

Seismological Research Letters
Data-driven seismicity models based on Voronoi diagrams
--Manuscript Draft--

Manuscript Number:	SRL-D-24-00428R3
Full Title:	Data-driven seismicity models based on Voronoi diagrams
Article Type:	Article - Regular Section
Corresponding Author:	Guillaume DANIEL Electricité de France (EDF) Aix-en-Provence cedex 2, FRANCE
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	Electricité de France (EDF)
Corresponding Author's Secondary Institution:	
First Author:	Guillaume DANIEL
First Author Secondary Information:	
Order of Authors:	Guillaume DANIEL Pierre ARROUCAU
Order of Authors Secondary Information:	
Manuscript Region of Origin:	FRANCE
Abstract:	<p>The mapping of seismic activity is often achieved by means of smoothing or interpolation methods. Over the years, non-parametric kernel estimation methods have become popular to estimate seismicity intensity levels from catalog data. Their use is however not completely devoid of arbitrary decisions, especially concerning the selection of kernel type and the parameterization of kernel bandwidth. This inconvenience thus calls for an objective optimization of kernel shapes, or for a sound justification of choices made by end-users. Such a quest for objectivity is particularly expected in (probabilistic) seismic hazard analysis (PSHA) and in earthquake forecasting experiments.</p> <p>We propose an alternative approach where the spatial distribution of the seismicity rates is based on a Voronoi diagram of epicentral locations. In doing so, we avoid the use of any artificial smoothing operator that would be associated with a given kernel shape. Instead, we introduce an uncertainty-driven smoothing process, in which the degree of smoothing is controlled by the range of location and magnitude uncertainties in the catalog data. We propagate these uncertainties using Monte-Carlo random sampling, where multiple realizations of the original catalog data are used, each with randomly perturbed earthquake locations and magnitudes. The method relies on simple geometrical concepts. We demonstrate its potential by considering two representative use-cases: a smooth seismicity model to be used for PSHA in continental France, and a past earthquake forecasting experiment in California. Both applications produce maps of seismicity rates matching closely those obtained with adaptive smoothing kernels. This algorithm thus results in a reliable, un-biased, and more transparent smoothing process based on catalog uncertainties. Compared to other approaches, we rely on a smaller number of operator-driven decisions to control the degree of smoothing. This data-driven smoothing approach thus constitutes a valuable alternative to non-parametric kernel estimation methods for the future development of seismicity models.</p>
Suggested Reviewers:	John Adams Seismologist, Geological Survey of Canada john.adams@canada.ca Proven expertise in the production of source models for probabilistic seismic hazard assessment

	<p>Matthew C. Gerstenberger Principal Scientist, GNS Science, New Zealand m.gerstenberger@gns.cri.nz Proven expertise in developing seismicity models for earthquake forecasting and PSHA</p>
	<p>Andrea Llenos Research Geophysicist, U.S. Geological Survey allenos@usgs.gov Proven expertise in the development of seismicity rates models</p>
Opposed Reviewers:	

1 Data-driven seismicity models based on

2 Voronoi diagrams

3 Guillaume Daniel¹ and Pierre Arroucau¹

⁴ ¹EDF - Direction Technique, Aix-en-Provence, France

April 11, 2025

⁶ Corresponding author:

⁷ G. Daniel, EDF - Direction Technique; - 180 rue du lieutenant Parayre -
⁸ 13290 Aix-en-Provence, France.

9 Abstract

10 The mapping of seismic activity is often achieved by means of smoothing or
11 interpolation methods. Depending on the tectonic context and on the objec-
12 tives of a study, end-users can select among several alternative methods to
13 produce a reliable representation of seismicity rate spatial distribution. Over
14 the long term, non-parametric kernel estimation methods have become popu-

lar to estimate the seismicity rate intensity (density) directly from the catalog data. Their use is however not completely devoid of arbitrary decisions, especially concerning the selection of kernel type and the parameterization of kernel bandwidth. This inconvenience thus calls for an objective optimization of kernel shapes, or for a sound justification of choices made by end-users. Such a quest for objectivity is particularly expected in (probabilistic) seismic hazard analysis (PSHA) and in earthquake forecasting experiments.

We propose an alternative approach where the spatial distribution of the seismicity rates is based on a Voronoi diagram of epicentral locations. In doing so, we avoid the use of any artificial smoothing operator that would be associated with a given kernel shape. Instead, we introduce an uncertainty-driven smoothing process, in which the degree pf smoothing is controlled by the range of location and magnitude uncertainties in the catalog data. We propagate these uncertainties using Monte-Carlo random sampling, where multiple realizations of the original catalog data are used, each with randomly perturbed earthquake locations and magnitudes. The method relies on simple geometrical concepts. We demonstrate its potential by considering two representative use-cases: a smooth seismicity model to be used for PSHA in continental France, and a past earthquake forecasting experiment in California. Both applications produce maps of seismicity rates matching closely those obtained with adaptive smoothing kernels. This algorithm thus results in a reliable, un-biased, and more transparent smoothing process based on catalog uncertainties. Compared to other approaches, we rely on a smaller

38 number of operator-driven decisions to control the degree of smoothing, or
39 to optimize the parameterization of the smoothing kernel. This data-driven
40 smoothing approach thus constitutes a valuable alternative to non-parametric
41 kernel estimation methods for the future development of seismicity models.

42 1 Introduction

43 Seismologists commonly rely on earthquake catalog data to produce interpo-
44 lated maps of seismic activity rates for the construction of seismicity models.
45 These models, completed by additional information on the geological and
46 tectonic contexts in a region of interest, can be used to elaborate earthquake
47 forecasts to estimate future seismic hazard intensity levels (e.g., Gersten-
48 berger et al., 2020). Note that in doing so, one often implicitly suggests
49 that future hazard levels will be determined by the geographical distribution
50 of past activity patterns (Kagan and Jackson, 1991; Woo, 1996; Kafka and
51 Walcott, 1998).

52 Smooth seismicity models are typically built using non-parametric kernel
53 density estimation methods (e.g., Frankel, 1995). This approach requires to
54 select an *a priori* (and *ad hoc*) smoothing kernel, such as a 2-D isotropic
55 Gaussian, or a 2-D power-law kernel, or any other example adapted to the
56 study. In each pixel, the local earthquake rate is then estimated from counts
57 of earthquakes located nearby (i.e., within the pixel and beyond), where each
58 event contributes with a weight that decreases as a function of the increasing

59 distance from the kernel centre. This decay is governed by the kernel shape
60 and bandwidth. Non-parametric smoothing has become a valuable tool for
61 the objective interpolation of seismicity rates in areas with sparse historical
62 or instrumental records. It also permits to represent the spatial variability of
63 earthquake occurrence. While smooth models were originally developed to
64 offer an alternative to area source models in PSHA (e.g., Frankel, 1995; Woo,
65 1996), they are nowadays more frequently combined with fault-source models
66 to represent a spatially-variable background seismic activity (e.g., Valentini
67 et al., 2017)

68 Several implementations for the smooth mapping of seismicity rates are
69 reported in the literature. They are associated with different strategies to
70 modulate the width of spatial kernels. The earliest use of such a smoothing
71 kernel was proposed by Frankel (1995). It was used to update seismic hazard
72 maps for the 1997 edition of the *NEHRP Recommended Provisions for Seis-*
73 *mic Regulations for New Buildings and Other Structures* (Holmes, 2000). In
74 their final model, seismic hazard levels were estimated from the mean of 4 al-
75 ternative source models, 2 of which being derived from the spatial smoothing
76 of seismicity using different kernels. These authors used constant-width 2-D
77 isotropic Gaussian kernels. At that time, their aim was to propose a more
78 objective alternative to the traditional approach used in seismic hazard as-
79 sessment, which was originally founded on area source models (Cornell, 1968;
80 Mc Guire, 1976). In doing so, they aimed to avoid expert judgment in the
81 definition of seismic source zones, where the assessment of tectonic features

82 is often based on the combined interpretation of seismological, geological and
83 geophysical observations.

84 Woo (1996) then proposed alternative formulations for the spatial kernel,
85 grounded on the concepts of fractal geometry and self-organized criticality.
86 In his study, kernel bandwidth scaled according to magnitude. His idea was
87 to provide a smoothing operator that accounts for long-range earthquake in-
88 teractions typical of many critical phenomena, and to offer an alternative
89 to the *magnitude-independent* area source definition process. In addition, he
90 emphasized that the smoothing operation avoids potential algebraic inconsis-
91 tencies that can occur when b -values are estimated from earthquake popula-
92 tions located in distinct area sources. He proposed two possible formulations
93 for the spatial kernel: an infinite-range power-law decay (Vere-Jones, 1992),
94 and a finite-range kernel based on the fractal dimension of epicenters. Cao
95 et al. (1996) proposed a slightly different kernel formulation, with a $\sim 1/r^\mu$
96 power-law decay where μ is a linear function of magnitude. These authors
97 used this latter kernel for a study in Southern California where seismic haz-
98 ard levels were estimated using a smoothed background seismicity model,
99 which was itself obtained from a declustered catalogue of earthquakes.

100 However, one of the major limitations of the aforementioned kernels is
101 when seismicity rates are mis-estimated in areas where earthquake epicen-
102 ters present sharp spatially-variable local features, such as sharp boundaries
103 between areas of high and low seismic activity. Under- and over-smoothing
104 can also occur as a result of an inappropriate kernel bandwidth selection (e.g.,

105 using a constant bandwidth over a large area...). Adaptive kernel methods
106 were developed to circumvent these issues, by involving kernels with spatially-
107 varying bandwidths. Such kernels have proven to be more appropriate to re-
108 solve local features of seismicity (Stock and Smith, 2002; Helmstetter et al.,
109 2007).

110 Adaptive smoothing methods have been implemented following different
111 approaches to define local kernel bandwidths. Stock and Smith (2002) and
112 Helmstetter (2006) use a two-step process where the local bandwidth for each
113 earthquake is based on the local rate obtained from a preliminary constant-
114 bandwidth smoothing operation. Helmstetter et al. (2007) propose a sim-
115 pler approach where the local bandwidth is given by the horizontal distance
116 between each earthquake and its n^{th} nearest neighbor. In doing so, local
117 bandwidths decrease when local seismicity rates increase, and conversely. In
118 their application, the number n is an adjustable parameter that is optimized
119 to increase the likelihood of their earthquake forecasting model.

