Insights on the Italian Seismic Network from location uncertainties

Salvatore Scudero · Carlo Marocchi ·
Antonino D'Alessandro

Received: 13 November 2020 / Accepted: 3 May 2021 / Published online: 5 June 2021
© The Author(s) 2021

Abstract Probabilistic earthquake locations provide confidence intervals for the hypocentre solutions such as errors encountered in the position, the origin time, and in magnitude. If the relationship of the parameters relative to the local arrangement of the seismic network is considered, such as the node distance, the number of stations, the seismic gap, and the quality of phase readings), the uncertainties can then provide insights on the location capability of the network. In this paper, we collect the earthquake data recorded from the Italian Seismic Network for a time span of 5 years. The data pertain to three different catalogues according to the progressive refinement phases of the location procedure: automatic location, revised location, and published location. By means of spatial analysis, we assess the distribution of the location-related and network-related estimators across the study area. These estimators are subsequently combined to assess the existence of spatial correlations at a local scale. The results indicate that the Italian network is generally able to provide robust locations at the national scale and for smaller earthquakes, and the elongated shape of Italy (and of its network) does not cause

systematic bias in the locations. However, we highlight the existence of subregions in which the performance of the network is weaker. At present, a unique 2D, 3-layer velocity model is used for the earthquake location procedure, and this could represent the main limitation for the improvement of the locations. Therefore, the assessment of locally optimized velocity models is the priority for the homogenization and the improvement of the Italian Seismic Network performance.

Keywords Probabilistic earthquake location · Seismic network · Spatial statistics

1 Introduction

An earthquake is defined by its focal parameters, and its estimation is a classical geophysical inverse problem; given the arrival times of P- and S-waves at seismic stations and given a velocity model, the unknowns are the spatio-temporal coordinates of the hypocentre. The solution (i.e. earthquake location) is computed by solving a system of equations where the residual between calculated and observed arrival times is minimized. Probabilistic earthquake locations provide a complete description of the uncertainty estimate in terms of probability density functions presented in Gaussian form (cf Tarantola and Valette 1982; Lomax et al. 2000) which allow the evaluation of the quality of the locations.

S. Scudero (✉) · A. D'Alessandro
Osservatorio Nazionale Terremoti, Istituto Nazionale
di Geofisica e Vulcanologia, Rome, Italy
e-mail: salvatore.scudero@ingv.it

C. Marocchi
Sezione Roma2, Istituto Nazionale di Geofisica
e Vulcanologia, Rome, Italy

Systematic and random errors can affect the location: while the former influence the accuracy (i.e. difference from the “true” location), the latter influence the precision (confidence of the location). Errors can be generally introduced by (i) the elementary model used in the source representation, (ii) simplified 1D seismic velocity models, (iii) the suboptimized distribution of the seismic stations, and (iv) the quality of the data themselves. As a consequence, the systematic analysis of the uncertainties can be useful to derive information about the abovementioned sources of error and assess the performance of a seismic network. Such assessment is of crucial importance in order to find possible “weak” zones and therefore address the future development of the network.

The performance of a seismic network is usually evaluated by numerical simulations (Schumann et al. 1988; D’Alessandro et al. 2011, 2012, 2013; D’Alessandro and Stickney 2012; D’Alessandro and Ruppert 2012; Möllhoff et al. 2019), with the a posteriori analysis of the quality of recorded earthquakes (Turino et al. 2010; Mahani et al. 2016; Saunders et al. 2016; Cattaneo et al. 2017; Peng et al. 2018; Michele et al. 2019), or also by evaluating the spatial fit with historical and instrumental seismicity (Siino et al. 2020). In this paper, we analyze earthquake locations recorded by the Italian Seismic Network in order to perform its evaluation. In particular, we analyze the content of three different earthquake catalogues resulting from the progressive refinement location procedure. The main objective is to explore the spatial correlations between the network-related and uncertainty-related estimators of the locations to find clues that indicate for network weakness. Secondly, examining the difference among the three locations, we can also discuss the steps of the location procedure.

