



1 **Scale-dependent spatial coherence between historical and instrumental earthquake catalogues at the**  
2 **global scale**

3  
4

5 Antonino D'Alessandro

6 Istituto Nazionale di Geofisica e Vulcanologia, Osservatorio Nazionale Terremoti, Rome, Italy

7 antonino.dalessandro@ingv.it

8  
9

10 **Abstract**

11 Historical earthquake catalogues extend seismic observations back by several centuries and are widely used in  
12 seismic hazard and tectonic studies, yet their global-scale informational content remains difficult to quantify  
13 due to strong spatial, temporal, and magnitude-dependent reporting biases. In this study, we present a  
14 quantitative, spatially explicit assessment of the consistency between global historical (1600–1899) and early  
15 instrumental (1900–1950,  $M \geq 5.5$ ) earthquake catalogues. Rather than relying on magnitude-based  
16 comparisons, we represent earthquake occurrence as spatial probability density fields obtained through  
17 Gaussian smoothing and define a scale-dependent spatial coherence metric based on the overlap between  
18 historical and instrumental distributions. This approach allows us to isolate large-scale tectonic signal from  
19 localized reporting artefacts and to systematically explore the role of spatial scale. Our results show that spatial  
20 coherence between historical and instrumental seismicity is low at small scales and increases monotonically  
21 with smoothing length, reaching moderate values only at regional to continental scales. Even at the largest  
22 scales considered, coherence remains well below unity, indicating that only a limited fraction of the  
23 instrumental spatial pattern is recoverable from historical data. Decomposition by tectonic domain reveals that  
24 subduction zones dominate the historical–instrumental agreement, while continental collision belts and  
25 intraplate regions contribute substantially less. These findings demonstrate that global historical earthquake  
26 catalogues contain a detectable but intrinsically limited imprint of tectonic structure. Meaningful use of  
27 historical seismicity at the global scale therefore requires explicit consideration of spatial scale and tectonic



28 context. The framework proposed here provides a transparent and reproducible basis for evaluating the  
29 reliability of historical earthquake data in seismic hazard and global seismotectonic applications.

30

31

## 32 **1. Introduction**

33 Earthquake catalogs constitute the primary empirical basis for seismic hazard assessment, tectonic studies, and  
34 long-term analyses of earthquake occurrence. While instrumental seismic networks provide spatially and  
35 temporally homogeneous observations only since the early–mid twentieth century, historical earthquake  
36 catalogs extend the temporal window of observation back by several centuries, offering the potential to  
37 investigate rare large events and long-term seismic patterns (e.g. Guidoboni and Stucchi, 2011; Albini et al.,  
38 2014).

39 Despite their importance, historical earthquake catalogs are affected by profound and well-known limitations.  
40 The reporting of pre-instrumental earthquakes depends on population density, literacy, administrative  
41 organization, and the survival of historical sources, resulting in highly heterogeneous spatial and temporal  
42 coverage (Ambraseys, 1989; Guidoboni et al., 2007). In addition, magnitude information is often incomplete,  
43 uncertain, or retrospectively inferred from macroseismic intensity data, introducing systematic biases that  
44 complicate quantitative analyses (Bakun and Wentworth, 1997; Stucchi et al., 2013). These limitations are  
45 particularly critical when historical data are used in a probabilistic seismic hazard context, where assumptions  
46 on completeness and representativeness directly affect hazard estimates (Albini et al., 2015).

47 At the global scale, several compilations of historical seismicity are currently available, including the Global  
48 Historical Earthquake Catalogue (GHEC; Albini et al., 2014), the European Pre-Instrumental Catalogue  
49 (EPICA; Rovida et al., 2022), and the NOAA Significant Earthquake Database (NOAA, 2023). These datasets  
50 differ substantially in scope, selection criteria, and methodological philosophy, but they share common  
51 limitations related to reporting completeness and spatial representativeness. Recent studies have emphasized  
52 that catalog completeness should not be regarded as a binary property, but rather as a scale-, magnitude-, and  
53 region-dependent quantity, whose uncertainty must be explicitly considered (Woessner et al., 2015; Gómez  
54 Capera et al., 2020). As a result, the degree to which historical catalogs capture the underlying tectonic signal  
55 of global seismicity remains an open and debated question.



56 Previous investigations have primarily addressed this issue at regional scales, often through qualitative  
57 comparisons or by focusing on specific well-documented areas such as Europe, the Mediterranean, or Japan  
58 (Ambraseys and Jackson, 1998; Guidoboni and Ebel, 2009). At the global scale, however, a systematic and  
59 quantitative assessment of the spatial consistency between historical and instrumental seismicity is still  
60 lacking. In particular, little attention has been paid to the spatial scales at which historical earthquake  
61 distributions may transition from being dominated by reporting artifacts to reflecting genuine tectonic  
62 structure.

63 A key challenge arises from the intrinsic mismatch between historical and instrumental datasets. Instrumental  
64 catalogs, even in their early phases, provide more uniform detection thresholds and spatial coverage, especially  
65 for moderate-to-large earthquakes (e.g. Gutenberg and Richter, 1954; Utsu, 2002). For this reason, **early**  
66 **instrumental seismicity can be used as an approximate reference for the large-scale spatial organization of**  
67 **earthquake occurrence. Spatially explicit and probabilistic comparison frameworks, already widely adopted in**  
68 **earthquake forecast testing and model evaluation (Schorlemmer et al., 2010; Helmstetter et al., 2007), offer a**  
69 **natural basis for such analyses.**

70 In this study, we introduce a quantitative, spatially explicit framework to assess the coherence between  
71 historical and instrumental earthquake catalogs at the global scale. Rather than relying on magnitude-based  
72 comparisons, which are strongly affected by historical incompleteness (Mignan and Woessner, 2012), we focus  
73 on the spatial distributions of earthquake occurrence. Historical (1600–1899) and early instrumental (1900–  
74 1950,  $M \geq 5.5$ ) seismicity are represented as spatial probability density fields obtained through Gaussian  
75 smoothing, allowing us to define a scale-dependent measure of spatial coherence.

76 By systematically varying the smoothing scale and decomposing the analysis by tectonic domain, we aim to  
77 answer three central questions:

- 78 i. At which spatial scales does historical seismicity exhibit statistically meaningful agreement with  
79 instrumental seismicity?
- 80 ii. How does this coherence depend on tectonic setting (e.g. subduction zones, continental collision belts,  
81 intraplate regions)?
- 82 iii. What are the implications of these results for the use of global historical earthquake catalogs in seismic  
83 hazard and tectonic studies?



84 By explicitly quantifying the limits and conditions under which historical earthquake data reflect tectonic  
85 reality, this work provides a robust methodological basis for their appropriate and informed use in global-scale  
86 analyses.

87

88

## 89 **2. Data**

90 This study relies exclusively on openly accessible earthquake catalogues in order to ensure transparency,  
91 reproducibility, and long-term traceability of the results. Two distinct classes of data are considered: historical  
92 earthquake catalogues documenting pre-instrumental seismicity, and an early instrumental catalogue used as a  
93 spatial reference for global earthquake occurrence.