120 Although more seldom, anisotropic smoothing kernels have also been con-
121 sidered in the literature to account for the presumed location of epicenters
122 along active faults. A rare example is the kernel proposed by Kagan and
123 Jackson (1994), and re-used by Jackson and Kagan (1999) for forecasting
124 purposes. Their approach consists in defining an anisotropic kernel shape
125 elongated in the main fault direction of each earthquake, based on the nodal
126 plane orientation deduced from the focal mechanism, when available (other-
127 wise, isotropy is assumed).

128 It is also worth mentioning more complex techniques, such as the combi-
129 nation of multiple kernels realized by Taroni and Akinci (2021) to improve
130 earthquake forecasts. Ogata et al. (2003) also proposed a model-based tech-
131 nique to better account for skewed spatial earthquake clustering patterns in
132 Japan, where the background rate of earthquakes presents a significant spa-
133 tial variability. They developed an objective spatial partitioning of the area
134 which relied on a Delaunay triangulation. Earthquake epicentral locations
135 were used as nodes for the tessellation, and thus constitute the junctions of
136 neighboring triangles in this tessellation. Each triangle is associated with a
137 collection of constant values used to define spatially-variable 2-D functions
138 for each of the 5 parameters of the seismicity rate model. In the end, these
139 spatially-variable functions were used as a prior in a Bayesian framework for
140 the estimation of the background seismicity rate.

141 All of these smoothing approaches rely on a variable number of operator-
142 driven decisions related to the selection of kernel shape, to the number of
143 kernels involved, to parameterization of its (their) bandwidth(s), etc... In
144 addition, even if an optimization process is put in place to select the best
145 among multiple kernels, one also has to define a testing strategy, to select
146 a reference catalog, and to define one, or several, reference time periods for
147 the calibration and testing phases. The optimal parametrization tends to be
148 very sensitive to each of these aspects, and it has also been documented that
149 a single solution is rarely confirmed if one changes any component of this
150 optimization procedure (e.g., Werner et al., 2011; Moschetti, 2015).

151 Our objective is here to introduce an alternative non-parametric, kernel-
152 free, approach for the spatial estimation of seismic intensity rates. This work
153 is motivated by the need for a more objective and transparent procedure,
154 requiring a reduced number of operator decisions, and where the degree of
155 smoothing is controlled by catalog uncertainties. This data-driven approach
156 is grounded on a Voronoi tessellation of earthquake epicenters, with several
157 iterations to propagate uncertainties using Monte-Carlo random sampling.
158 We apply this method to compare with two past case-studies: a smooth
159 seismicity model for a PSHA in France and an earthquake forecasting ex-
160 periment in California. Our results confirm the suitability of the method to
161 deliver objective maps of seismic intensity rates for any region. For the two
162 case-studies, we obtain spatial patterns of seismicity matching closely those
163 produced by adaptive smoothing kernels. This technique gives also some
164 statistical sense to the degree of smoothing implemented over the map, as it
165 directly reflects the uncertainties reported in the catalog data. Recalling Woo
166 (1996) who recognized that kernel smoothing techniques are not devoid of
167 subjectivity, we believe that our approach can improve existing methods, in
168 that it removes the need for a prior and subjective definition of the smooth-
169 ing kernel. In addition, the proposed approach enables to account explicitly
170 for catalogue uncertainties.

₁₇₁ **2 Method**

₁₇₂ Consider a catalog \mathcal{C} that contains N_e earthquakes in a spatial domain \mathcal{D} over
₁₇₃ a given time period. Our aim is to produce an objective, spatially-continuous,
₁₇₄ estimate of the earthquake activity rate over \mathcal{D} based on this catalog. For
₁₇₅ this purpose, we partition the spatial domain \mathcal{D} into a tessellation of N_e
₁₇₆ Voronoi cells (Dirichlet, 1850; Voronoi, 1908). According to the construction
₁₇₇ principle of such diagrams, each Voronoi cell is built around a single point (or
₁₇₈ 'germ', or 'seed'...), in such a way that it encompasses the partition of space
₁₇₉ that is located closer to this germ than to any other in the domain. Voronoi
₁₈₀ cells thus constitute, by definition, the "nearest-neighbor" areas around each
₁₈₁ germ.

₁₈₂ Each earthquake in \mathcal{C} is characterized by an occurrence time t_k , an epicen-
₁₈₃ tral location \mathbf{x}_k , a magnitude m_k and uncertainties related to these quantities,
₁₈₄ for $k = 1, \dots, N_e$. m_0 is the minimum magnitude considered in the catalog for
₁₈₅ this analysis. We define N_m magnitude bins of width dm , bounded between
₁₈₆ m_0 and $m_0 + N_m \cdot dm$. Each bin contains events with magnitudes in the
₁₈₇ $[m_0 + dm \cdot (j - 1); m_0 + dm \cdot j]$ interval, for $j = 1, \dots, N_m$.

₁₈₈ Each interval can also be associated with a specific completeness epoch
₁₈₉ of duration T_j , during which the collection of events in \mathcal{C} is assumed to
₁₉₀ be exhaustive. Note that this bin-wise description of catalog completeness
₁₉₁ makes it easier to work with historical catalogs, which are often described
₁₉₂ using such magnitude-dependent completeness thresholds (Stepp, 1972). In

193 the following, all earthquakes with $m < m_0$, or events which occurred before
194 their completeness period, are excluded from our analysis.

195

196 In the following sections we describe a two-step approach for the eval-
197 uation of seismicity rates over \mathcal{D} , expressed in terms of a spatially-variable
198 a -value in the Gutenberg and Richter (1944) formula. First, we evaluate
199 seismicity rates over a regular grid of pixels, where the value in each pixel is
200 obtained from the contributions of a collection of Voronoi cells intersecting its
201 area. One grid is created for each magnitude interval, as detailed in section
202 2.1. Then, in a second step, we combine estimated rates for all N_m magnitude
203 intervals, so as to reconstruct a frequency-magnitude distribution (FMD) for
204 every pixel. This FMD is used in turn to estimate the two parameters of the
205 Gutenberg-Richter formula in each cell (namely, a and b), using a maximum-
206 likelihood algorithm (e.g., Weichert, 1980; Dutfoy, 2020), see section 2.3.
207 This process is illustrated here by a synthetic example, where event counts
208 per cell and event magnitudes are randomly drawn from a Gutenberg-Richter
209 model with spatially variable a and b parameters. Given the local event rate,
210 epicentral locations are uniformly distributed within each pixel. Parameter
211 values were sampled in space according to a pseudo-bivariate normal dis-
212 tribution, configured with an almost-constant b -value over the domain and
213 variable a -values in each cell. Our simulation resulted in a synthetic cata-
214 logue of 50 390 events, with magnitudes ranging from 2.0 to 6.8. Epicentral
215 locations for events with magnitude included in the [3.0; 3.5] interval are

216 presented on the top panel of Figure 1.

217 2.1 Seismicity rate maps

218 For every magnitude interval j , we produce a diagram which is composed
219 of N_j Voronoi polygons surrounding each epicenter with magnitude ranging
220 within this interval. As an illustrative example, we present such a diagram
221 in the middle panel of Figure 1 for the [3.0; 3.5] magnitude interval of our
222 synthetic catalogue. Given that each Voronoi polygon contains only one
223 epicenter, the local seismicity rate in each polygon λ_{pj} (where $p = 1, \dots, N_j$) is
224 equal to the inverse of the product of the polygon area A_p by the completeness
225 period for this interval (e.g., Ord, 1978; Nicholson et al., 2000; Barr and
226 Schoenberg, 2010):

$$\lambda_{pj} = \frac{1}{A_p \cdot T_j}$$

Smooth seismicity maps usually consist of a set of values computed over a regular mesh of N_c pixels. Conversely, the natural spatial clustering of earthquake epicenters makes the geometries and distribution of Voronoi polygons very irregular, as shown in Figure 1. This implies that each pixel of the mesh intersects at least one polygon, and possibly more. For each polygon p , we denote A_{pi} the fraction of its area intersecting pixel i . In each pixel i the rate of earthquakes in the j -th magnitude interval is thus obtained by summing

all individual portions of polygons that intersect its (rectangular) area:

$$\lambda_{ij} = \sum_p \lambda_{pj} \left(\frac{A_{pi}}{A_p} \right)$$

227 A gridded seismic density map is presented in the bottom panel of Figure
228 1 for illustrative purposes. This result is based on the collection of Voronoi
229 polygons presented in the middle panel. In this example, one can distinguish
230 the slight smoothing of spatial patterns related to the grid discretization
231 step, in the transition from the central to the bottom panels of Figure 1,
232 in particular for the most active cells. The feature can be amplified, or
233 diminished, by setting respectively a larger, or a smaller grid spacing, and is
234 illustrated on Figure A in the Supplementary Material section. The selection
235 of the optimal grid spacing should thus represent a compromise between
236 the highest expected spatial resolution and the most reasonably acceptable
237 level of uncertainty in density estimates. Such interaction between the grid
238 discretization step and the variability of a and b estimates is documented in
239 the Supplementary Material section, see Figures B and C. Using a simple
240 synthetic example based on an area with a homogeneous seismicity rate, we
241 could characterize that, independently of the grid spacing, the average pixel-
242 wise estimates of a and b parameters are robust and unbiased, but that the
243 variability of these estimates increases significantly when the grid spacing
244 diminishes. Once obtained such earthquake density maps, the conversion to
245 seismicity rates, as illustrated in Figure 1 (middle and bottom panels), is

246 made by dividing densities by the observation period (which must be shorter
247 or equal to the completeness epoch).

248 **2.2 Consideration of uncertainties**

249 Uncertainties on earthquake attributes (occurrence time, location, magni-
250 tude) can be propagated in the process using Monte-Carlo random sampling.
251 This process is illustrated in Figure 2. Events locations and magnitudes
252 $(\mathbf{x}_k, m_k)_{k=1, \dots, N_e}$ are randomly drawn from their respective domain of uncer-
253 tainty, here modelled by independent univariate distributions. Note that the
254 sampling of origin times does not need to be necessarily implemented, as
255 such uncertainties are most certainly negligible compared to those on the
256 durations of completeness periods. For each realization of the Monte Carlo
257 process, we first draw N_e random magnitudes from normal distributions cen-
258 tered on $m_k - \sigma_k^2 \beta / 2$ for each earthquake, with $\beta = b \cdot \ln(10)$. This corrects
259 for the positive bias in FMD introduced by the symmetrical random per-
260 turbation of event magnitudes (Tinti and Mulargia, 1985; Rhoades, 1996;
261 Dutfoy, 2023), see also Figure D of the Supplementary Material. Then, for
262 each bin in magnitude, the number of events is perturbed according to a
263 Poisson variable, leading to a random decimation, or replication, of events
264 included in the bin, depending on the drawn event count. Event locations are
265 also randomly perturbed based on their respective ellipsoids of uncertainty.