2 Data

The Italian Seismic Network is managed by the “Istituto Nazionale di Geofisica e Vulcanologia” (INGV). During the last 30 years, the INGV built an extensive seismic network, which has been growing both in number of sites and in the quality of the installed instruments (Amato and Mele 2008; Michelini et al. 2016); today, it counts more than 550 seismic stations. All of the seismic signals are transmitted in real time to the seismic control room located in the headquarters

of the INGV in Rome (Italy). The first communication of the earthquake epicentral coordinates and magnitude must be accomplished in a few minutes. For this purpose, an optimized automatic procedure, able to pick seismic phases and compute the location promptly, is implemented at the seismic room. These automatic locations can generate large errors, and therefore, the technical staff rapidly check the automatic data processing and provide a more accurate location within few minutes (revised location). Last, in a more relaxed time, a further revision is performed; it includes a careful analysis of the waveforms associated with a single event in order to calculate the highest-quality, ultimate locations. This step is generally performed for all the relevant events ($M_L > 2.5$) (Amato and Mele 2008). However, from the final locations are excluded all those events falling outside the country’s border (except surrounding seas): the automatic procedure locates all the events, but when the event is confirmed and (revised location) in another country, this is not included in the final revision. Events such as quarry blasts and detonations of bombs are also ignored when recognized; these “events” could be as many as hundreds per year (Gulia 2010; Mele et al. 2010a). Unwanted, but unavoidable, is the cut of events as a consequence of the extraordinary rate of earthquakes during seismic sequences. The list of all the expert-revised earthquakes is published every four months in the “Bollettino Sismico Italiano”. The details about the entire procedure can be found in Amato and Mele (2008), Mele et al. (2010a), and Michelini et al. (2016).

Seismic data are processed in real time by means of the “Earthworm” software package. It was originally proposed by Johnson et al. (1995) and continuously developed (see www.earthwormcentral.org). It is a robust and well-tested tool that is largely adopted worldwide. It implements the picker originally proposed by Allen (1978, 1982) with parameters that can be conveniently calibrated. Mele et al. (2010b) performed the configuration of the package, tuning the code’s parameters to account for the specific context of the Italian Seismic Network (types of sensors, sampling frequency, background noise at the sites, shape of the network, etc.). The velocity model consists of a two-layered crust (thickness of 11.1 and 26.9 km, V_p of 5 and 6.5 km/s, V_s of 2.89 and 3.75 km/s). The mantle is represented by a half-space with $V_p = 8.05$ km/s and $V_s = 4.65$ km/s. Even though it is able to provide

acceptable locations, this model represents an evident simplification, especially considering the geodynamic complexity of Italy. Details about the implementation of the location system of the Italian Seismic Network are in Mele et al. (2010a) and Michelini et al. (2016))

The data set used in this work has been compiled by selecting from the automatic (AUTO hereafter), revised (REV hereafter), and published (FINAL hereafter) catalogues all of the seismic events with magnitude $M \geq 2.5$ that occurred from May 2012 to May 2017. The magnitude threshold at 2.5 ensures a complete and homogeneous catalogue. In fact, the completeness magnitude for the Italian Seismic Network is between ≈ 1.7 and ≈ 2.5 despite some spatial and temporal variations (Marchetti et al. 2004; Amato

and Mele 2008; Schorlemmer et al. 2010; Chiarabba et al. 2015). In 2012, the procedure for the compilation of the catalogues was revised and, since then, their content can be considered homogeneous. The earthquake data range from 7°E to 19°E and from 36°N to 47°N (centred on the Italian territory). The three catalogues contain 6318, 6318, and 3149 events. The seismicity is not uniformly distributed in space (Fig. 1a) and in time (Fig. 1c). It well marks the Apennine belt and the southern Tyrrhenian Sea, while it is scattered in the other regions and almost totally lacking in Sardinia and in the central Tyrrhenian basin. The earthquakes are mainly crustal (depth < 20–25 km), but intermediate and deep seismicity (down to approximately 450 km) characterizes the southern Tyrrhenian