94 Historical seismicity is represented by three major compilations covering the period 1600–1899: the Global  
95 Historical Earthquake Catalogue (GHEC; Albini et al., 2014), the European Pre-Instrumental Earthquake  
96 Catalogue (EPICA; Rovida et al., 2022), and the NOAA Significant Earthquake Database (NOAA, 2023).  
97 GHEC provides a globally harmonized dataset derived from critically reviewed regional and national  
98 catalogues, with explicit documentation of sources and selection criteria. EPICA represents the most up-to-  
99 date and systematically revised compilation of pre-instrumental earthquakes for Europe and the Mediterranean  
100 region. The NOAA database complements these resources by including globally distributed significant  
101 historical earthquakes, albeit with heterogeneous reporting standards and variable parameter completeness.

102 All historical events reported between 1600 and 1899 were initially retained. Quality control procedures were  
103 applied to exclude events lacking geographic coordinates or a defined year of occurrence. No attempt was  
104 made to homogenize historical magnitude estimates or to convert macroseismic intensities into magnitudes, as  
105 such procedures introduce additional epistemic uncertainty and are known to be strongly dependent on region-  
106 and period-specific assumptions (Bakun and Wentworth, 1997; Stucchi et al., 2013). Consequently, historical  
107 data are treated exclusively in terms of spatial occurrence, without using magnitude as a quantitative variable.

108 To provide a reference representation of the large-scale spatial organization of global seismicity, an early  
109 instrumental earthquake catalogue covering the period 1900–1950 was constructed from the United States  
110 Geological Survey (USGS) database. A conservative magnitude threshold of  $M \geq 5.5$  was adopted to ensure  
111 near-global detectability even during the early instrumental era (Gutenberg and Richter, 1954; Utsu, 2002).



112 Although this catalogue does not achieve the completeness of modern global datasets, it offers substantially  
113 more homogeneous spatial coverage than historical records and is therefore used here as a proxy for the  
114 underlying tectonic control on earthquake occurrence at regional to continental scales.

115 All catalogues were projected onto a common global geographic grid and processed using identical spatial  
116 discretization. Potential duplicate events across catalogues were identified based on spatial and temporal  
117 proximity and flagged for diagnostic purposes, but not forcibly merged, in order to preserve the original  
118 information content of each dataset. Subsequent analyses are based on spatial probability density fields derived  
119 from event locations, rather than on individual earthquake parameters. This strategy minimizes sensitivity to  
120 catalogue-specific uncertainties in magnitude, depth, and origin time, and ensures that the results reflect robust  
121 spatial patterns rather than artefacts of catalogue construction.

122

123

### 124 3. Methods

125 The methodological framework adopted in this study is designed to quantify the spatial consistency between  
126 historical and instrumental earthquake catalogues in a scale-dependent and tectonically informed manner. The  
127 analysis is entirely based on earthquake locations and avoids the direct use of magnitude information, which  
128 is known to be highly heterogeneous and incomplete in historical datasets.

129 Both historical (1600–1899) and instrumental (1900–1950) earthquake catalogues are first represented as  
130 discrete spatial point processes on a common global geographic grid. To enable a meaningful comparison  
131 between datasets characterized by different levels of spatial completeness and reporting noise, earthquake  
132 locations are transformed into continuous spatial probability density fields through isotropic Gaussian kernel  
133 smoothing. For a given catalogue  $k$  (historical or instrumental), the smoothed spatial density field  $p_k^\sigma(\mathbf{x})$  is  
134 defined as

$$136 \quad p_k^\sigma(\mathbf{x}) = \frac{1}{N_k} \sum_{i=1}^{N_k} \frac{1}{2\pi\sigma^2} \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}_i\|^2}{2\sigma^2}\right), \quad (1)$$

135



137 where  $\mathbf{x}_i$  denotes the epicentral location of the  $i$ -th earthquake in catalogue  $k$ ,  $N_k$  is the total number of events  
138 in that catalogue, and  $\sigma$  is the Gaussian smoothing length scale. This normalization ensures that each density  
139 field integrates to unity over the global domain and can therefore be interpreted as a spatial probability density  
140 function.

141 The smoothing scale  $\sigma$  is treated as a free parameter and systematically varied in the range 50–300 km. This  
142 interval spans scales from sub-regional, where historical reporting noise is expected to dominate, to continental  
143 scales, where the imprint of large-scale tectonic structure should emerge. All analyses are performed on  
144 identical grids and with identical smoothing parameters for both historical and instrumental datasets, ensuring  
145 methodological consistency.

146 To quantify the spatial agreement between the two density fields, we define a scale-dependent spatial coherence  
147 metric  $C(\sigma)$  as the total overlap between the historical and instrumental probability density functions,

148

$$150 \quad C(\sigma) = \int_{\Omega} \min(p_{\text{hist}}^{\sigma}(\mathbf{x}), p_{\text{inst}}^{\sigma}(\mathbf{x})) \, d\mathbf{x}, \quad (2)$$

149

151 where  $\Omega$  denotes the global spatial domain. By construction,  $0 \leq C(\sigma) \leq 1$ , with  $C(\sigma) = 1$  corresponding to  
152 identical spatial distributions and  $C(\sigma) = 0$  indicating no spatial overlap. This metric emphasizes shared large-  
153 scale spatial structure while being insensitive to localized discrepancies, making it particularly suitable for  
154 comparing datasets with heterogeneous completeness.

155 To investigate the role of tectonic setting, the global coherence is further decomposed into contributions from  
156 distinct tectonic domains. Tectonic regions are defined using the PB2002 global plate boundary model (Bird,  
157 2003), which provides a geometrically consistent representation of plate boundaries and their kinematic  
158 classification. Each grid cell is assigned to one of three broad tectonic domains: subduction zones, continental  
159 collision belts, or intraplate regions. Grid cells located within 200 km of a PB2002 boundary segment are  
160 classified according to the dominant boundary type, while remaining cells are assigned to the intraplate  
161 domain.

162 For each tectonic domain  $d$ , a domain-specific coherence contribution  $C_d(\sigma)$  is computed as

163



165 
$$C_d(\sigma) = \int_{\Omega_d} \min(p_{\text{hist}}^\sigma(\mathbf{x}), p_{\text{inst}}^\sigma(\mathbf{x})) \, d\mathbf{x}, \quad (3)$$

164

166 where  $\Omega_d \subset \Omega$  denotes the spatial domain associated with tectonic setting  $d$ . By construction, the sum of the  
167 domain-specific contributions exactly reproduces the global coherence,

168

170 
$$C(\sigma) = \sum_d C_d(\sigma).$$

169

171 This decomposition allows the relative contribution of different tectonic environments to the overall historical–  
172 instrumental agreement to be quantified as a function of spatial scale.

173 No temporal smoothing is applied beyond the selection of the analysis periods, and no declustering is  
174 performed. This choice reflects the focus of the study on long-term spatial patterns rather than on short-term  
175 seismic interactions. Sensitivity to catalogue size and smoothing scale is addressed explicitly through the  
176 systematic exploration of  $\sigma$ , while uncertainties related to magnitude, depth, and focal mechanism are  
177 intentionally excluded from the analysis to avoid introducing poorly constrained assumptions.

