, S. Sun, A. Wang, Bull. Seismol. Soc. Am. **96**, 757 (2006)
31. F. Huang, M. Li, Y. Ma, Y. Han, L. Tian, W. Yan, X. Li, Geod. Geodyn. **8**, 1 (2017)
32. E. Mearns, D. Sornette *The Great 1976 Tangshan Earthquake: Learning from the 1966-1976 Chinese earthquake prediction projects* (book in press)
33. H. Kanamori, in *Earthquake Prediction: An International Review* (1981), Vol. 4, pp. 1–19
34. V. Keilis-Borok, Annu. Rev. Earth Planet. Sci. **30**, 1 (2002)
35. R.E. Buskirk, C. Frohlich, G.V. Latham, Rev. Geophys. Space Phys. **19**, 247 (1981)
36. M. Wyss, R.E. Habermann, Pure Appl. Geophys. **126**, 319 (1988)
37. S.C. Jaumé, L.R. Sykes, Pure Appl. Geophys. **155**, 279 (1999)
38. A. Mignan, Tectonophysics **505**, 1 (2011)
39. A. Mignan, Sci. Rep. **4**, 4099 (2014)
40. D. Sornette, Phys. Rep. **313**, 238 (1999)
41. D Sornette, Mechanochemistry: an hypothesis for shallow earthquakes, in *Earthquake thermodynamics and phase transitions in the Earth's interior*, edited by R. Teisseyre, E. Majewski (Cambridge University Press, 2000), pp. 329–366
42. Y. Ben-Zion, V. Lyakhovsky, Pure Appl. Geophys. **159**, 2385 (2002)

43. S. Pulinets, K.A. Boyarchuk, *Ionospheric Precursors of Earthquakes* (Springer, Berlin, Germany, 2004)
44. R.D. Cicerone, J.E. Ebel, J. Britton, Tectonophysics **476**, 371 (2009)
45. T. Ma, Z. Wu, Int. J. Geophys. **2012**, 583097 (2012)
46. T.H. Jordan et al., Ann. Geophys. **54**, 315 (2011)
47. A. Mignan, M. Broccardo, Seismol. Res. Lett. **91**, 2330 (2020)
48. M.C. Gerstenberg, S. Wiemer, L.M. Jones, P.A. Reasenberg, Nature **435**, 328 (2005)
49. K. Wang, Q.-F. Chen, S. Sun, A. Wang, Bull. Seismol. Soc. Am. **96**, 757 (2006)
50. D. Sornette, Towards a truly multidisciplinary approach to earthquake prediction, Nature debate April 1999, in *Is the reliable prediction of individual earthquakes a realistic scientific goal?* (1999)
51. H.-J. Chen, C.-C. Chen, G. Ouillon, D. Sornette, Eur. Phys. J. Special Topics **230**, 67 (2021)
52. H.-J. Chen, C.-C. Chen, G. Ouillon, D. Sornette, Eur. Phys. J. Special Topics **230**, 381 (2021)
53. R. Yan, L. Wang, M. Parrot, X. Zhang, Z. Hu, Eur. Phys. J. Special Topics **230**, 179 (2021)
54. J. Zhuang, M. Matsu'ura, P. Han, Eur. Phys. J. Special Topics **230**, 409 (2021)
55. F.T. Freund, J. A. Heraud, V. A. Centa, J. Scoville, Eur. Phys. J. Special Topics **230**, 47 (2021)
56. K. Qin, Z. Shuo, L. Wu, Y. Wang, Eur. Phys. J. Special Topics **230**, 263 (2021)
57. M. Parrot, V. Tramutoli, T.J.Y. Liu, S. Pulinets, D. Ouzounov, N. Genzano, M. Lisi, K. Hattori, A. Namgaladze, Eur. Phys. J. Special Topics **230**, 197 (2021)
58. L. Piroddi, Eur. Phys. J. Special Topics **230**, 111 (2021)
59. J. Scoville, F. Freund, Eur. Phys. J. Special Topics **230**, 85 (2021)
60. Y. Zhang, Q. Meng, G. Ouillon, L. Zhang, D. Hu, W. Ma, D. Sornette, Eur. Phys. J. Special Topics **230**, 133 (2021)
61. Q. Meng, Y. Zhang, Eur. Phys. J. Special Topics **230**, 247 (2021)
62. F. Nemec, M. Liska, M. Parrot, Eur. Phys. J. Special Topics **230**, 227 (2021)
63. S.M. Potirakis, Y. Contoyiannis, A. Schekotov, K. Eftaxias, M. Hayakawa, Eur. Phys. J. Special Topics **230**, 151 (2021)
64. G. Lai, C. Jiang, W. Wang, L. Han, S. Deng, Eur. Phys. J. Special Topics **230**, 275 (2021)
65. M.R. Mansouri Daneshvar, F.T. Freund, Eur. Phys. J. Special Topics **230**, 335 (2021)
66. P. Kalenda, L. Neumann, Eur. Phys. J. Special Topics **230**, 353 (2021)
67. G. Anagnostopoulos, I. Spyroglou, A. Rigas, P. Preka-Papadema, H. Mavromichalaki, I. Kiosses, Eur. Phys. J. Special Topics **230**, 287 (2021)
68. S. Nandan, Y. Kamer, G. Ouillon, S. Hiemer, D. Sornette, Eur. Phys. J. Special Topics **230**, 425 (2021)