266 Each perturbed catalog thus produces a Voronoi diagram with slightly
267 modified polygons shapes. This, in turn, affects the distribution of seismic

268 density and seismicity rates in each pixel, see Figure 2. Beyond a sufficient
269 number of perturbations, the average spatial pattern of seismicity rates re-
270 mains constant over the domain \mathcal{D} . Note however that the optimal number of
271 realizations, N , required for convergence, depends strongly on the properties
272 and uncertainties in the original catalog. As such, it needs to be carefully
273 assessed by end-users for each study. As a result of this process, while the
274 large-scale geographical patterns of the seismicity rate are preserved by the
275 Monte-Carlo process, the smaller-scale features related to sharp contrasts in
276 seismic activity will undergo a data-driven smoothing operator, which will
277 be more pronounced around earthquakes with uncertain location or with
278 uncertain magnitude, see Figure 2. In addition, the level of smoothing is
279 affected by the number of sources of uncertainty. This effect is presented
280 in Figure 3, where an enhanced smoothing can be observed when one intro-
281 duces the perturbation of catalogue magnitudes (bottom panel) in addition
282 to the perturbation of epicentral locations (middle panel). This amounts to
283 a data-driven smoothing process.

284 In the end, this approach provides users with a complete distribution of
285 the seismic density or seismicity rate in each pixel, which can then be further
286 processed to produce maps of average quantities and associated standard de-
287 viations. It is worth mentioning that consistency between obtained seismicity
288 models and real catalogue of earthquakes can be assessed and quantified us-
289 ing a variety of goodness-of-fit tests, such as likelihood tests (e.g., Savran
290 et al., 2022) or residual analysis (Clemens et al., 2012; Bray et al., 2014).

291 **2.3 Modelling Frequency-Magnitude Distribution**

292 Beyond seismic densities, or seismicity rates, it can be useful and convenient
293 to express seismic activity levels in terms of the properties of an FMD. For
294 example, one may summarize results by presenting maps of the a and b
295 parameters in the Gutenberg-Richter (GR) formula (Gutenberg and Richter,
296 1944). This is often used to describe activity rates with smooth seismicity
297 models in PSHA, for instance. However, there is no restriction here and any
298 model could be used to parameterize FMDs, in principle.

299 The main outcome of the proposed methodology, as explained above, is
300 a collection of magnitude bin-wise estimates of densities/rates for each pixel.
301 FMDs are reconstructed in each pixel by repeating the steps described in
302 section 2.1 and 2.2 for each of the N_m magnitude intervals. This process is
303 illustrated on Figure 4.

304 The FMD modelling is done in a distinct and subsequent step. Parameter
305 estimation procedures can differ from one application to the other. If both
306 parameters a and b are sought after, one can use the maximum-likelihood
307 algorithm of Weichert (1980) in each pixel (see also Dutfoy, 2020, for a com-
308 plementary work on the distribution of the estimators). Depending on the
309 number of events per cell, it may be necessary to introduce *a priori* con-
310 straints on b -values. This can be implemented using a Bayesian framework,
311 for instance. Or by fixing arbitrarily its value, based on external sources
312 of information. In this latter case, the resolution is straightforward, as the
313 optimal a parameter can be derived analytically. In the next sections, we

314 present two applications where parameters of the GR formula are estimated
315 under varying hypotheses. For the France case-study (see section 3.1) we
316 consider domain-wise constant b -values, while for the California case-study
317 (see section 3.2) we use a constant b -value of 1.0 over the whole area. These
318 choices are made to mimic, as closely as possible (except for the smooth-
319 ing algorithm), the implementation of the original authors (i.e., Helmstetter
320 et al., 2007; Drouet et al., 2020). They are not meant to be recommended
321 for other applications.

322 3 An alternative to smoothing kernels

323 We illustrate the capabilities of the method by reproducing two past studies,
324 which both made use of (adaptive) smoothing kernels for different purposes.
325 The first use-case is taken from Drouet et al. (2020) and consists in the
326 construction of a smooth seismicity model for a probabilistic seismic hazard
327 assessment (PSHA, see Cornell, 1968; Mc Guire, 1976) in France, see section
328 3.1. Then, we introduce a second use-case in section 3.2, based on Helmstet-
329 ter et al. (2007), where we use Voronoi diagrams to compute earthquake
330 density maps for California, in the framework of an earthquake forecasting
331 testing exercise. These remakes are not meant to challenge results, nor to
332 update conclusions made by these authors. Our aim is rather to provide the
333 reader with a sensitivity study on the choice of the smoothing method, when
334 our method is used in replacement of smoothing kernels (under strict similar

335 working hypotheses and input conditions).

336 3.1 Smooth seismicity model for France

337 In their original PSHA study for mainland France, Drouet et al. (2020) in-
338 cluded a smooth seismicity model built using a 2-D isotropic Gaussian adap-
339 tive kernel. For each event, the kernel bandwidth was set equal to the distance
340 to its 10th nearest-neighbor. These authors also tested different kernel con-
341 figurations (i.e., adaptive vs. magnitude-dependent, different minimum M_W ,
342 different bandwidth parameterizations), without any significant variability in
343 the outcomes.

344

345 Here, we implement the approach described in section 2 to produce an
346 alternative smooth model based on Voronoi diagrams. Earthquake epicenters
347 are taken from the declustered earthquake catalogue of Drouet et al. (2020)
348 which covers mainland France over the AD 463 - 2016 period, see Figure
349 5c. Their catalogue is based on the FCAT-17 catalogue (Baumont et al.,
350 2018; Manchuel et al., 2018; Traversa et al., 2018), which was was extended
351 geographically beyond the original 20 km buffer around the French borders.
352 This extension was realized using the European SHEEC catalogue (Stucchi
353 et al., 2013) and events published in the CEA-LDG bulletins during the 2010-
354 2016 period. The catalog was declustered using the algorithm of Gardner and
355 Knopoff (1974), with the spatio-temporal window defined by Burkhard and
356 Grünthal (2009).

357 In a first step, we compute earthquake density maps for 7 magnitude bins
358 between 3.0 and 6.5, with a width of 0.5 magnitude unit. These bins were se-
359 lected based on the different completeness periods identified by Drouet et al.
360 (2020). Uncertainties are propagated by conducting a Monte-Carlo random
361 sampling of magnitudes and event locations (200 realizations). Then, we
362 characterize recurrence relationships over the area by adjusting parameters
363 of the Gutenberg-Richter relationship in each pixel, as described in Figure 4.
364 We re-used the constant b -values in each of the macro-domains as defined by
365 Drouet et al. (2020). This spatial distribution of b -values for France is pre-
366 sented in Figure E of the Supplementary Material. In the end, we obtain a
367 map of (normalized) a -values representing the spatial distribution of the seis-
368 micity rate over the area, see Figure 5a. These values have been obtained by
369 combining a maximum-likelihood technique (Weichert, 1980; Dutfoy, 2020)
370 with a strong prior on b . For the interested reader, note that we also present
371 an alternative parameterization of the (a , b) estimation problem in the Sup-
372 plementary Material, where no *a priori* constraint is set on b -values, see
373 Figure F.

374 The resulting smooth model is consistent with the location of seismic ac-
375 tivity in mainland France. Areas with higher activity rates can be identified
376 in the Alps region (at the French-Italian border), in the Pyrénées mountains
377 and also to a lesser extent in the West, in the southern part of the Armorican
378 Massif. Areas of lower activity encompass mainly the Paris basin and the
379 South-Western region of mainland France. Interestingly, we can also notice

380 finer details in the spatial distribution of the most active regions, compared
381 to the less active areas. This observation is fully consistent, and expected,
382 due to the use of Voronoi polygons which are self-adaptive features.

383

384 For comparison purposes, the original smooth seismicity model of Drouet
385 et al. (2020) is presented on Figure 5b. Normalized differences between with
386 our study are presented on 5d. Overall, both maps are relatively similar, with
387 a good agreement on the locations of the least and most seismically active
388 regions at a large-scale (e.g., Pyrénées, Alps). Both approaches produce
389 similarly smooth patterns at large and intermediate scales. However, patterns
390 deviate at a finer scale due to a tendency to spread higher-activity patches
391 in the method used by Drouet et al. (2020). The map of relative differences
392 also confirms the tendency of our method to produce more localized, higher
393 seismic levels in active areas, and smaller, smoother, levels in the lower active
394 areas (e.g., Paris basin and North-Western region, South-Western mainland
395 region). We quantified this observation by computing the value of the Gini
396 coefficient (Gini, 1936) for these two models. This coefficient is a measure of
397 statistical dispersion and produces values in the [0; 1] interval. A coefficient
398 value of 0 denotes perfect equality among samples of the population, while
399 a value of 1 reflects a situation of maximal inequality. The Gini coefficient
400 value for the the Voronoi-based source model is 0.052, while the coefficient for
401 the Drouet et al. (2020) model is 0.032. This quantifies a higher inequality
402 between cells of the former model, and by extension, confirms our observation

403 of more localized seismic patches in active areas.

404 **3.2 Earthquake forecast in California**

405 Since the end of the 1990's, several studies have made use of smoothing ker-
406 nels to produce seismicity rate maps for the purpose of earthquake forecasting
407 (e.g., Kagan and Jackson, 1994; Helmstetter, 2006; Helmstetter et al., 2007;
408 Helmstetter and Werner, 2012; Taroni and Akinci, 2021, ...). The rationale
409 being that future events are more likely to occur in areas of past and recent
410 seismic activity (Kagan and Jackson, 1991).

411 In this section, our aim is to assess how the forecast obtained using the
412 method described above (see section 2) would compare to a previous study
413 carried out with smoothing kernels.