178

179

#### 180 **4. Results**

181 The spatial distribution of historical earthquake reporting exhibits pronounced heterogeneity at the global  
182 scale. **Figure 1** illustrates the geographic coverage of historical seismicity between 1600 and 1899 as derived  
183 from the combined GHEC, EPICA, and NOAA datasets. Reported events are strongly concentrated in Europe  
184 and the Mediterranean region, Japan, parts of East Asia, and selected areas of the Americas, whereas vast  
185 regions, including most of Africa, oceanic domains, and large intraplate areas, display sparse or absent  
186 reporting. This pattern reflects the combined influence of tectonic activity and historically contingent reporting  
187 conditions rather than a homogeneous representation of global seismicity.

188 The temporal evolution of historical reporting further emphasizes its non-stationary character. Figure 2 shows  
189 the smoothed temporal evolution of the number of reported historical earthquakes using a wide moving  
190 window. **The gradual increase in reporting rate over time, particularly from the eighteenth century onward,**



191 does not exhibit clear episodic behaviour that could be robustly attributed to changes in global seismicity rates.  
192 Instead, the observed trends are consistent with progressive improvements in documentation, communication,  
193 and archival preservation. This temporal variability highlights the difficulty of interpreting raw historical event  
194 counts in terms of physical seismic processes.

195 Spatial probability density maps of historical seismicity, constructed through Gaussian smoothing, provide  
196 further insight into the scale-dependent nature of historical information (Fig. 3). At relatively small smoothing  
197 scales, the density maps are dominated by localized clusters associated with well-documented regions, while  
198 at larger scales persistent large-scale features emerge. These features broadly correspond to major tectonic  
199 structures, such as subduction zones and active plate boundaries, suggesting that historical earthquake  
200 locations contain a non-random tectonic signal when considered at sufficiently large spatial scales. However,  
201 even at these scales, the density fields remain far from globally uniform, underscoring the persistent influence  
202 of reporting bias.

203 A direct comparison between historical and instrumental spatial density fields is presented in Figure 4. The  
204 instrumental reference catalogue (1900–1950,  $M \geq 5.5$ ) displays a spatial pattern closely aligned with global  
205 plate boundaries, whereas the historical density field shows only partial correspondence. The resulting overlap  
206 between the two distributions is spatially limited and concentrated in a small number of regions. The global  
207 spatial coherence, quantified as the integrated overlap between the two density fields, remains substantially  
208 below unity, indicating that only a fraction of the instrumental spatial pattern is captured by historical reporting.

209 The limitations associated with historical magnitude information are illustrated in Figure 5. The fraction of  
210 historical events with reported magnitudes varies strongly through time and across catalogues, with a marked  
211 increase only in the late nineteenth century. Complementary cumulative magnitude distributions reveal  
212 substantial inconsistencies between catalogues, reflecting differences in magnitude assignment practices and  
213 data availability. These results confirm that historical magnitude data are unsuitable for direct quantitative  
214 comparison with instrumental catalogues at the global scale and justify the exclusive focus on spatial  
215 information in the present analysis.

216 The dependence of historical–instrumental spatial coherence on the smoothing scale is quantified in Figure 6.  
217 The coherence  $C(\sigma)$  increases monotonically with increasing  $\sigma$ , from values of approximately 0.13 at 50 km  
218 to about 0.19 at 300 km. This behaviour indicates that small-scale discrepancies, dominated by reporting noise



219 and localized incompleteness, progressively diminish as spatial averaging increases. No clear saturation is  
220 observed at the smallest scales, whereas a gradual approach to an asymptotic value occurs at larger scales,  
221 suggesting that only a limited fraction of the instrumental spatial structure is recoverable from historical data  
222 even after substantial smoothing.

223 The role of tectonic setting in controlling spatial coherence is further examined in **Figure 7**, where the global  
224 coherence is decomposed into contributions from subduction zones, continental collision belts, and intraplate  
225 regions. Subduction zones provide the dominant contribution to the total coherence across all smoothing scales,  
226 followed by intraplate regions, while collision zones contribute comparatively less. The relative contribution  
227 of each tectonic domain remains stable with increasing  $\sigma$ , indicating that the scale dependence of coherence is  
228 largely driven by the emergence of large-scale tectonic patterns rather than by changes in the relative  
229 importance of different tectonic environments.

230 Overall, the results demonstrate that historical earthquake catalogues contain a detectable but limited imprint  
231 of global tectonic structure. This imprint becomes apparent only when spatial information is analysed at  
232 sufficiently large scales and remains strongly modulated by tectonic setting. At smaller scales, historical  
233 seismicity is dominated by reporting artefacts, whereas at larger scales a partial convergence toward the  
234 instrumental spatial pattern is observed.

235

236

## 237 **6. Conclusions**

238 This study presents a quantitative, spatially explicit assessment of the consistency between global historical  
239 and early instrumental earthquake catalogues, with the aim of clarifying the extent to which historical  
240 seismicity reflects the underlying tectonic organization of earthquake occurrence. By adopting a scale-  
241 dependent approach based on spatial probability density fields and a robust coherence metric, we explicitly  
242 separate the influence of reporting bias from large-scale tectonic signal.

243 The results demonstrate that global historical earthquake catalogues contain a detectable but limited imprint  
244 of tectonic structure. Spatial coherence between historical (1600–1899) and instrumental (1900–1950,  $M \geq$   
245 5.5) seismicity is systematically low at small spatial scales and increases monotonically with increasing  
246 smoothing length, reaching moderate values only at regional to continental scales. This behaviour indicates



247 that historical earthquake locations are dominated by reporting artefacts at fine spatial resolution, while  
248 meaningful tectonic information emerges only after substantial spatial averaging.  
249 The decomposition of spatial coherence by tectonic domain highlights a strong control exerted by tectonic  
250 setting. Subduction zones provide the dominant contribution to historical–instrumental agreement at all scales,  
251 reflecting both their high seismic productivity and their long-standing documentation in historical sources.  
252 Continental collision belts and intraplate regions exhibit markedly lower coherence, underscoring the  
253 combined effects of distributed deformation, lower seismicity rates, and heterogeneous historical reporting.  
254 These findings demonstrate that the informational content of historical catalogues is not uniform across  
255 tectonic environments and must be interpreted accordingly.  
256 The pronounced inconsistencies in historical magnitude reporting further confirm that global historical  
257 earthquake catalogues are unsuitable for direct magnitude-based quantitative analyses. Consequently, spatial  
258 representations that do not rely on magnitude information provide a more robust basis for extracting physically  
259 meaningful information from historical data at the global scale.  
260 From a broader perspective, the results place an objective upper bound on the amount of tectonic information  
261 that can be recovered from global historical earthquake catalogues. While these datasets remain invaluable for  
262 regional studies and for documenting rare destructive events, their use in global-scale seismic hazard or  
263 tectonic analyses requires explicit consideration of spatial scale, tectonic context, and reporting bias.  
264 Approaches that neglect these factors risk overinterpreting patterns that are primarily controlled by non-  
265 geophysical processes.  
266 The framework developed here offers a transparent and reproducible methodology for evaluating the spatial  
267 reliability of historical seismicity and can be readily extended to other catalogues, time periods, or alternative  
268 spatial representations. More generally, this study contributes to a clearer and more quantitative understanding  
269 of the role and limitations of historical earthquake data in modern seismological research.