Fig. 1 **a** Spatial distribution of *AUTO*, *REV*, and *FINAL* catalogues classified according to magnitude (size of the circle) and depth (colour). The triangles indicate the stations of the Italian Seismic Network. **b** Log-plot of the normalized frequency

distribution of earthquake magnitude for the three catalogues. **c** Temporal distribution for *AUTO* and *REV*, and *FINAL* catalogues. **d** Simplified tectonic map of Italy. The solid line indicates the outer front of the Africa-Eurasia collision

basin, where the Ionian slab is subducting (Chiarabba et al. 2005). Beyond the almost constant background seismicity, the catalogues include two relevant seismic sequences, namely, the 2012 Emilia sequence and the 2016–2017 central Italy sequence with main shocks of $M = 5.9$ and $M = 6.5$, respectively. In particular, the 2016–2017 sequence, with hundreds of events per day (Fig. 1c), represented a challenging task for the seismic analysts (Marchetti et al. 2016). Because of the massive amount of events that occurred, the refinement of the automatic locations (*AUTO*) was neglected for many smaller events that were not listed in the ultimate (*FINAL*) catalogue.

The data set contains the common information about the event location and the associated uncertainties. In particular, we selected eight parameters, splitting them into two groups: a group of “network” estimators (seismic gap, minimum distance, number of stations, and average phase weights) and a group that includes the “uncertainty” estimators (horizontal error, vertical error, time error, and magnitude error). Table 1 and Fig. 2 summarize and compare the main statistical indexes and the frequency distribution of the estimators for the three catalogues. In general, for all of the estimators, the differences in the statistical indexes (i.e. the mean and standard deviation) are greater passing from *AUTO* to *REV*, since the *REV* and

FINAL catalogues are usually very similar (Table 1 and Fig. 2). The details about the indexes are provided in the following paragraphs.

2.1 Earthquake magnitude

For a more rapid estimate of the magnitude, M_L is also automatically computed also for the larger events but by using only a subset of stations (Amato and Mele 2008). For this reason, the *AUTO* catalogue includes the M_L for 99.1% of events, while *REV* lists 97% of M_L because M_w and mb are computed for the strongest events. The magnitude-frequency relationship of the three seismic catalogues follows the linear trend of the Gutenberg-Richter law and they can be considered complete (Fig. 1b). Generally, the *AUTO* and *REV* catalogues tend to slightly overestimate the earthquakes’ magnitude with respect to the *FINAL*. In detail, for approximately 50% of the earthquakes the difference is less than ± 0.1 M; the magnitude is over-(under-)estimated by more than 0.1 for 42% (8%) of the events, and the average value is $+0.19$ (-0.35). The *REV* and *FINAL* catalogues show an almost perfect overlap of magnitude values).

We calculated Pearson’s correlation coefficients of earthquake magnitude (M) against the eight estimators (Table 2). They are all uncorrelated, except for a weak,

Table 1 Number of observations, average value, standard deviation, and minimum and maximum values for each selected parameter for the three catalogues (*AUTO*, *REV*, and *FINAL*)

	GAP (°)	MIN (km)	STA	WPH	ERH (km)	ERZ (km)	RMS (s)	ERM
n_{AUTO}	6318	6318	6318	6264	6318	6318	6318	6318
mean _{AUTO}	71.97	19.08	53.90	1.53	2.22	4.88	0.82	0.41
St. dev _{AUTO}	53.75	31.93	30.89	0.26	5.80	14.81	0.80	0.18
min _{AUTO}	15.00	0.00	4.00	0.00	0.23	0.15	0.00	0.00
max _{AUTO}	329	513	204	3	99	99	13.29	5.28
n_{REV}	6318	6318	6318	6264	6318	5711	6318	6193
mean _{REV}	98.221	17.803	49.381	2.062	1.043	0.845	0.322	0.280
St. dev _{REV}	62.164	30.745	31.541	0.815	1.416	1.058	0.151	0.064
min _{REV}	17.00	0.00	3.00	0.00	0.00	0.10	0.02	0.00
max _{REV}	355	509	207	3.75	37.87	31.61	2.23	1.72
n_{FINAL}	3149	3149	3149	3149	3149	2961	3149	3149
mean _{FINAL}	87.01	19.08	46.71	1.13	0.91	0.79	0.33	0.27
St. dev _{FINAL}	58.32	26.41	38.08	0.49	0.99	0.68	0.10	0.06
min _{FINAL}	13.0	0.0	5.0	0.0	0.0	0.1	0.0	0.0
max _{FINAL}	338	420	372	3.248	11.14	6.65	1.151	0.6

Fig. 2 Comparison of kernel density plots between the tree catalogues for the network estimators (top row) and for the uncertainty estimators (bottom row). Bandwidths are selected according to Scott's rule (Scott 1992)

expected correlation with the number of stations. The same observation for the Italian catalogue has also been reported by Michele et al. (2019). Because of the lack of correlation between the estimators and the magnitude, it is fair to analyze the data regardless of the size of the earthquakes.