69. F. Freund, G. Ouillon, J. Scoville, D. Sornette, Eur. Phys. J. Special Topics **230**, 7 (2021)
70. A. Bhardwaj, L. Sam, F.J. Martin-Torres, Eur. Phys. J. Special Topics **230**, 367 (2021)
71. Y. Kamer, S. Nandan, G. Ouillon, S. Hiemer, D. Sornette, Eur. Phys. J. Special Topics **230**, 451 (2021)
72. D. Sornette, G. Ouillon, A. Mignan, F. Freund, Eur. Phys. J. Special Topics **230**, 1 (2021)
73. Q. Kong, R.M. Allen, L. Schreier, Y.-W. Kwon, Sci. Adv. **2**, e1501055 (2016)
74. C.G. Sammis, D. Sornette, PNAS **99**, 2501 (2002)
75. F. Freund, D. Sornette, Tectonophysics **431**, 33 (2007)
76. A. Mignan, Geophys. Res. Lett. **39**, L21308 (2012)
77. A. Mignan, Nonlin. Processes Geophys. **23**, 107 (2016)
78. A. Mignan, Nonlin. Processes Geophys. **25**, 241 (2018)
79. S. Pulinets, D. Ouzounov, J. Asian Earth Sci. **41**, 371 (2011)
80. A. Mignan, C. Jiang, J.D. Zechar, S. Wiemer, Z. Wu, Z. Huang, Bull. Seismol. Soc. Am. **103**, 845 (2013)
81. A. Helmstetter, D. Sornette, J. Geophys. Res. **108**, 2457 (2003)

82. A. Helmstetter, D. Sornette, J.-R. Grasso, J. Geophys. Res. **108**, 2046 (2003)
83. T. van Stiphout, D. Schorlemmer, S. Wiemer, Bull. Seismol. Soc. Am. **101**, 482 (2011)
84. J. Zhuang, Y. Ogata, D. Vere-Jones, J. Am. Stat. Assoc. **97**, 369 (2002)
85. G.M. Molchan, Phys. Earth Planet. Int. **61**, 84 (1990)
86. A. Eleftheriou, C. Filizzola, N. Genzano, T. Lacava, M. Lisi, R. Paciello, N. Pergola, F. Vallianatos, V. Tramutoli, Pure Appl. Geophys. **173**, 285 (2016)
87. S. Mei, J. Phys. Earth **34**, S193 (1986)
88. S. Wheatley, A. Wehrli, D. Sornette, Quant. Finance **19**, 1165 (2019)
89. S. Hainzl, T. Kraft, J. Wassermann, H. Igel, E. Schmedes, Geophys. Res. Lett. **33**, L19303 (2006)
90. H.-J. Chen, C.-C. Chen, Nat. Hazards **84**, 877 (2016)
91. D. Sornette, C.G. Sammis, J. Phys. I. France **5**, 607 (1995)
92. G. Ouillon, D. Sornette, Geophys. J. Int. **143**, 454 (2000)
93. X.-H. Yin, X.-Z. Chen, Z.-P. Song, C. Yin, Pure Appl. Geophys. **145**, 701 (1995)
94. Y. Ogata, Ann. Inst. Statist. Math. **50**, 379 (1998)
95. S. Seif, A. Mignan, J.D. Zechar, M.J. Werner, S. Wiemer, J. Geophys. Res. Solid Earth **122**, 449 (2017)
96. S. Nandan, G. Ouillon, S. Wiemer, D. Sornette, J. Geophys. Res. Solid Earth **122**, 5118 (2017)
97. G. Ouillon, D. Sornette, J. Geophys. Res. **110**, B04306 (2005)
98. D. Sornette, G. Ouillon, Phys. Rev. Lett. **94**, 038501 (2005)
99. D.R. Cox, *Principles of Statistical Inference* (Cambridge University Press, 2006)
100. P. Riley, Nature **572**, 27 (2019)
101. A. Saichev, D. Sornette, Phys. Rev. Lett. **97**, 078501 (2006)
102. M. Werner, D. Sornette, Phys. Rev. Lett. **99**, 179801 (2007)
103. A. Mignan, Tectonophysics **452**, 42 (2008)
104. Y. Kamer, S. Hiemer, J. Geophys. Res. Solid Earth **120**, 5191 (2015)
105. J.P. Romano, M. Wolf, Econometrica **73**, 1237 (2005)
106. J.P. Romano, M. Wolf, Stat. Proba. Lett. **113**, 38 (2016)
107. D. Sornette, V. Pisarenko, Eos Transactions AGU **92**, 64 (2011)
108. S. Wheatley, D. Sornette, Physica A, submitted (14 Oct 2019), arXiv:1904.05662
109. A. Mignan, M. Broccardo, Earth Planet. Sci. Lett. **544**, 116402 (2020)
110. J. Pearl, Commun. ACM **62**, 54 (2019)
111. F.J. Anscombe, Am. Stat. **27**, 17 (1973)
112. Y. Malevergne, D. Sornette *Extreme financial risks (from dependence to risk management)* (Springer Heidelberg 2006)
113. T. Parsons, Geophys. Res. Lett., in press (2020)
114. S. Seif, J.D. Zechar, A. Mignan, S. Nandan, S. Wiemer, Bull. Seismol. Soc. Am. **109**, 1 (2019)
115. C. L. P. Chen CLP, C.-Y. Zhang, Inf. Sci. **275**, 314 (2014)
116. M. Hilbert, P. López, Science **332**, 60 (2011)
117. A. Mignan, Geophys. J. Int. **219**, 1348 (2019)
118. A. Mignan, M. Broccardo, Nature **574**, E1 (2019)