414 For this purpose, we replay the emblematic high-resolution, time-independent
415 earthquake forecast experiment of Helmstetter et al. (2007) in California.
416 Their study was realized in the framework of the Working Group of the Re-
417 gional Earthquake Likelihood Models (RELM) forecast test (Kagan et al.,
418 2003; Schorlemmer et al., 2007). It proposed a model-independent long-term
419 forecast based on a smooth map of declustered seismic activity in the area.
420 To our knowledge, it was also the first study to use an adaptive kernel for
421 smoothing purposes, i.e., with a kernel bandwidth that scales with respect to
422 the distance to its n^{th} nearest-neighbor. Note that alternative parameteriza-
423 tions for adaptive kernels have been proposed in the literature, often with a
424 more explicit relationship between the bandwidth and the local earthquake

425 rate in each pixel (e.g., Stock and Smith, 2002; Helmstetter and Werner,
 426 2012). Here, for the sake of simplicity, we only compare our results with
 427 those obtained using a power-law nearest-neighbor adaptive kernel. How-
 428 ever, we recognize that a more systematic study of every adaptive kernel
 429 category would potentially shed more light on strengths and weaknesses of
 430 each approach.

431 We re-use the same input dataset, input parameters and working hy-
 432 potheses, except for the method involved in the computation of density
 433 maps. The input catalogue of earthquakes with $M \geq 2$ for California covers
 434 a period extending from Jan 1st, 1981 to Aug 23rd, 2005, and was declus-
 435 tered using Reasenberg's (1985) algorithm. We consider several magnitude
 436 bins, and for each bin we produce a gridded count of events with magnitude
 437 $M_i \leq M \leq 8.0$, where M_i varies from 2.0 to 7.9, in steps of 0.1. Note that we
 438 rely upon estimates of magnitude of completeness $M_c(\vec{r})$ computed by these
 439 authors over the RELM area (see Figure 2 in their manuscript). These values
 440 are used to extrapolate the count of $M \geq 2.0$ events in each pixel using a
 441 M_c -dependent correction factor (based on equation 5 in their article):

$$N_{\geq 2.0}(\vec{r}) = N_{\geq M_c}(\vec{r}) \cdot 10^{b(M_c(\vec{r}) - 2.0)} \quad (1)$$

442 where b is taken equal to 1.0 over the whole RELM area. Our constant b -
 443 value hypothesis is a minor deviation from the original study, as the authors
 444 considered also a local b -value of 1.94 for the Geysers area in Northern Cali-

445 fornia. The resulting density map $M \geq 2.0$ events is presented in Figure 6.
446 Our results are shown on the left panel, next to their results presented in
447 the middle panel. Normalized differences are represented on the right panel
448 and illustrate the consistency between both approaches. Only minor and
449 localized discrepancies can be identified for pixels with darker colors, almost
450 exclusively located in areas of lower activity. This confirms that absolute
451 differences between both approaches are very weak, and that our approach
452 leads to very similar outcomes.

453

454 The objective of Helmstetter et al. (2007) was to propose a forecast model
455 for $M \geq 5.0$ events over the 5-year period following the publication date of
456 their article. Their preferred model was chosen among a collection of 21
457 candidate models. Each model was evaluated based on a comparison of
458 extrapolated (i.e., forecasted) rates of $M \geq M_{min}$ events for a target 1996
459 - 2005 period, against actual earthquake counts for the same period in the
460 Advanced National Seismic System (ANSS) catalog. Model performance
461 was expressed in terms of a log-likelihood \mathcal{L} , and of a probability gain per
462 earthquake G relative to a model with a (spatially) uniform density of events.
463 Here, we only consider their preferred model, referred to as "Model #21" in
464 Table 1 of their article. Importantly, note that the nearest-neighbor rank
465 parameter n involved in the definition of their kernel bandwidth was also
466 optimized by the authors for each candidate model, so as to maximize model
467 log-likelihood in the target time interval. They retained the optimal $n = 6$

468 value in their analysis.

469 Similarly, we evaluated the log-likelihood and probability gain for our
470 results. These metrics ($\mathcal{L} = -3216$, $G = 6.71$) are reported in Table 1 to-
471 gether with those associated with their Model #21, for n between 1 and 6
472 (*A. Helmstetter, pers. comm., Jan. 2024*). Performance metrics for both
473 studies are in good agreement, albeit less favorable for our approach. Our
474 approach best matches with a power-law adaptive smoothing kernel where
475 the bandwidth is set equal to the distance to the closest neighbor ($n = 1$).
476 Note however that their two metrics do not deviate significantly from this
477 optimum at larger n values. The consistency between both approaches is also
478 in good agreement with the density evaluation process of Voronoi diagrams
479 (Nicholson et al., 2000). Indeed, as each Voronoi polygon contains a single
480 event, it also constitutes, by construction, a nearest-neighbour area to this
481 event. The approach proposed in section 2 is intrinsically adaptive. At this
482 point, it worth mentioning that we could not propagate location and magni-
483 tude uncertainties in this case-study for California. Because the information
484 was not available in the original ANSS catalogue used by Helmstetter et al.
485 (2007). However, if such information were used, one can anticipate that we
486 would have obtained a smoother seismicity map, that would potentially, in
487 turn, lead to results closer to those obtained by these authors, but not for
488 the same reasons. In their case, smoothing is imposed by the method; in our
489 case, it would be a consequence of uncertainty propagation.

4 Discussion and Conclusion

We have presented in section 2 a methodological framework based on Voronoi diagrams for the construction of seismicity models. The partitioning of spatial domain using the concept of Voronoi diagrams, or Dirichlet tessellation, is not particularly new (Dirichlet, 1850; Ord, 1978), but has proven to be useful for applications aimed at mapping point occurrences into spatially-continuous intensity maps of any quantity of physical interest (e.g., Nicholson et al., 2000; Moradi et al., 2019). In particular, we have illustrated this method with two representative use-cases: a smooth seismicity model for PSHA; and a density map used to replicate the RELM earthquake forecasting experiment in California. While past studies on such topics favored the use of smoothing kernels, we promote here a more objective, data-driven, approach, which would be less sensitive to end-user decisions.

503

We have implemented the method in section 3, with the intention to illustrate how our outcomes compare with those obtained using (adaptive) smoothing kernels, all other things being equal. We have found that results obtained using Voronoi diagrams are very similar to those obtained with adaptive smoothing kernels. Some small-scale discrepancies could be identified in very localized areas, where earthquake density is high. It is worth mentioning here that our results were obtained at a significantly lower computation cost than with adaptive smoothing kernels, especially due to

512 the required need for bandwidth and kernel type optimization. On the
513 contrary, Voronoi diagrams produce a self-adaptive, nearest-neighbor, tes-
514 sellation which is a pure data-driven process. In addition, by introducing
515 catalogue uncertainties in this process, the smoothness of the final map re-
516 flects a tangible property of the input catalogue. The smoothing pattern
517 is only governed by event density and by the level of uncertainty associated
518 with earthquake locations and magnitudes. This property differs significantly
519 from approaches involving smoothing kernels, where smoothness is also af-
520 fected by the specific kernel shape (e.g., power-law, Gaussian...) and its
521 bandwidth (e.g., constant-width, magnitude-dependent, adaptive...).

522

523 Voronoi polygons are, by construction, separated by sharp boundaries.
524 But in the proposed method the due propagation of uncertainties will likely
525 smooth contrasts in the final maps of seismic activity, enabling a statistically
526 sound interpretation of the smoothing pattern of the outcome. However, one
527 has to keep in mind that in some circumstances the approach may preserve
528 sharp transitions, when the data justifies a locally strong spatial variability
529 with a high confidence level (i.e., low uncertainties). We should also re-
530 mind end-users that intensity maps constructed from Voronoi diagrams are
531 not completely devoid of bias. It is known that they tend to over-smooth
532 intensity in low-density areas, and, respectively, to under-smooth intensity
533 in high-density areas. This matches with the discrepancies we observed be-
534 tween our outcomes and those obtained using adaptive smoothing kernels,

see sections 3.1 and 3.2. Advanced techniques have been proposed in the literature to tackle these effects. For example, Moradi et al. (2019) proposed to implement the smoothing of a Voronoi diagram by a resampling of the input events using an independent random thinning procedure. Other examples involve iterative techniques to enhance the smoothing of Voronoi diagrams, such as Centroidal Voronoi Tessellations (CVT) (Du et al., 1999), or Incomplete Centroidal Voronoi Tessellations (ICVT, Xiong et al., 2019). Their algorithm updates, for each iteration, the location of Voronoi seeds (or base points) to the location of the centroid in the corresponding Voronoi cell. The next iteration then starts with an update of the Voronoi diagram itself based on updated Voronoi seeds. The CVT algorithm stops when convergence is met between the locations of seeds and centroids. On the contrary, the ICVT algorithm stops after only few iterations, as soon as a target log-likelihood function reaches a first optimum. Finally, another potential issue may also appear when Voronoi diagrams are built from a very limited number of seeds, such as for magnitude bins containing only few events (i.e., for large magnitude events). This situation remains problematic and would spread Voronoi cells over possibly unrealistically large regions. It would thus potentially over-estimate the occurrence rate of such events in seismically quiet areas, and significantly under-estimate their occurrence rate in their epicentral area. Note that the slope of reconstructed FMD may be slightly affected under such circumstances. This point calls for further developments of the method, which should promote the (objective) concentration of earth-

558 quake density in the vicinity of (uncertain) epicentral locations.

559

560 In summary, we want to emphasize that a spatial partitioning method
561 based on Voronoi diagrams associated with a due propagation of catalogue
562 uncertainties using Monte-Carlo random sampling, can provide a valuable
563 alternative to smoothing kernels. This is particularly important for seis-
564 mological applications seeking objectivity and independence from arbitrary
565 operator decisions, such as PSHA or earthquake forecasting. Although one
566 can argue that implementing this technique would only lead to minor changes
567 in comparison to results obtained using adaptive smoothing kernels, we stress
568 here that its key advantage remains its ability to deliver a mapping entirely
569 driven by the statistical properties of the input catalogue.

570 5 Data and Resources

571 All data used in this paper came from published sources listed in the refer-
572 ences. The source-code of computer-programs used in this study in Python
573 language is available at <https://github.com/guyomd/vsm/tree/main>. The
574 Supplemental Material for this article includes a copy of this source-code,
575 with additional sensitivity analyses and figures.

576

⁵⁷⁷ **6 Declaration of competing interests**

⁵⁷⁸ The authors acknowledge there are no conflicts of interest recorded.