2.2 Network estimators

The network estimators provide information about the local network (i.e. the set of stations that contributed to locating the earthquake) for each event and can be considered as input specifications in the location procedure. They are as follows: seismic gap (GAP), minimum distance (MIN), number of stations (STA), and average phase weights (WPH).

2.2.1 GAP

The azimuthal gap (GAP) is defined as the largest angle between two recording stations as seen from the epicentre. It represents a concise measure of the geometry of the network around each seismic event: high gap values ($>180^\circ$) usually indicate locations of reduced quality. The average GAP value increases for 84% of the events passing from *AUTO* to *REV* (average of 43°); conversely it decreases from *REV* to *FINAL* for 68% of the events (average of -38°).

2.2.2 MIN

The minimum distance (MIN) is the least among the distances between the calculated epicentre and the

Table 2 Pearson's correlation coefficients between earthquake magnitude and the estimators from the three locations. STA is the only estimator with a certain degree of correlation with the earthquake magnitude

	GAP	MIN	STA	WPH	ERH	ERZ	RMS	ERM
M_{AUTO}	0.06	0.3	0.47	-0.07	0.17	0.16	0.37	0.11
M_{REV}	0.02	0.27	0.49	0.24	0.13	0.2	0.34	0.04
M_{FINAL}	-0.21	0.03	0.63	0.13	-0.11	-0.14	-0.11	0.08

receiving stations. If MIN is of the same order as the earthquake depth, it ensures a robust constraint on the hypocentre depth. MIN has similar average and standard deviation values among the catalogues (Fig. 2). Only approximately 21% of the events have MIN > 20 km and it is unchanged for most of the events passing from *AUTO* to *REV*, while from *REV* to *FINAL*, it is lower for 75% of the events (average of -7.4 km). Overall, MIN is lower in the *FINAL* than in the *REV* (average value of -2.9 km). This happens because phase readings from the closest stations are usually included in this step.

2.2.3 STA

The number of stations used in the location procedure (STA) is another factor that could affect the quality and the precision of the locations. It depends on the density of the network around each event, but it is also controlled by the earthquake magnitude. The distribution of STA is very similar throughout *AUTO* and *REV* catalogues (Fig. 2), while it is lower for 68% of the events (18 fewer stations on the average) in the *FINAL* one.

2.2.4 WPH

The phase weight is a quality index that provides a measure of the clarity of the readings of the arrivals of the seismic phases (both P- and S-phases) essentially on the basis of the background noise level at the event onset (see Allen (1978) for the details). Good-quality data with clear wave arrivals are given the quality label “0”; labels “1”, “2”, and “3” are assigned to increasingly ambiguous readings. When the temporal uncertainty of the phase arrival is too high, it is given the quality label “4”. The lower the average quality of phase arrivals (WPH) is, the more accurate the location will be. For each event, the labels can be different from station to station. WPH almost exclusively falls between “1” and “2” in the *AUTO* catalogue (Fig. 2), certainly because the automatic procedures preferentially assign the scores “1” and “2”. Moving to the *REV* catalogue, the data are more widespread and usually (72% of the events) have higher values (i.e. are assigned lower quality); in the passage from *REV* to *FINAL*, the values of WPH are

again lower. In general, the quality is better for 83% of the events with average gain of -0.69 . This factor is strongly controlled by the revision procedure in which weights can be re-assigned and new phases included in the location even with higher weights (i.e. lower quality), while some other phases may be excluded.

2.3 Uncertainty estimators

The main outputs of an earthquake location are the coordinates, the origin time, and the magnitude of the event. Their associated uncertainties provide information about the precision and reliability of locations themselves which are dependent on the above described “network estimators”.