270

271

## 272 **Code availability**

273 All data processing, analysis, and figure generation were performed using custom Python scripts developed  
274 specifically for this study. The code implements the full workflow described in the Methods section, including



275 data download, quality control, spatial smoothing, computation of scale-dependent spatial coherence, and  
276 tectonic-domain decomposition. The scripts are available from the authors upon reasonable request.

277

278

#### 279 **Data availability**

280 This study is based exclusively on openly accessible earthquake catalogues. Historical seismicity data were  
281 obtained from the Global Historical Earthquake Catalogue (GHEC), the European Pre-Instrumental  
282 Earthquake Catalogue (EPICA), and the NOAA Significant Earthquake Database. Early instrumental  
283 earthquake data (1900–1950) were retrieved from the United States Geological Survey (USGS) earthquake  
284 catalogue. Tectonic domain definitions are based on the PB2002 global plate boundary model. All datasets  
285 used in this study are publicly available from the original data providers through their respective online  
286 repositories, as cited in the References section. No proprietary or restricted-access data were used.

287

288

#### 289 **Author contribution**

290 Antonino D’Alessandro conceptualized the study and designed the methodological framework. He compiled  
291 and curated the historical and instrumental earthquake datasets, developed the analysis code, and performed  
292 all data processing, statistical analyses, and figure generation. Antonino D’Alessandro interpreted the results  
293 and wrote the manuscript (Conceptualization, Methodology, Software, Data curation, Formal analysis,  
294 Visualization, Writing – original draft, Writing – review & editing)

295

296

#### 297 **Competing interests**

298 The author declare that he has no conflict of interest.

299

300

#### 301 **Disclaimer**



302 The views and interpretations expressed in this manuscript are those of the author and do not necessarily reflect  
303 the official policies or positions of the institutions with which the author is affiliated. The use of publicly  
304 available datasets does not imply endorsement by the data providers, who bear no responsibility for the  
305 analyses, interpretations, or conclusions presented in this study.

306

307

### 308 **Acknowledgements**

309 The author acknowledges the institutions and individuals who contributed, directly or indirectly, to this  
310 research. The study benefited from openly accessible earthquake catalogues and related resources provided by  
311 international and national data centers, including the Global Historical Earthquake Catalogue (GHEC), the  
312 European Pre-Instrumental Earthquake Catalogue (EPICA), the NOAA Significant Earthquake Database, and  
313 the United States Geological Survey (USGS). The author also acknowledges the availability of the PB2002  
314 global plate boundary model, which was essential for the tectonic-domain analyses.

315 This work made use of research infrastructure and computational resources available at the author's host  
316 institution, which supported data processing, analysis, and visualization. The author is grateful to colleagues  
317 for informal discussions that helped clarify methodological choices and interpret the results. Any remaining  
318 errors or omissions are the sole responsibility of the author.

319

320

### 321 **Financial support**

322 This research received no specific grant from any funding agency in the public, commercial, or not-for-profit  
323 sectors.

324

325

### 326 **7. References**

327 Albini, P., Musson, R. M. W., Rovida, A., Locati, M., Gomez Capera, A. A., and Stucchi, M.: The Global  
328 Historical Earthquake Catalogue (GHEC), *Ann. Geophys.*, **57**, S0435, <https://doi.org/10.4401/ag-6309>, 2014.

329



- 330 Albini, P., Rovida, A., Locati, M., Gomez Capera, A. A., and Stucchi, M.: The challenge of assessing seismic  
331 hazard in historical time, *Ann. Geophys.*, **58**, S0101, <https://doi.org/10.4401/ag-6623>, 2015.
- 332
- 333 Ambraseys, N. N.: *Earthquakes in the Mediterranean and Middle East*, Cambridge University Press,  
334 Cambridge, UK, 1989.
- 335
- 336 Ambraseys, N. N. and Jackson, J. A.: Faulting associated with historical and recent earthquakes in the Eastern  
337 Mediterranean region, *Geophys. J. Int.*, **133**, 390–406, <https://doi.org/10.1046/j.1365-246X.1998.00508.x>,  
338 1998.
- 339
- 340 Bakun, W. H. and Wentworth, C. M.: Estimating earthquake location and magnitude from seismic intensity  
341 data, *Bull. Seismol. Soc. Am.*, **87**, 1502–1521, 1997.
- 342
- 343 Bird, P.: An updated digital model of plate boundaries, *Geochem. Geophys. Geosyst.*, **4**(3), 1027,  
344 <https://doi.org/10.1029/2001GC000252>, 2003.
- 345
- 346 Gómez Capera, A. A., Albini, P., Stucchi, M., and Rovida, A.: Completeness of historical earthquake catalogues  
347 and its impact on seismic hazard, *Nat. Hazards Earth Syst. Sci.*, **20**, 1905–1923, [https://doi.org/10.5194/nhess-](https://doi.org/10.5194/nhess-20-1905-2020)  
348 [20-1905-2020](https://doi.org/10.5194/nhess-20-1905-2020), 2020.
- 349
- 350 Guidoboni, E. and Ebel, J.: *Earthquakes and Tsunamis in the Past: A Guide to Techniques in Historical*  
351 *Seismology*, Cambridge University Press, Cambridge, UK, 2009.
- 352
- 353 Guidoboni, E. and Stucchi, M.: Seismic history and seismic hazard, in: *Seismic Hazard Assessment*, Springer,  
354 Dordrecht, 1–24, [https://doi.org/10.1007/978-94-007-0418-4\\_1](https://doi.org/10.1007/978-94-007-0418-4_1), 2011.
- 355



356 Guidoboni, E., Ferrari, G., Gasperini, P., Mariotti, D., Comastri, A., Tarabusi, G., and Valensise, G.:  
357 CFTI4Med, Catalogue of Strong Earthquakes in Italy (461 B.C.–1997), INGV,  
358 <https://doi.org/10.6092/INGV.IT-CFTI4MED>, 2007.

359

360 Gutenberg, B. and Richter, C. F.: *Seismicity of the Earth and Associated Phenomena*, 2nd edn., Princeton  
361 University Press, Princeton, USA, 1954.

362

363 Helmstetter, A., Kagan, Y. Y., and Jackson, D. D.: Comparison of short-term and long-term earthquake forecast  
364 models for southern California, *Bull. Seismol. Soc. Am.*, **97**, 405–425, <https://doi.org/10.1785/0120060090>,  
365 2007.

366

367 Mignan, A. and Woessner, J.: Estimating the magnitude of completeness for earthquake catalogs, *J. Geophys.*  
368 *Res.*, **117**, B08302, <https://doi.org/10.1029/2011JB008888>, 2012.