⁵⁷⁹ **7 Acknowledgments**

⁵⁸⁰ We thank S. Drouet for providing us with the earthquake catalogue, the
⁵⁸¹ smooth seismicity model and all parameters required to run the PSHA study
⁵⁸² for mainland France presented in Drouet et al. (2020). We also thank A.
⁵⁸³ Helmstetter for sending us the input data and computer programs used to
⁵⁸⁴ build forecast models in their original study (Helmstetter et al., 2007). We
⁵⁸⁵ would like to express our gratitude to A. Llenos and an anonymous reviewer
⁵⁸⁶ for their thorough reviews, comments and questions, which helped at signifi-
⁵⁸⁷ cantly improving our manuscript. These also triggered inspiration for future
⁵⁸⁸ applications and developments of the method.

⁵⁸⁹ Figures were produced using the GMT software (Wessel et al., 2019) and the
⁵⁹⁰ Python PyGMT module (Uieda et al., 2022). Computer programs were writ-
⁵⁹¹ ten in Python language and made significant use of the NumPy (Harris et al.,
⁵⁹² 2020) and Shapely (Gillies et al., 2022; GEOS contributors, 2024) packages.
⁵⁹³ Figures were prepared using the Generic Mapping Toolbox (Wessel et al.,
⁵⁹⁴ 2019) and the Python PyGMT package (Uieda et al., 2022). This manuscript
⁵⁹⁵ was prepared using the Overleaf online editor for L^AT_EX(www.overleaf.com).

596 **References**

- 597 C. D. Barr and F. P. Schoenberg. On the Voronoi estimator for the intensity
598 of an inhomogeneous planar Poisson process. *Biometrika*, 97(4):977–984,
599 2010. ISSN 00063444, 14643510. URL <http://www.jstor.org/stable/29777151>. Publisher: [Oxford University Press, Biometrika Trust].
600
- 601 D. Baumont, K. Manchuel, P. Traversa, C. Durouchoux, E. Nayman, and
602 G. Ameri. Intensity predictive attenuation models calibrated in Mw for
603 metropolitan France. *Bulletin of Earthquake Engineering*, 16(6):2285–
604 2310, June 2018. ISSN 1573-1456. doi: 10.1007/s10518-018-0344-6. URL
605 <https://doi.org/10.1007/s10518-018-0344-6>.
- 606 A. Bray, K. Wong, C. D. Barr, and F. P. Schoenberg. Voronoi residual
607 analysis of spatial point process models with applications to California
608 earthquake forecasts. *The Annals of Applied Statistics*, 8(4):2247–2267,
609 2014. doi: 10.1214/14-AOAS767. URL <https://arxiv.org/pdf/1501.06387.pdf>.
- 610
- 611 M. Burkhard and G. Grünthal. Seismic source zone characterization for
612 the seismic hazard assessment project PEGASOS by the Expert Group 2
613 (EG1b). *Swiss Journal of Geosciences*, 102(1):149–188, May 2009. ISSN
614 1661-8726, 1661-8734. doi: 10.1007/s00015-009-1307-3. URL <http://link.springer.com/10.1007/s00015-009-1307-3>.
- 615
- 616 T. Cao, M. D. Petersen, and M. S. Reichle. Seismic hazard estimate from

- 617 background seismicity in southern California. *Bulletin of the Seismological
618 Society of America*, 86(5):1372–1381, Oct. 1996. ISSN 1943-3573, 0037-
619 1106. doi: 10.1785/BSSA0860051372.
- 620 R. A. Clemens, F. P. Schoenberg, and A. Veen. Evaluation of spacetime
621 point process models using super-thinning. *Environmetrics*, 23:606–616,
622 2012. doi: 10.1002/env.2168.
- 623 C. A. Cornell. Engineering seismic risk analysis. *Bulletin of the Seismological
624 Society of America*, 58(5):1583–1606, Oct. 1968.
- 625 G. L. Dirichlet. Über die Reduction der Positiven Quadratischen Formen mit
626 Drei Unbestimmten Ganzen Zahlen. *J. reine Angew. Math.*, 40:209–227,
627 1850.
- 628 S. Drouet, G. Ameri, K. Le Dortz, R. Secanell, and G. Senfaute. A probabilis-
629 tic seismic hazard map for the metropolitan France. *Bulletin of Earthquake
630 Engineering*, 18(5):1865–1898, Mar. 2020. ISSN 1570-761X, 1573-1456.
631 doi: 10.1007/s10518-020-00790-7. URL <http://link.springer.com/10.1007/s10518-020-00790-7>.
- 633 Q. Du, V. Faber, and M. Gunzburger. Centroidal voronoi tessellations:
634 Applications and algorithms. *SIAM Rev*, 41(4):637–676, 1999. doi:
635 10.1137/S0036144599352836.
- 636 A. Dutfoy. Estimation of the GutenbergRichter Earthquake Recurrence
637 Parameters for Unequal Observation Periods and Imprecise Magnitudes.

- 638 *Pure and Applied Geophysics*, 177(10):4597–4606, Oct. 2020. ISSN 0033-
639 4553, 1420-9136. doi: 10.1007/s00024-020-02551-8. URL <http://link.springer.com/10.1007/s00024-020-02551-8>.
- 641 A. Dutfoy. Uncertainty on Estimated Magnitudes: A New Approach Based
642 on a Poisson Point Process of Dimension 2. *Pure Appl. Geophys.*, 180(3):
643 919–933, 2023. doi: 10.1007/s00024-022-03222-6.
- 644 A. Frankel. Mapping Seismic Hazard in the Central and Eastern United
645 States. *Seismol. Res. Letters*, 66(4):14, 1995.
- 646 J. K. Gardner and L. Knopoff. Is the sequence of earthquakes in Southern
647 California, with aftershocks removed, Poissonian? *Bulletin of the Seismo-*
648 *logical Society of America*, 64(5):1363–1367, Oct. 1974.
- 649 GEOS contributors. *GEOS computational geometry library*. Open Source
650 Geospatial Foundation, 2024. URL <https://libgeos.org/>.
- 651 M. C. Gerstenberger, W. Marzocchi, T. Allen, M. Pagani, J. Adams, L. Dan-
652 ciu, E. H. Field, H. Fujiwara, N. Luco, K. Ma, C. Meletti, and M. D.
653 Petersen. Probabilistic Seismic Hazard Analysis at Regional and Na-
654 tional Scales: State of the Art and Future Challenges. *Reviews of Geo-*
655 *physics*, 58(2):e2019RG000653, June 2020. ISSN 8755-1209, 1944-9208.
656 doi: 10.1029/2019RG000653.
- 657 S. Gillies, C. van der Wel, J. Van den Bossche, M. W. Taves, J. Arnott,

- 658 B. C. Ward, et al. Shapely, Dec. 2022. URL <https://doi.org/10.5281/zenodo.7428463>.
- 660 C. Gini. On the Measure of Concentration with Special Reference to Income
661 and Statistics. *Colorado College Publication, General Series*, 208:73–79,
662 1936.
- 663 B. Gutenberg and C. F. Richter. Frequency of Earthquakes in California.
664 *Bulletin of the Seismological Society of America*, 34(4):185–188, 1944.
- 665 C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen,
666 D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith,
667 R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane,
668 J. F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard,
669 T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, and T. E. Oliphant.
670 Array programming with NumPy. *Nature*, 585(7825):357–362, Sept.
671 2020. doi: 10.1038/s41586-020-2649-2. URL <https://doi.org/10.1038/s41586-020-2649-2>.
- 673 A. Helmstetter. Comparison of Short-Term and Time-Independent Earthquake Forecast Models for Southern California. *Bulletin of the Seismological Society of America*, 96(1):90–106, Feb. 2006. ISSN 0037-1106. doi:
674 10.1785/0120050067. URL <https://pubs.geoscienceworld.org/bssa/article/96/1/90-106/146804>.
- 678 A. Helmstetter and M. J. Werner. Adaptive Spatiotemporal Smoothing

- 679 of Seismicity for Long-Term Earthquake Forecasts in California. *Bulletin
680 of the Seismological Society of America*, 102(6):2518–2529, Dec.
681 2012. ISSN 0037-1106. doi: 10.1785/0120120062. URL <https://pubs.geoscienceworld.org/bssa/article/102/6/2518-2529/331555>.
- 682
- 683 A. Helmstetter, Y. Y. Kagan, and D. D. Jackson. High-resolution Time-
684 independent Grid-based Forecast for $M \geq 5$ Earthquakes in California.
685 *Seismological Research Letters*, 78(1):78–86, Jan. 2007. doi: 10.1785/gssrl.
686 78.1.78.
- 687 W. T. Holmes. The 1997 NEHRP Recommended Provisions for Seismic
688 Regulations for New Buildings and Other Structures. *Earthquake Spectra*,
689 16(1):101–114, Feb. 2000. doi: 10.1193/1.1586085.
- 690 D. D. Jackson and Y. Y. Kagan. Testable Earthquake Forecasts for 1999.
691 *Seismological Research Letters*, 70(4):393–403, July 1999. doi: 10.1785/
692 gssrl.70.4.393.
- 693 A. L. Kafka and J. R. Walcott. How Well Does the Spatial Distribution
694 of Smaller Earthquakes Forecast the Locations of Larger Earthquakes in
695 the Northeastern United States? *Seismological Research Letters*, 69(5):
696 428–440, Sept. 1998. doi: 10.1785/gssrl.69.5.428.
- 697 Y. Y. Kagan and D. D. Jackson. Seismic Gap Hypothesis: Ten years after.
698 *Journal of Geophysical Research: Solid Earth*, 96(B13):21419–21431, Dec.
699 1991. doi: 10.1029/91JB02210.