2.3.1 ERH and ERZ

The horizontal and vertical errors (ERH and ERZ, respectively) represent the standard confidence intervals, expressed in kilometers, of the horizontal and vertical location coordinates, respectively. They are usually of the order of a few kilometres; values are lower and less scattered moving from *AUTO* to *REV* and *REV* to *FINAL* (Table 1 and Fig. 2). Please note that for ERZ the number of data is reduced in *REV* and in *FINAL* because all the events in which the depth was forcibly fixed (only in *REV* and *FINAL*) have been neglected. Such cases arise when the depth cannot be sufficiently constrained (i.e. no stations close to the epicentre), and therefore, a default value is assigned in order to compute the location. Such case occurs for 9.6% of events in *REV* and for 6.0% in *FINAL*. These events have been excluded in all the following analyses that consider the vertical uncertainty ERZ. ERH and ERZ have very similar density distributions (Fig. 2), and the values decrease almost systematically passing from *AUTO* to *FINAL* (79.1% and 25% respectively). The average improvement is greater from *AUTO* to *REV* (1.9 km and 5.2 km, respectively) than from *REV* to *FINAL* (0.7 km and 0.3 km, respectively). Only a few events (<4%) have a worse estimation of the epicentre coordinates in the *FINAL* catalogue rather than in the *AUTO* one. As we will observe later, the values of ERH and ERZ are unusually higher in the lower Tyrrhenian

sea, because of the deep seismicity occurring in this area (Fig. 1).

2.3.2 RMS

For each event, the origin time represents the best estimation of the moment in which the earthquake nucleated and the seismic waves started to propagate. When the origin time is well determined, the precision is usually of the order of a tenth of second or even lower. The root mean square (RMS) is defined as the arithmetic mean of the squares of the temporal residuals between the observed arrival times at the stations and the calculated arrival times given the velocity model. Indicating the misfit between the theoretical data and the empirical observation, the RMS (in seconds) is a measure of the accuracy of the location. RMS values are lower and less scattered passing from the *AUTO* to the *REV* and *FINAL* locations (Table 1 and Fig. 2). The RMS improves from *AUTO* to *REV* for the most of the events (-0.61 s on average), except for a small group of events (approximately 9%); the improvement is not so sharp from *REV* to *FINAL* (for 55% of the events and an average improvement of 0.13 s).

2.3.3 ERM

The earthquake magnitude provides an estimate of the energy associated with the seismic event. For most of the events, the catalogues report the local magnitude (M_L) based on measures of the maximum amplitudes of the signals normalized according to their epicentral distance. A measure of uncertainty (ERM) is associated with the earthquake magnitude estimation. It corresponds to the confidence interval of the magnitude values observed at the various stations, and therefore, it depends on the number and on the geometry of the stations that observed the event. The average ERM is lower passing from both *AUTO* to *REV* (95%, average value of -0.13), while the ERM distribution is essentially unchanged from *REV* to *FINAL*. It is noteworthy that the difference in the earthquake magnitude values for the same event between the catalogues can be much higher than ERM. In fact, the ranges of the absolute differences in magnitude are 0–2.54 and 0–0.24, respectively, for the *AUTO-REV* and *REV-FINAL* steps. The corresponding ranges of ERM are 0–0.6 and 0–0.7

3 Methodology

In this paper, data are analyzed by means of descriptive and spatial statistics. The methodologies are briefly described in the following paragraphs in terms of formula, meaning, and usage with a particular focus on the proposed case study. The analysis is performed with the statistical software R (R Development Core Team 2005), exploiting the libraries by Baddeley and Turner (2005) and Chiodi and Adelfio (2014).

3.1 Spatial interpolation

Spatial interpolation is a procedure to calculate unobserved values of a function defined in a finite spatial region given a set of observations with known location. Inverse distance weighting (IDW) is a deterministic method for spatial interpolation that assumes that each observation has a local influence that is inversely related to the distances between the prediction location and the sampled observations. The spatial prediction at a given position is calculated as the weighted average of a subset of observations that includes all the observations within a given distance (neighbourhood), a fixed number of minimum observations, or both.