369

370 NOAA: Significant Earthquake Database, National Centers for Environmental Information (NCEI),  
371 <https://www.ngdc.noaa.gov/hazel/view/hazards/earthquake/search>, 2023.

372

373 Rovida, A., Locati, M., Camassi, R., Lolli, B., and Gasperini, P.: The European Pre-Instrumental Earthquake  
374 Catalogue (EPICA), *Earth Syst. Sci. Data*, **14**, 5369–5389, <https://doi.org/10.5194/essd-14-5369-2022>, 2022.

375

376 Schorlemmer, D., Zechar, J. D., Werner, M. J., Field, E. H., Jackson, D. D., and Jordan, T. H.: First results of  
377 the Regional Earthquake Likelihood Models experiment, *Seismol. Res. Lett.*, **81**, 338–343,  
378 <https://doi.org/10.1785/gssrl.81.3.338>, 2010.

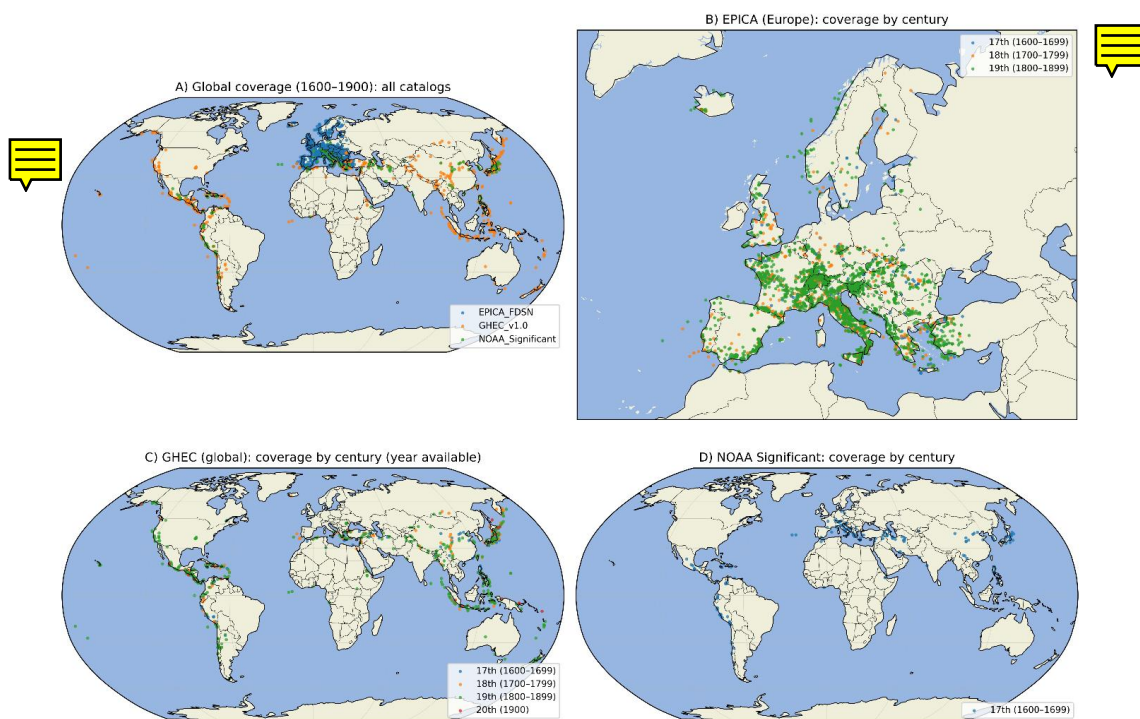
379

380 Stucchi, M., Albin, P., Rovida, A., Locati, M., Gomez Capera, A. A., and Viganò, D.: Assessing the  
381 completeness of macroseismic intensity data, *Bull. Seismol. Soc. Am.*, **103**, 243–256,  
382 <https://doi.org/10.1785/0120120022>, 2013.

383

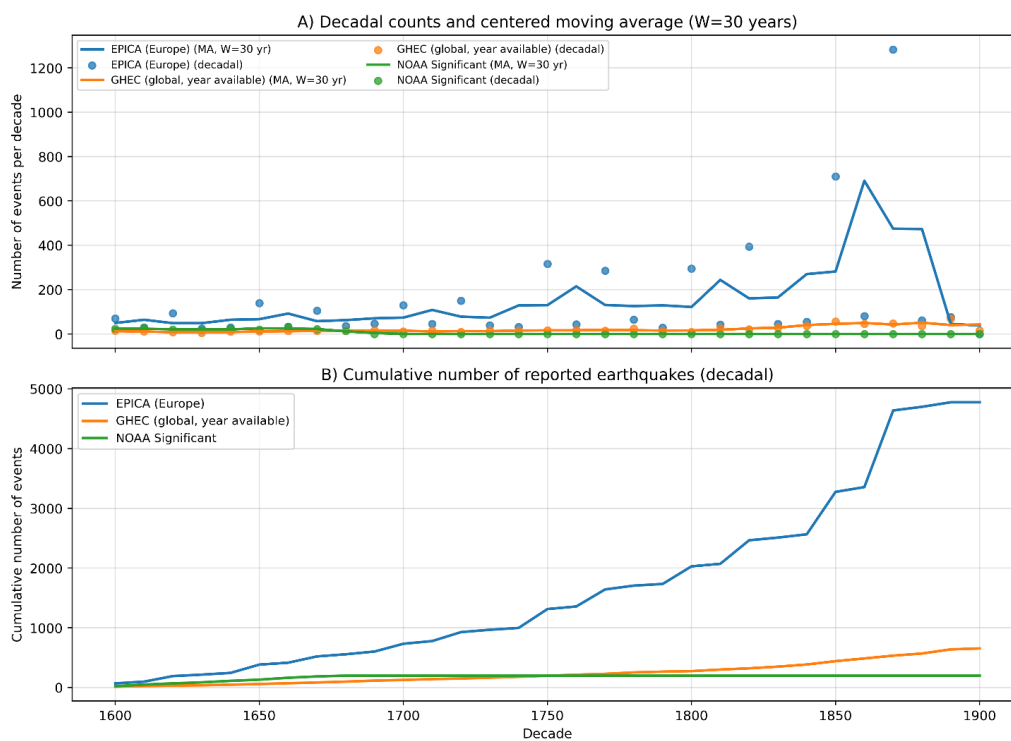


- 384 Utsu, T.: Statistical features of seismicity, in: *Int. Handb. Earthquake Eng. Seismol.*, Part A, Academic Press,  
385 719–732, [https://doi.org/10.1016/S0074-6142\(02\)80210-5](https://doi.org/10.1016/S0074-6142(02)80210-5), 2002.
- 386
- 387 Woessner, J., Laurentiu, D., Giardini, D., Crowley, H., Cotton, F., Grünthal, G., et al.: The 2014 European  
388 Seismic Hazard Model: key components and results, *Bull. Earthquake Eng.*, **13**, 3553–3596,  
389 <https://doi.org/10.1007/s10518-015-9795-1>, 2015.