- 700 Y. Y. Kagan and D. D. Jackson. Long-term probabilistic forecasting of
701 earthquakes. *Journal of Geophysical Research: Solid Earth*, 99(B7):13685–
702 13700, July 1994. doi: 10.1029/94JB00500.
- 703 Y. Y. Kagan, D. D. Jackson, D. Schorlemmer, and M. Gerstenberger. Testing
704 hypotheses of earthquake occurrence. In *AGU Fall Meeting Abstracts*,
705 pages S31G–01, 2003.
- 706 K. Manchuel, P. Traversa, D. Baumont, M. Cara, E. Nayman, and C. Durou-
707 choux. The French seismic CATalogue (FCAT-17). *Bulletin of Earthquake
708 Engineering*, 16(6):2227–2251, June 2018. ISSN 1570-761X, 1573-1456.
709 doi: 10.1007/s10518-017-0236-1. URL <http://link.springer.com/10.1007/s10518-017-0236-1>.
- 711 R. K. Mc Guire. FORTRAN Computer Program for Seismic Risk Analysis.
712 United States Geological Survey Open-File Report 76-67, U.S. Geologi-
713 cal Survey, Washington, D.C., 1976. URL [https://doi.org/10.3133/
714 ofr7667](https://doi.org/10.3133/ofr7667).
- 715 M. M. Moradi, O. Cronie, E. Rubak, R. Lachieze-Rey, J. Mateu, and A. Bad-
716 deley. Resample-smoothing of Voronoi intensity estimators. *Statistics
717 and Computing*, 29(5):995–1010, Sept. 2019. ISSN 0960-3174, 1573-1375.
718 doi: 10.1007/s11222-018-09850-0. URL <http://link.springer.com/10.1007/s11222-018-09850-0>.
- 720 M. P. Moschetti. A long-term earthquake rate model for the central and

721 eastern united states from smoothed seismicity. *Bulletin of the Seismological*
722 *Society of America*, 105(6):2928–2941, 11 2015. ISSN 0037-1106. doi:
723 10.1785/0120140370.

724 T. Nicholson, M. Sambridge, and O. Gudmundsson. On entropy and clus-
725 tering in earthquake hypocentre distributions. *Geophysical Journal Inter-*
726 *national*, 142(1):37–51, July 2000. doi: 10.1046/j.1365-246x.2000.00113.x.

727 Y. Ogata, K. Katsura, and M. Tanemura. Modelling Heterogeneous Space-
728 Time Occurrences of Earthquakes and its Residual Analysis. *Journal of the*
729 *Royal Statistical Society Series C: Applied Statistics*, 52(4):499–509, Oct.
730 2003. ISSN 0035-9254, 1467-9876. doi: 10.1111/1467-9876.00420. URL
731 <https://academic.oup.com/jrsssc/article/52/4/499/7112953>.

732 K. Ord. How Many Trees in a Forest? *The Mathematical Scientist*, 3:23–33,
733 1978.

734 P. Reasenberg. Second-order moment of Central California Seismicity, 1969–
735 1982. *Journal of Geophysical Research: Solid Earth*, 90(B7):5479–5495,
736 June 1985.

737 D. Rhoades. Estimation of the Gutenberg-Richter relation allowing for in-
738 dividual earthquake magnitude uncertainties. *Tectonophysics*, 258(1-4):
739 71–83, 1996.

740 W. H. Savran, J. A. Bayona, P. Iturrieta, K. M. Asim, H. Bao,
741 K. Bayliss, M. Herrmann, D. Schorlemmer, P. J. Maechling, and

- 742 M. J. Werner. pyCSEP: A Python Toolkit for Earthquake Fore-
743 cast Developers. *Seismological Research Letters*, 93(5):2858–2870,
744 Sept. 2022. ISSN 0895-0695, 1938-2057. doi: 10.1785/0220220033.
745 URL <https://pubs.geoscienceworld.org/srl/article/93/5/2858/615824/pyCSEP-A-Python-Toolkit-for-Earthquake-Forecast>.
746
- 747 D. Schorlemmer, M. C. Gerstenberger, S. Wiemer, D. D. Jackson, and D. A.
748 Rhoades. Earthquake Likelihood Model Testing. *Seismological Research
749 Letters*, 78(1):17–29, 01 2007. doi: 10.1785/gssrl.78.1.17.
- 750 J. C. Stepp. Analysis of Completeness of the Earthquake Sample in the
751 Puget Sound Area and Its Effect on Statistical Estimates of Earthquake
752 Hazard. In *Proc. of the 1st Int. Conf. on Microzonation*, volume 2, pages
753 pp. 897–910, Seattle, 1972.
- 754 C. Stock and E. Smith. Adaptive Kernel Estimation and Continuous Prob-
755 ability Representation of Historical Earthquake Catalogs. *Bulletin of
756 the Seismological Society of America*, 92(3):904–912, Apr. 2002. doi:
757 10.1785/0120000233.
- 758 M. Stucchi, A. Rovida, A. A. Gomez Capera, P. Alexandre, T. Camel-
759 beeck, M. B. Demircioglu, P. Gasperini, V. Kouskouna, R. M. W. Mus-
760 son, M. Radulian, K. Sesetyan, S. Vilanova, D. Baumont, H. Bungum,
761 D. Fäh, W. Lenhardt, K. Makropoulos, J. M. Martinez Solares, O. Scotti,
762 M. ivi, P. Albini, J. Batllo, C. Papaioannou, R. Tatevossian, M. Locati,

- 763 C. Meletti, D. Viganò, and D. Giardini. The SHARE European Earth-
764 quake Catalogue (SHEEC) 10001899. *Journal of Seismology*, 17(2):523–
765 544, Apr. 2013. ISSN 1573-157X. doi: 10.1007/s10950-012-9335-2. URL
766 <https://doi.org/10.1007/s10950-012-9335-2>.
- 767 M. Taroni and A. Akinci. A New Smoothed Seismicity Approach to Include
768 Aftershocks and Foreshocks in Spatial Earthquake Forecasting: Applica-
769 tion to the Global $Mw \geq 5.5$ Seismicity. *Applied Sciences*, 11(22):10899,
770 Nov. 2021. doi: 10.3390/app112210899.
- 771 S. Tinti and F. Mulargia. Effects of magnitude uncertainties on estimating
772 the parameters in the Gutenberg-Richter frequency-magnitude law. *Bull.*
773 *Seismol. Soc. Am.*, 75(6):1681–1697, 1985.
- 774 P. Traversa, D. Baumont, K. Manchuel, E. Nayman, and C. Durouchoux. Ex-
775 ploration tree approach to estimate historical earthquakes Mw and depth,
776 test cases from the French past seismicity. *Bulletin of Earthquake Engi-*
777 *neering*, 16(6):2169–2193, June 2018. ISSN 1570-761X, 1573-1456. doi:
778 10.1007/s10518-017-0178-7. URL <http://link.springer.com/10.1007/s10518-017-0178-7>.
- 780 L. Uieda, D. Tian, W. J. Leong, M. Jones, W. Schlitzer, M. Grund,
781 L. Toney, Y. Fröhlich, J. Yao, Y. Magen, K. Materna, A. Belem, T. New-
782 ton, A. Anant, M. Ziebarth, J. Quinn, and P. Wessel. PyGMT: A
783 Python interface for the Generic Mapping Tools, Dec. 2022. URL <https://doi.org/10.5281/zenodo.7481934>.

- 785 A. Valentini, F. Visini, and B. Pace. Integrating faults and past earth-
786 quakes into a probabilistic seismic hazard model for peninsular Italy. *Nat-*
787 *ural Hazards and Earth System Sciences*, 17(11):2017–2039, Nov. 2017.
788 ISSN 1684-9981. doi: 10.5194/nhess-17-2017-2017. URL [https://nhess.](https://nhess.copernicus.org/articles/17/2017/2017/)
789 [copernicus.org/articles/17/2017/2017/](https://nhess.copernicus.org/articles/17/2017/2017/).
- 790 D. Vere-Jones. Statistical methods for the description and display of earth-
791 quake catalogs. In *Statistics in the Environmental and Earth Sciences*,,
792 pages 220–246. Hodder Education, London, a. t. walden and p. guttorp
793 (editors) edition, 1992.
- 794 M. G. Voronoi. Nouvelles applications des paramètres continus à la théorie
795 des formes quadratiques. *J. reine Angew. Math.*, 134:198–287, 1908.
- 796 D. H. Weichert. Estimation of the earthquake recurrence parame-
797 ters for unequal observation periods for different magnitudes. *Bul-*
798 *letin of the Seismological Society of America*, 70(4):1337–1346, Aug.
799 1980. ISSN 1943-3573, 0037-1106. doi: 10.1785/BSSA0700041337.
800 URL <https://pubs.geoscienceworld.org/bssa/article/70/4/1337/118111/Estimation-of-the-earthquake-recurrence-parameters>.
- 802 M. J. Werner, A. Helmstetter, D. D. Jackson, and Y. Y. Kagan. High-
803 Resolution Long-Term and Short-Term Earthquake Forecasts for Califor-
804 nia. *Bulletin of the Seismological Society of America*, 101(4):1630–1648,
805 Aug. 2011. ISSN 0037-1106. doi: 10.1785/0120090340.

- 806 P. Wessel, J. F. Luis, L. Uieda, R. Scharroo, F. Wobbe, W. H. F. Smith,
807 and D. Tian. The generic mapping tools version 6. *Geochemistry, Geo-*
808 *physics, Geosystems*, 20:55565564, 2019. doi: <https://doi.org/10.1029/2019GC008515>.
- 810 G. Woo. Kernel Estimation Methods for Seismic Hazard Area Source Model-
811 ing. *Bulletin of the Seismological Society of America*, 86(2):353–362, 1996.
- 812 Z. Xiong, J. Zhuang, and S. Zhou. LongTerm Earthquake Haz-
813 ard in North China Estimated from a Modern Catalog. *Bul-*
814 *letin of the Seismological Society of America*, 109(6):2340–2355, Dec.
815 2019. ISSN 0037-1106, 1943-3573. doi: 10.1785/0120190066.
816 URL <https://pubs.geoscienceworld.org/ssa/bssa/article/109/6/2340/574616/LongTerm-Earthquake-Hazard-in-North-China>.
817

⁸¹⁸ Mailing addresses:

⁸¹⁹ G.D. and P. A.

⁸²⁰ Electricité de France (EDF)

⁸²¹ Direction Technique

⁸²² 180 rue du lieutenant Parayre

⁸²³ 13290 Aix-en-Provence

⁸²⁴ France

	\mathcal{L}	G
HKJ 2007		
$n = 1$	-3 213	6.71
$n = 2$	-3 102	6.99
$n = 3$	-3 075	7.06
$n = 4$	-3 071	7.07
$n = 5$	-3 075	7.06
$n = 6$	-3 069	7.08
This study	-3 216	6.71

Table 1: Comparison of the forecast model performance proposed by Helmstetter et al. (2007) (case #21 in Table 1), for different configurations of their adaptive smoothing approach (power-law kernel), and for our method. The model performance is expressed in terms of a log-likelihood (\mathcal{L}) and of a probability gain per earthquake (G) for each configuration. Note that both approaches present similar performance. Performance indicators for our method best match with those of an adaptive power-law kernel with a bandwidth equal to the first nearest-neighbor epicentral distance ($n = 1$).