IDW is usually adopted to interpolate highly variable data and it is more sensitive to their pattern with respect to other interpolation methods (Shepard 1968; Lu and Wong 2008). The assumption that close observations are more alike than those that are farther apart is appropriate in the present case study where the values to interpolate are mainly constrained by the configuration of a fixed observational network. Moreover, IDW is relatively fast to compute and the results are straightforward to interpret. In this specific case study, IDW is more easier to generalize for all the estimators at once, allowing their comparison.

Considering a subset of observations $\mathcal{N}(x) \subset \{x_1, x_2, \dots, x_n\}$ for the neighbourhood of $x \in \mathcal{W}$, the relative weights at each location x are:

$$w_i(x) = \delta(x, x_i)^{-p} > 0 \quad (1)$$

where $\delta(x, x_i)$ are the distances between x and the neighbouring observations $x_i \in \mathcal{N}(x)$. The weights are normalized to sum to 1 constraining:

$$\omega_i(x) = \frac{w_i(x)}{\sum_{x_i \in \mathcal{N}(x)} w_i(x)} \quad (2)$$

The interpolated value $\hat{y}(x)$ will be a weighted combination of the subset of observations as follows:

$$\hat{y}(x) = \sum_{x_i \in \mathcal{N}(x)} \omega_i(x) y_i \quad (3)$$

Performing IDW requires the selection of the power $p \geq 1$ and the definition of a search procedure to determine the neighbourhood (1). The exponent p is a parameter that controls the strength of the distance-decay relationship: small values return predictions as averages of the observations, while large values give greater weights to the nearest observation.

To define the neighbourhood, we explored the spatial autocorrelation structure of the data. In general, the level of spatial autocorrelation decreases as spatial lag increases. The upper distance bound has been determined by computing and modelling the experimental semivariograms of the estimators (spherical model) to infer the distance at which the semivariance no longer depends on the inter-distances between the observations. However, since the observations are not evenly distributed, we also considered a maximum number of observations within the search distance in order to also resolve the areas where the observations are more clustered. Moreover, a limited number of observations prevents to weighting distant observations when the closest ones are available, reducing the residuals of the model.

As the error estimation, the root mean square error (RMSE) has been calculated for each IDW model to check its accuracy. RMSE is the normalized mean of the squared difference between the observations and the predicted values, given by the following: $RMSE = \sqrt{\sum_{i=1}^n (y_i - \hat{y}_i)^2 n^{-1}}$ where y_i is the observed value for the i -th observation and $\hat{y}_i$ is the calculated one.

3.2 Spatial correlation

As illustrated in the previous paragraph, the considered estimators can be subdivided into two groups: one pertains to the local network, and the other pertains to the uncertainties of the locations. The estimators in each group have been combined to assess the spatial dependence between the two groups.

The network (I_N) and quality (I_Q) indexes for each event e are calculated according to the relationship

proposed by Michele et al. (2019) exactly for seismic events located by the Italian Seismic Network:

$$I(e) = \sqrt{\frac{\sum_{i=1}^N (w_i y_{norm(e)})^2}{N}} \quad (4)$$

where y_{norm} is the value of the n estimators, which has been normalized to its maximum value, and w is the weight associated with each estimator; N represents the number of employed estimators. As suggested by Michele et al. (2019), we assumed $w = 1$ for all the estimators. In contrast to all the other estimators, the value of STA was normalized to its minimum value, rather than its maximum. The values of $I_N(e)$ and $I_Q(e)$ range in the interval between 0 and 1. The greater the value of I_N is, the lower the “capability” of the network; the greater the value of I_Q is, the higher will be the uncertainty associated with a given event.

After the calculation of I_N and I_Q , we removed the outliers by means of the interquartile range (IQR) method. IQR is the difference between the 75th and the 25th percentile of a distribution. IQR identifies as an outlier an observation outside the interval $\pm 1.5 * IQR$. Later, considering the IDW models for both I_N and I_Q , we checked for spatial variations of their correlation. Around each map unit, we defined a focal squared area of 25×25 cells and recorded the correlation between all of the 625 pairs of values in the central cell with a procedure recently adopted by Siino et al. (2020) for the evaluation of monitoring networks.