390

391 **Figure 1.** Spatial coverage of historical earthquake catalogs considered in this study for the period 1600–  
392 1900. (A) Global distribution of events from the three datasets used: the Global Historical Earthquake  
393 Catalogue (GHEC v1.0), the European Pre-Instrumental Earthquake Catalogue (EPICA), and the NOAA  
394 Significant Earthquake Database. (B) Spatial distribution of EPICA events over Europe, coloured by century  
395 (17th, 18th, and 19th centuries). (C) Global distribution of GHEC events coloured by century, restricted to  
396 earthquakes for which a reliable occurrence year could be extracted from the original metadata. (D) Global  
397 distribution of NOAA Significant Earthquake Database events, coloured by century. All maps are shown using  
398 equal-area cartographic projections with coastlines for geographical reference. Differences in spatial  
399 coverage between panels primarily reflect intrinsic heterogeneities in historical documentation, catalog  
400 compilation criteria, and temporal completeness rather than true variations in seismicity.

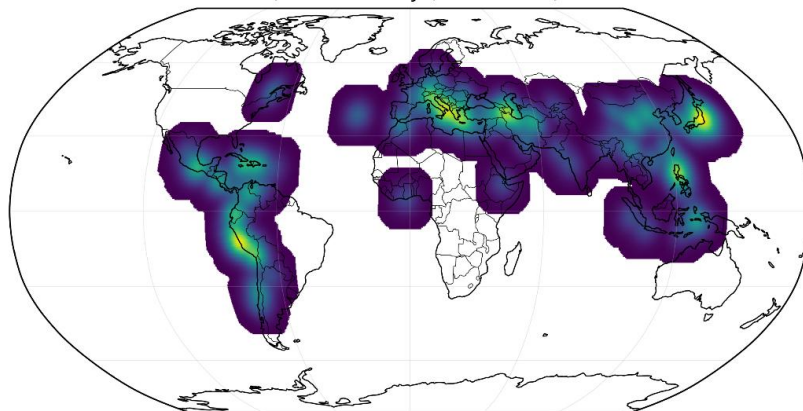


401

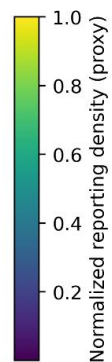
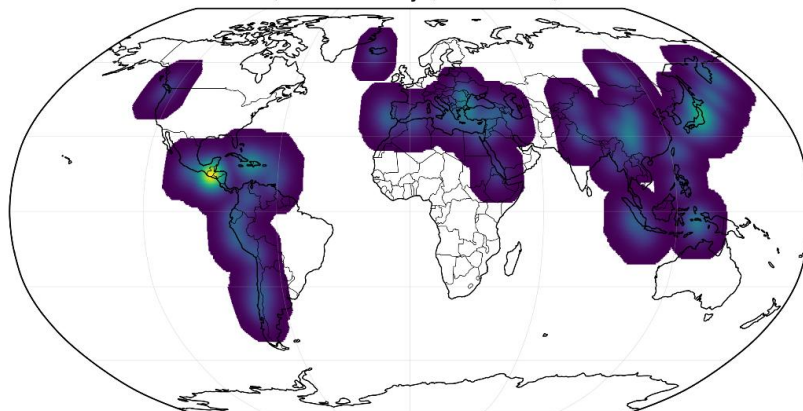
402 **Figure 2.** Temporal evolution of reported historical earthquakes in the study period 1600–1900. (A) Decadal  
403 counts  $N(t)$  for each catalog (symbols) together with a centred moving-average estimate  $\bar{N}_W(t)$  (solid lines;  
404 window length  $W = 30$  years) to stabilise temporal variability associated with historical dating uncertainty  
405 and heterogeneous reporting practices. (B) Cumulative number of reported earthquakes computed from the  
406 decadal counts, highlighting differences in long-term catalog growth. EPICA represents European pre-  
407 instrumental reporting, whereas GHEC and NOAA provide global compilations with different inclusion  
408 criteria; consequently, the curves primarily reflect temporal completeness and documentary/compilation  
409 effects rather than true changes in seismicity rates.



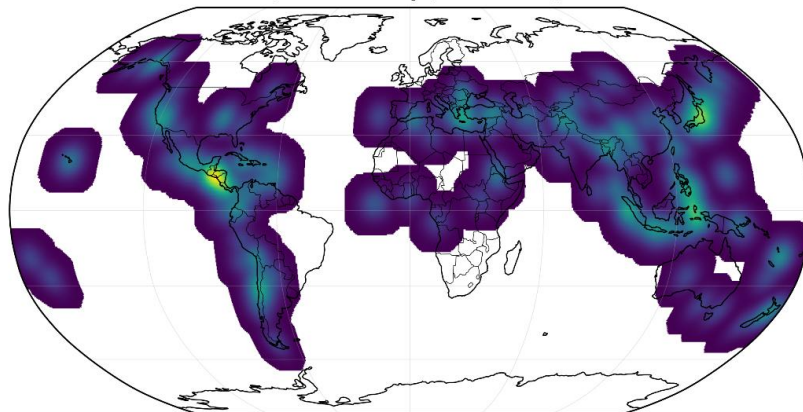
A) 17th century (1600-1699)



B) 18th century (1700-1799)



C) 19th century (1800-1899)



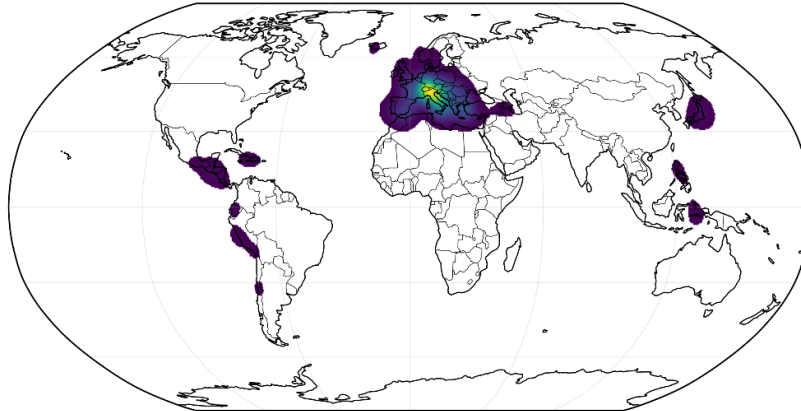


411 **Figure 3.** *Spatial patterns of historical earthquake reporting across centuries (1600–1899) derived from the*  
412 *global catalogs used in this study. Panels show the kernel-smoothed spatial density of reported earthquakes*  
413 *(coverage proxy) for (A) 1600–1699, (B) 1700–1799, and (C) 1800–1899, mapped on a global Robinson*  
414 *projection with coastlines and political borders for reference. For each century, the density field is computed*  
415 *from event locations on a regular geographic grid, smoothed to highlight coherent reporting clusters, and then*  
416 *normalized to the maximum value within that century to enable qualitative comparison of spatial coverage*  
417 *patterns. White areas indicate grid cells with negligible or null reporting density. The resulting maps primarily*  
418 *reflect the evolving spatial heterogeneity of historical documentation and catalog compilation, rather than*  
419 *physical variations in seismicity.*