825 **List of Figures**

826	1	Mapping earthquake density with Voronoi diagrams. Illustrative example based on a synthetic catalogue of earthquakes with magnitudes ranging from 2.0 to 6.8. (top) Epicentral map of earthquakes with magnitude included in the [3.0; 3.5] interval. (center) Corresponding Voronoi diagram for the same magnitude bin. (bottom) Gridded seismic density map over the domain. In each cell with a 0.1 x 0.1 square-degree area, the density is obtained by summing all portions of Voronoi polygons that intersect the cell. The colorscaling depends on earthquake density in each cell, or polygon, normalized for an area of 1 000 km ² . Parameters were randomly sampled in space according to a pseudo-bivariate normal distribution. <i>b</i> -values are almost-constant over the domain, with an average <i>b</i> -value of 1.0 and standard deviation $\sigma_b = 10^{-8}$. <i>a</i> -values have an average logarithm (base 10) of $L = 1.0$ (per unit time), with standard-deviation $\sigma_L = 0.2$ over the whole domain \mathcal{D} . Random earhquake counts in each cell were drawn for a period of 10,000 time units.	46
843			

863	4	Construction of a discrete frequency-magnitude distribution for each pixel of the spatial domain. For each bin in magni- tude, the seismicity rate is read from the corresponding pixel in a density map obtained following the steps described in section 2.1 and Figure 1.	49
864			
865			
866			
867			
868	5	Distributed seismicity model for mainland France, expressed in terms of a -values of the Gutenberg-Richter formula, nor- malized for a 10^6 km^2 area. (a) Model based on Voronoi di- agrams, as described in section 2, accounting for earthquake location and magnitude uncertainties (200 random perturba- tions). (b) Model published in Drouet et al. (2020) (D2020), obtained using the same catalogue and an adaptive smoothing kernel method. (c) $M \geq 3$ epicenters used for this study, from the historical catalogue of D2020. (d) Normalized differences between panels <i>a</i>) and <i>b</i>), in percent. Positive differences oc- cur where the Voronoi-based approach predicts higher rates than D2020, and conversely.	50
869			
870			
871			
872			
873			
874			
875			
876			
877			
878			
879			

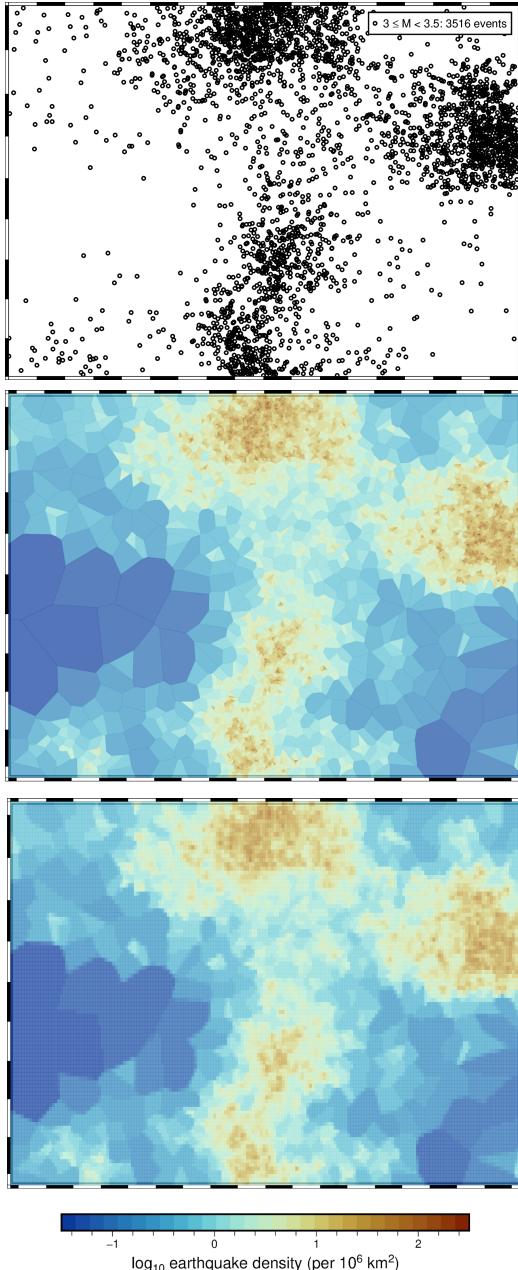


Figure 1: Mapping earthquake density with Voronoi diagrams. Illustrative example based on a synthetic catalogue of earthquakes with magnitudes ranging from 2.0 to 6.8. **(top)** Epicentral map of earthquakes with magnitude included in the [3.0; 3.5] interval. **(center)** Corresponding Voronoi diagram for the same magnitude bin. **(bottom)** Gridded seismic density map over the domain. In each cell with a 0.1×0.1 square-degree area, the density is obtained by summing all portions of Voronoi polygons that intersect the cell. The colorscaling depends on earthquake density in each cell, or polygon, normalized for an area of 1000 km^2 . Parameters were randomly sampled in space according to a pseudo-bivariate normal distribution. b -values are almost-constant over the domain, with an average b -value of 1.0 and standard deviation $\sigma_b = 10^{-8}$. a -values have an average logarithm (base 10) of $\log_{10}(a) = 0$ ($a = 1$), with a standard deviation of $\sigma_a = 0.05$.

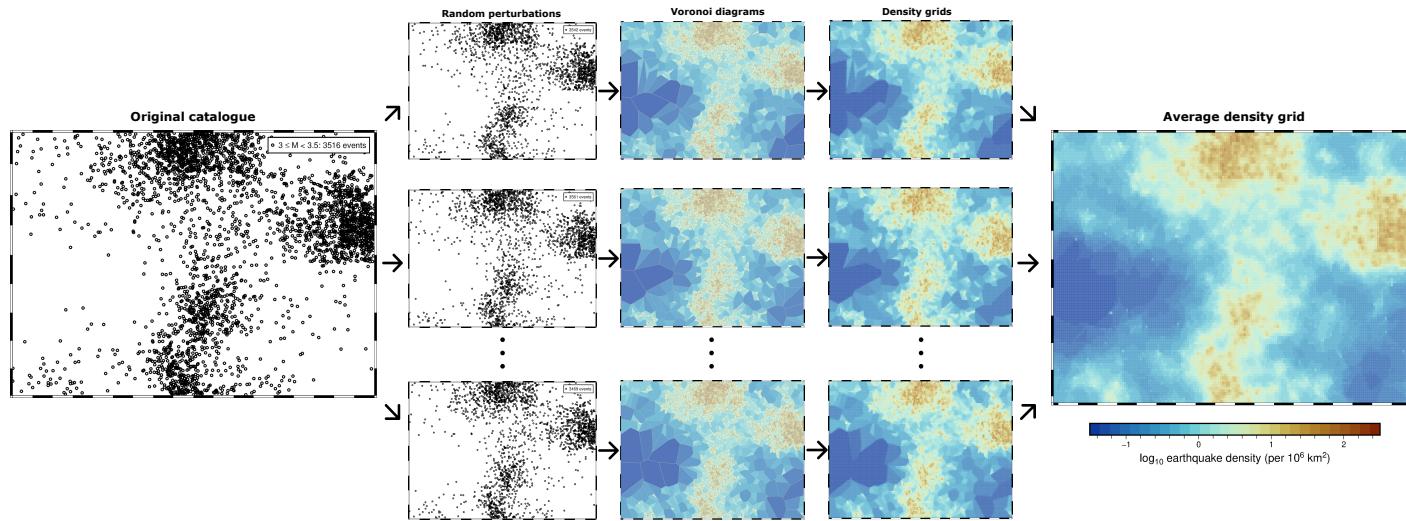


Figure 2: Monte-Carlo sampling process for the propagation of uncertainties on density maps. This figure illustrates the process for a single bin in magnitude. For each realization we use random perturbations of epicentral locations and magnitudes based on uncertainties provided in the original catalogue. Note also that the variability of event counts in each bin is modelled as a random Poisson variable. The final map is an average of density maps obtained for all realizations.

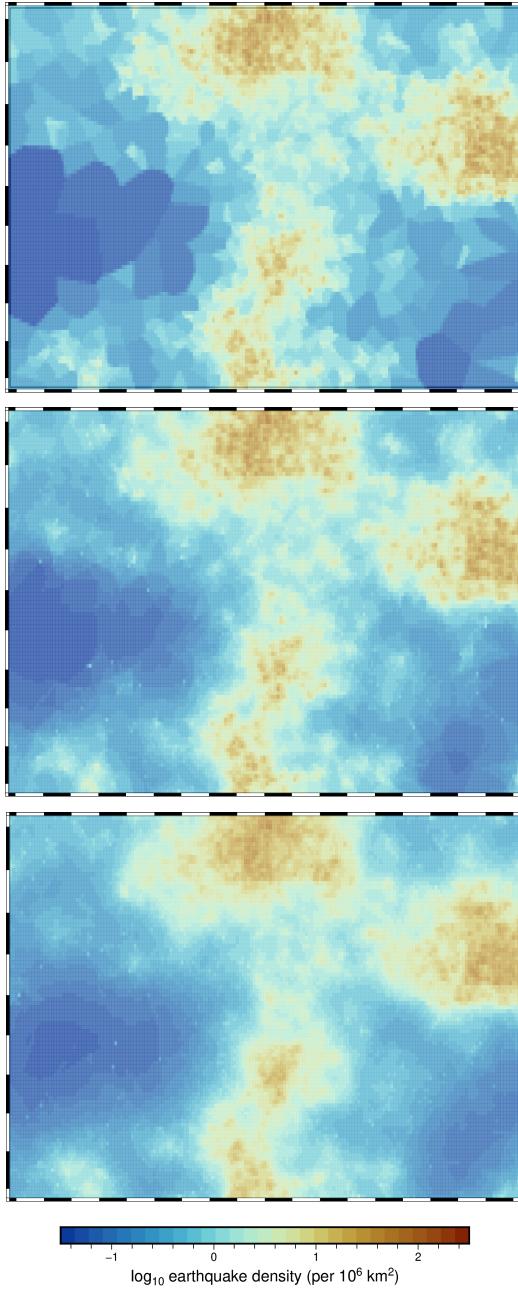


Figure 3: Impact of uncertainties on the smoothing of earthquake density maps, for grid cells with a 0.1×0.1 square-degree area. (**top**) Density map obtained without accounting for uncertainties (this panel is similar to the bottom panel of Figure 1). (**middle**) Map averaged over 100 Monte-Carlo random realizations of perturbed earthquake locations. (**bottom**) Map averaged over 100 Monte-Carlo random⁴⁸realizations of perturbed earthquake locations and magnitudes. Note that, in both cases, random perturbations of the catalogue account for the variability of Poisson-distributed event counts, see section 2.2.