All the computed spatial models are shown only where the coverage of the observations can guarantee their reliability. Three windows (one for each catalogue) have been drawn after calculating the 2D kernel intensity of the respective point patterns (i.e. earthquake distributions) considering a Gaussian kernel, a bandwidth according to Scott’s rule (Scott 1992), and a density threshold of 5 events per unit area. At this density value, the contours cover approximately the same region for the three catalogues, and it is the highest density value at which contour lines do not present enclaves.

4 Results and discussion

IDW models have been calculated for all the estimators, with the following details. The IDW models are

mapped with a grid size of $0.05^\circ \times 0.05^\circ$; the corresponding map unit measures approximately 23 km^2 . The grid size is congruous with the adopted IDW parameters and small enough to image the expected spatial variations. Because we are interested in large-scale features, we fixed $p = 1$ (1): the influence of the observations increases in the surroundings of the observations, highlighting variations at a smaller scale. The value of the search radius is the average range of the omnidirectional semivariogram model of the estimators, and it is equal to 0.95° ; the value has been rounded to 1° . The choice of the maximum

number of observations is a trade-off between residual error and desired resolution: the range of values between 5 and 50 has been tested, and the value has been set to 20.

The IDW models of the estimators include the outliers that have not been removed before the calculation of the models. Especially for the *AUTO* catalogue (Fig. 2), the outliers represent a considerable portion the data set, and they are precisely the direct result of the compilation procedure of the catalogue itself. In fact, the extreme values are clustered together rather than randomly distributed throughout the study area.

Fig. 3 IDW models of the network estimators: GAP, MIN, STA, and WPH, and the values of RMSE are reported. The black triangles indicate the seismic network. See the text for details

This is also true also for *REV* and *FINAL* catalogues, even though to a lesser extent.

The IDW models for all the estimators are shown in Figs. 3 and 4. The colour bars were conveniently set in order to allow the comparison among the three different models for each estimator (their ranges do not always overlap exactly) and to highlight variations at different scales. The higher variability of the *AUTO* locations is also reflected in the values of RMSE that are systematically higher with respect to the *REV* and *FINAL* locations.

For the network estimators, GAP and MIN mainly show a clear concentric pattern with lower values in the inner part of the Italian territory and lower values moving outward to the coast and beyond; these reflect the seismic station distribution (see Fig. 1a). Conversely, STA and WPH mainly show a patchy pattern (Fig. 3); in both cases, there are some exceptions. The general pattern across the three catalogues is similar; however, there are some differences at a smaller scale. The values of WPH_{AUTO} are almost constant, the pattern only emerges after the procedure of

Fig. 4 IDW models of the uncertainty estimators: ERH, ERZ, RMS, and ERM, and the values of RMSE are reported. The black triangles indicate the seismic network. See the text for details

revision. Finally, most of the maps show a region in the central part of northern Italy that clearly contrasts with the general pattern.

For the uncertainty estimators, the most relevant feature is the evident difference between the *AUTO* catalogue with respect to the two others. The former shows a patchy pattern with abrupt changes of the value, while *REV* and *FINAL* catalogues show a more regular pattern with smoother variations at a larger scale that are consistent with each other (Fig. 4). In the *REV* and *FINAL* catalogues, the values range within a relatively narrow band, except for a restricted area in southern Italy for ERZ. The anomalous area in ERH_{REV} and ERH_{FINAL} (Fig. 4, 1st row), with higher values than its surroundings, may result from the combined effect of the following: (i) higher GAP values (Fig. 3) and (ii) deep seismicity occurring in the area (Fig. 1a). In particular, we consider the latter effect as the most likely since other regions with high GAP values occur elsewhere. Therefore, the deep seismicity results in relatively higher horizontal errors (ERH) rather than in vertical ones (ERZ). Finally, the anomalous area in the central part of northern Italy emerges again with a different intensity in the four estimators (especially in the *REV* catalogue).

We also computed the IDW models for the epicentre shifts, i.e. the change of position of the epicentre after the two steps of revision (Fig. 5). Despite that ERH for the *AUTO* location is of a few km on average (2.2 km), the shift of the epicentres moving from

AUTO to *REV* is of 13.1 km on average, while the average shift from *REV* to *FINAL* (3.3 km) is of the same order of magnitude as the associated uncertainty. The azimuthal distribution of the epicentre shift through the whole revision procedure (*AUTO* to *FINAL*) is homogeneous, without evidence of preferential directions (Fig. 5, right). Because of the narrow and long shape of Italy that constrain the shape of the seismic network, we would have expected the occurrence of some preferential directions in the epicentre shift. Conversely, the azimuthal distribution is almost uniform, suggesting that the effects of the revisions are independent of the network configuration, at least at the country scale.