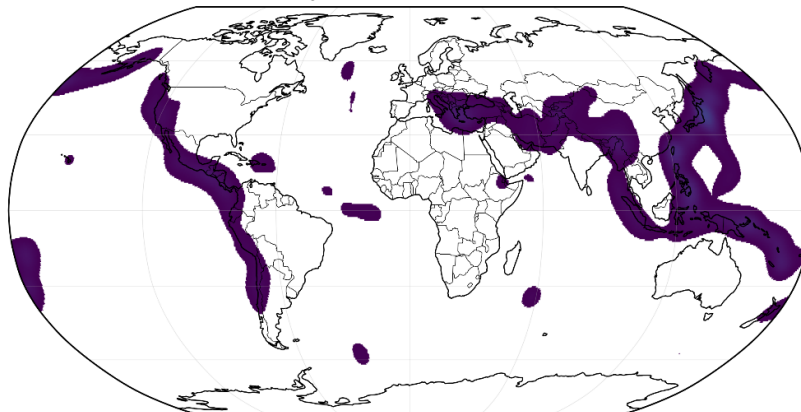


420

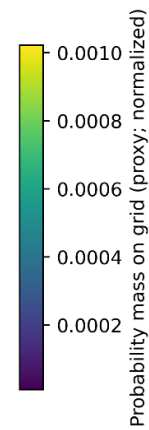
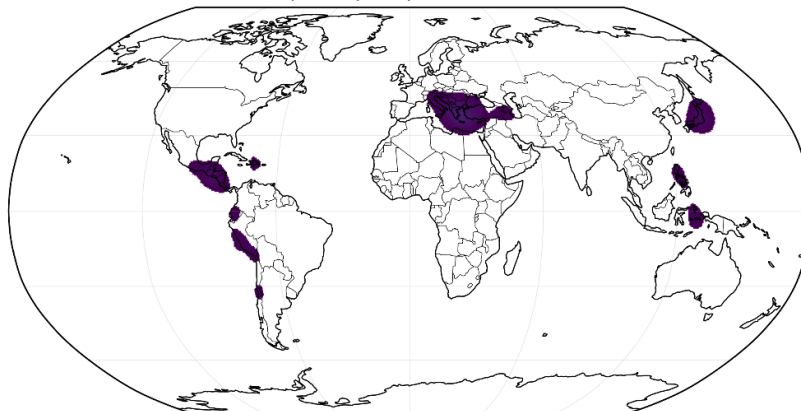
A) Historical reporting density (1600-1899), normalized



B) Instrumental density (USGS, 1900-1950,  $M \geq 5.5$ ), normalized



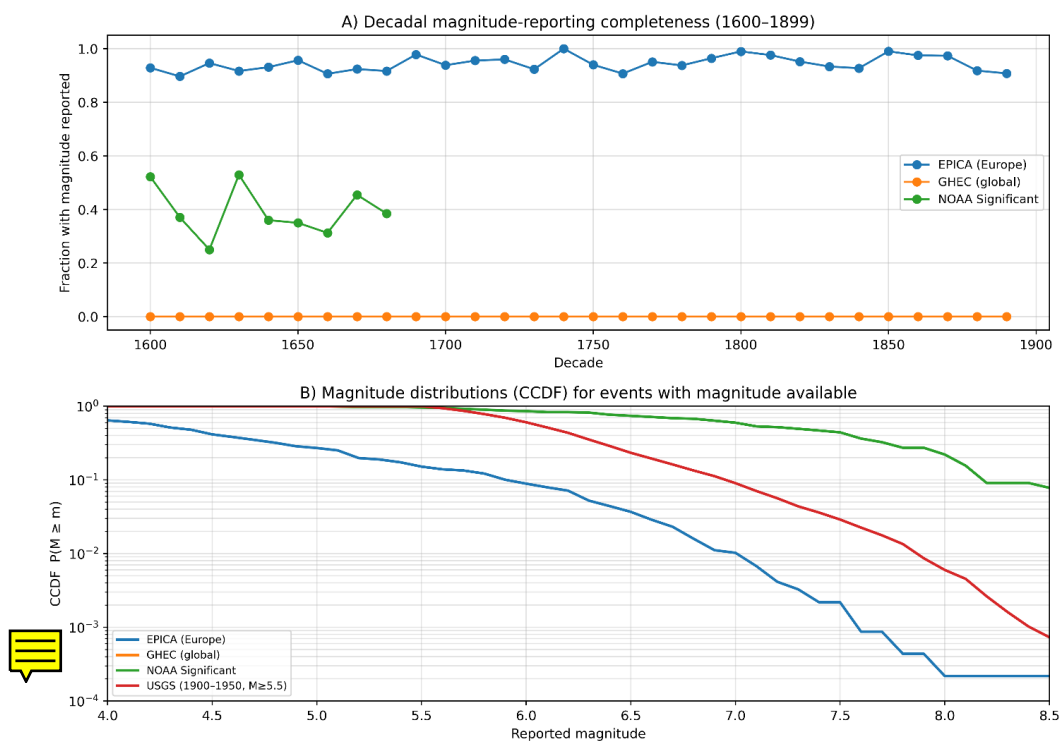
C) Local overlap  $\min(\rho_{hist}, \rho_{inst})$ ; coherence  $C = 0.188$