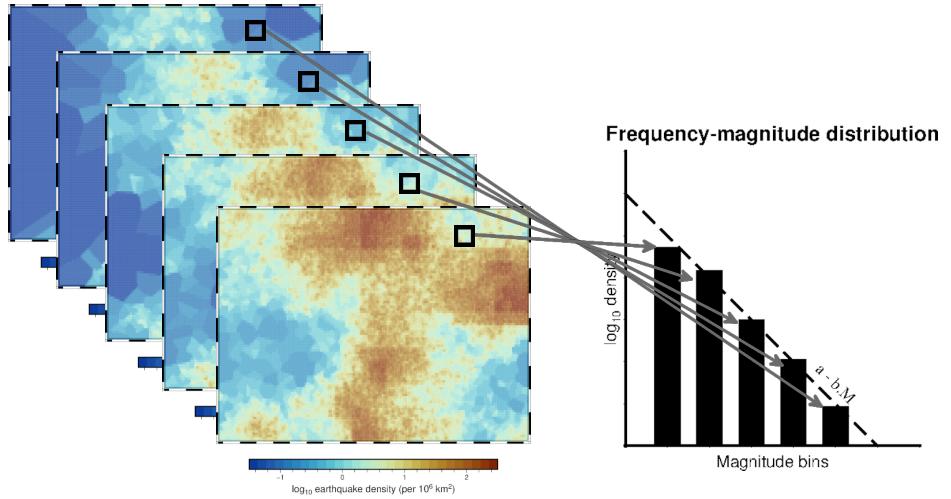


Figure 4: Construction of a discrete frequency-magnitude distribution for each pixel of the spatial domain. For each bin in magnitude, the seismicity rate is read from the corresponding pixel in a density map obtained following the steps described in section 2.1 and Figure 1.

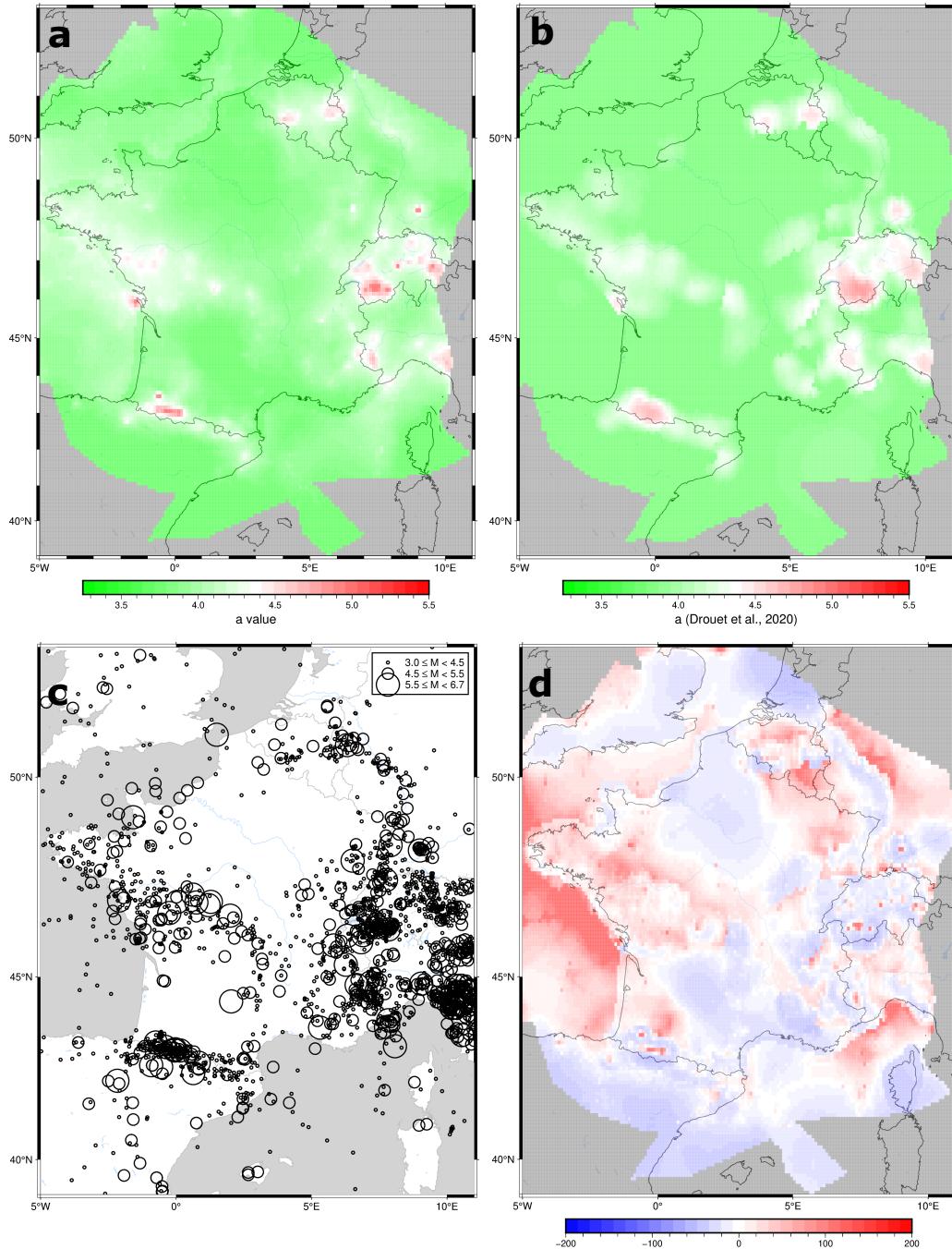


Figure 5: Distributed seismicity model for mainland France, expressed in terms of a -values of the Gutenberg-Richter formula, normalized for a 10^6 km^2 area. (a) Model based on Voronoi diagrams, as described in section 2, accounting for earthquake location and magnitude uncertainties (200 random perturbations). (b) Model published in Drouet et al. (2020) (D2020), obtained using the same catalogue and an adaptive smoothing kernel method. (c) $M \geq 3$ epicenters used for this study, from the historical catalogue of D2020. (d) Normalized differences between panels a) and b), in percent. Positive differences occur where the Voronoi-based approach predicts higher rates than D2020, and conversely.

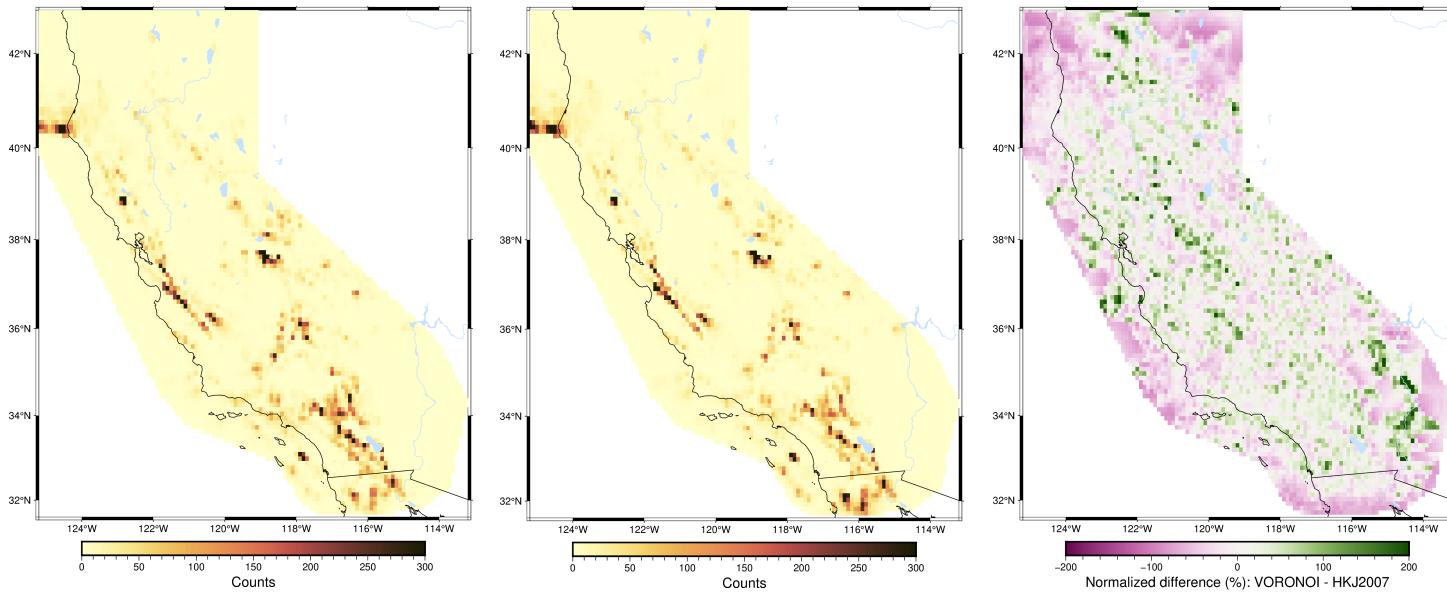


Figure 6: Counts of earthquakes with $M \geq 2.0$ for California, based on inputs from the RELM experiment. Note that earthquake counts were extrapolated in each cell based on completeness levels estimated by (Helmstetter et al., 2007) using a b -value of 1.0. Grid cells have a 0.1×0.1 square-degree area. **(left)** Earthquake counts obtained using the method described in section 2. **(middle)** Original earthquake counts obtained by Helmstetter et al. (2007) using an optimized smoothing power-law isotropic kernel. **(right)** Difference in earthquake counts per cell between the two approaches, normalized by counts of Helmstetter et al. (2007).

[Click here to access/download](#)

**Supplemental Material (All Other Files, i.e. Movie, Zip,
csv)
Electronic_Supplement.zip**