421

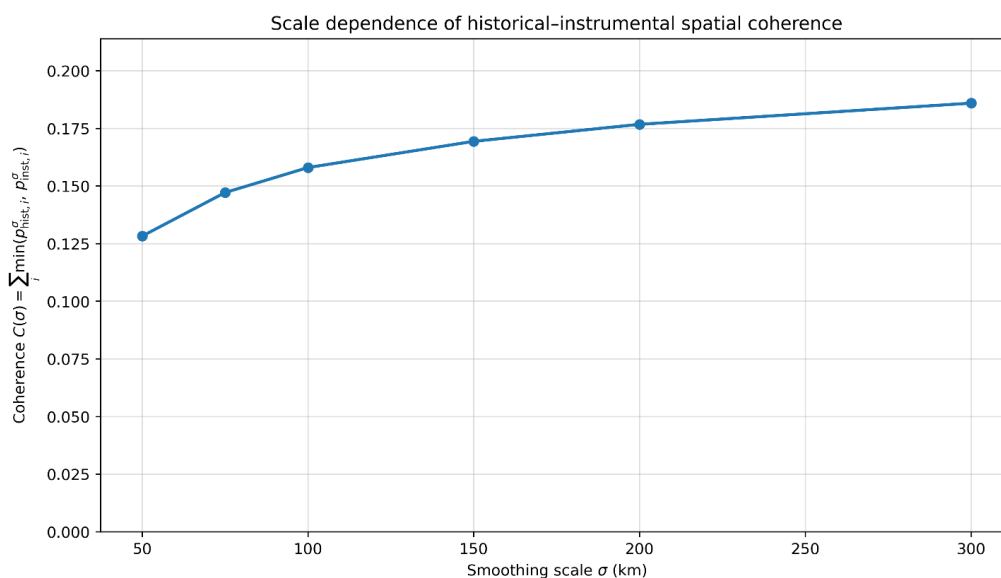


422 **Figure 4.** Comparison between historical and early instrumental earthquake reporting at the global scale. (A)  
423 Kernel-smoothed spatial density of reported historical earthquakes for the period 1600–1899, normalized to  
424 a probability mass function on a regular geographic grid. (B) Corresponding normalized spatial density for  
425 early instrumental seismicity derived from the USGS catalog for 1900–1950, considering events with  
426 magnitude  $M \geq 5.5$ . (C) Local spatial overlap between historical and instrumental reporting, computed as  
427  $\min(p_{\text{hist}}, p_{\text{inst}})$  at each grid cell. The scalar coherence index  $C = \sum_i \min(p_{\text{hist},i}, p_{\text{inst},i})$  quantifies the  
428 global spatial consistency between the two distributions. All maps are shown in Robinson projection with  
429 coastlines and political borders for reference; white areas indicate negligible probability mass. The limited  
430 overlap highlights the strong spatial bias of historical earthquake documentation relative to the early  
431 instrumental record, reflecting differences in population density, archival preservation, and catalog  
432 compilation rather than physical variations in seismicity.



433

434 **Figure 5.** Magnitude reporting bias in historical earthquake catalogs. (A) Decadal fraction of earthquakes  
 435 with a reported magnitude for the main historical catalogs considered in this study (EPICA, GHEC, NOAA  
 436 Significant) over the period 1600–1899. This metric quantifies the temporal completeness of magnitude  
 437 parametrization and highlights substantial differences in reporting practices across datasets. (B)  
 438 Complementary cumulative distribution functions (CCDFs) of reported magnitudes for events with available  
 439 magnitude estimates. Historical catalogs are compared with an early instrumental reference dataset (USGS,  
 440 1900–1950,  $M \geq 5.5$ ). Deviations from linear Gutenberg–Richter scaling and systematic offsets between  
 441 catalogs reflect reporting thresholds, selective documentation, and heterogeneous magnitude definitions,  
 442 emphasizing that historical magnitude distributions primarily represent reporting bias rather than underlying  
 443 seismicity.



444

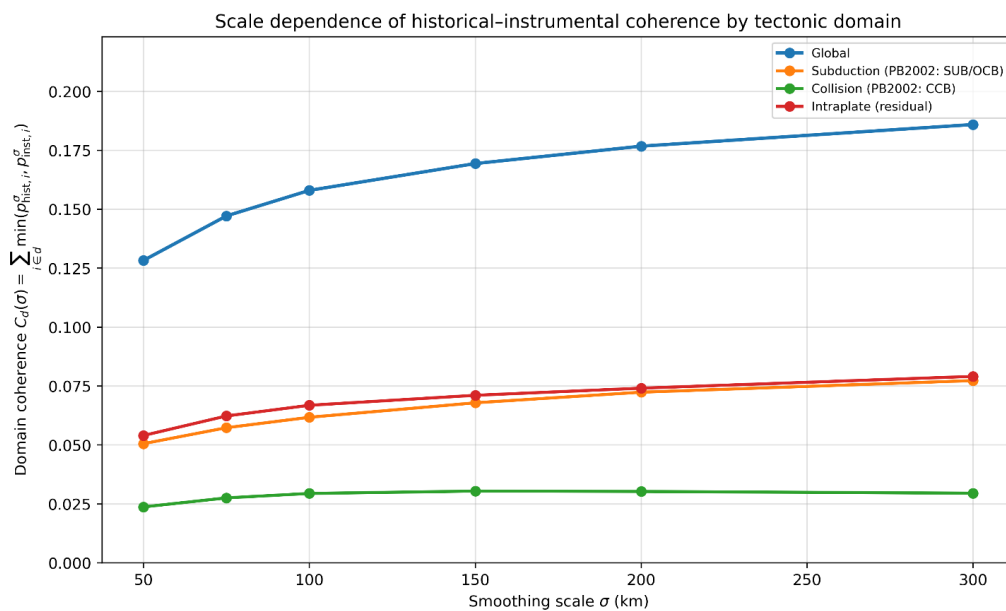
445 **Figure 6.** *Scale dependence of historical–instrumental spatial coherence. Scale dependence of the spatial*  
 446 *coherence  $C(\sigma)$  between historical (1600–1899) and instrumental (1900–1950,  $M \geq 5.5$ ) earthquake*  
 447 *reporting densities as a function of the Gaussian smoothing scale  $\sigma$ . For each value of  $\sigma$ , both catalogs are*  
 448 *converted into normalized spatial probability density functions on a common global grid using isotropic*  
 449 *Gaussian kernel smoothing. Coherence is quantified as*

450

$$452 \quad C(\sigma) = \int \min(p_{\text{hist}}^{\sigma}(\mathbf{x}), p_{\text{inst}}^{\sigma}(\mathbf{x})) d\mathbf{x},$$

451

453 *which measures the total spatial overlap between the two density fields. The coherence increases*  
 454 *monotonically with smoothing scale, indicating that agreement between historical and instrumental spatial*  
 455 *patterns emerges progressively as small-scale heterogeneities and reporting noise are averaged out. The*  
 456 *absence of saturation at small  $\sigma$  highlights the strong influence of local historical reporting biases, whereas*  
 457 *the gradual approach to an asymptotic value at larger scales reflects the recovery of a common large-scale*  
 458 *tectonic signal shared by both datasets.*



459

460 **Figure 7.** Scale dependence of historical–instrumental spatial coherence by tectonic domain. Scale  
 461 dependence of the spatial coherence  $C_d(\sigma)$  between historical (1600–1899) and instrumental (1900–1950,  
 462  $M \geq 5.5$ ) earthquake reporting densities, evaluated separately for major tectonic domains.  
 463 For each smoothing scale  $\sigma$ , coherence is defined as

464

466

$$C_d(\sigma) = \sum_{i \in d} \min(p_{hist,i}^\sigma, p_{inst,i}^\sigma),$$

465

467 where  $p_{hist,i}^\sigma$  and  $p_{inst,i}^\sigma$  are the normalized, Gaussian-smoothed reporting densities in grid cell  $i$ , and the sum  
 468 is restricted to cells belonging to domain  $d$ . Tectonic domains are defined using the PB2002 global plate-  
 469 boundary model (Bird, 2003): subduction zones (SUB/OCB), continental collision zones (CCB), and intraplate  
 470 regions (residual). Grid cells are assigned to a boundary-related domain within a 200 km influence distance  
 471 from the nearest PB2002 boundary segment; remaining cells are classified as intraplate. The global coherence  
 472 (blue curve) increases smoothly with  $\sigma$ , reflecting enhanced overlap at larger spatial scales. Subduction and  
 473 intraplate domains provide the dominant contributions to the total coherence across all scales, whereas  
 474 collision zones exhibit systematically lower and weakly scale-dependent coherence. The sum of the domain-



475 *specific coherences exactly reproduces the global coherence at each scale, confirming internal consistency of*  
476 *the decomposition.*