At the local scale, the behaviour could be different. For this reason, we first calculate the spatial IDW models for I_N and I_Q that combine the two groups of estimators by means of Eq. 4 and then evaluate their local correlation. For the IDW models, the values of $p = 1$ and $n = 20$ are the same as those used to compute the spatial models of the estimators. Conversely, the research radius has not been fixed because it has been difficult to obtain good theoretical variograms. The IDW models are shown in Fig. 6. The I_N displays a rough concentric pattern, as also suggested by the contour lines, with lower values in the inland regions and higher values moving outwards; some local variations occur at a smaller scale. The I_Q is irregularly distributed on the territory, but variations are relatively smooth. A region of relatively higher values of I_Q

Fig. 5 IDW models of the epicentral shift between the two steps of the location's refinement. The black triangles indicate the seismic network, and the values of RMSE are reported. The

bars represent the frequency distribution of the direction of the epicentral shift from *AUTO* to *FINAL*, and the horizontal line indicates the average value

Fig. 6 IDW models of I_N (left column) and I_Q (right column) calculated with Eq. 4: the lower the value is, the better the index. The contour lines (dashed lines) refer to the mean value of each model, and the values of RMSE are reported. The black triangles indicate the seismic network

marks the area characterized by deep seismicity in the southern Tyrrhenian Sea (Fig. 1a).

The local correlation between the three pairs of I_N and I_Q was calculated with a focal area of 25×25 maps units. This resolution has been selected to highlight features at the scale of hundreds of km. We aim to explain the results in relation to the Italian National Seismic Network; a finer resolution of the local correlation would not be consistent with the scale of the network. The overall Pearson's correlation coefficients between I_N and I_Q are 0.073, 0.081, and -0.018 for *AUTO*, *REV*, and *FINAL*, respectively. Therefore, the

two indexes are totally uncorrelated, but locally, the correlation emerges (Fig. 7). As a reference to identify the areas with higher correlation, we have drawn the contour lines at -0.5 (red) and $+0.5$ (blue).

The negative correlation can be due to the combination between high I_N and low I_Q and the inverse scenario. The two cases are as follows: (i) increasing the reliability of the locations despite the decay of the network capability and (ii) increasing the network capability despite decreasing the reliability of the locations. The positive correlation arises from the combination of both increasing (or decreasing) both

Fig. 7 Local correlation between I_N and low I_Q , as in Fig. 6, computed with a focal area of 25×25 map units for the three catalogues. The contour lines are at values of ± 0.5 ; the stars indicate the detected anomalous areas

I_N and I_Q . Therefore, (i) a lower reliability of the location is obtained with a decrease of the network capability, and (ii) an improvement of the locations is obtained with an increase of the network potential. At a glance, the local correlation maps show an increasing negative correlation moving from the *AUTO* to the *FINAL* catalogue (Fig. 7). The possible explanation relies on the progressively improved estimations of the locations, fixed the network estimators.

It is not possible to discriminate between the two pairs of combinations that lead to the same conclusion (i.e. negative or positive correlation); however, some considerations can be achieved. First, it is more interesting to focus on the areas with negative correlation to explore the cause of the inconsistency between the two indexes, which, in contrast, can be considered, the normal (and desired) condition. Then, taking into account that the general spatial distribution of I_N is approximately concentric (Fig. 6), it is possible to separate the red areas lying in the inner part of the study area from the other red areas lying in the outer, peripheral regions. The former among these two cases represents the advantageous situation of good locations despite the poorer network configuration; the second case represents the unfavourable situation of bad locations notwithstanding a good network configuration. Obviously, the latter condition should be avoided. Therefore, the areas where this condition is verified should be considered target areas of where to improve the reliability of the seismic network. Such

areas are marked with black stars in Fig. 7; some of them appear in all of the catalogues, others arise only in a single catalogue.

5 Conclusions

In this paper, we address with the progressive earthquake locations calculated with the procedure adopted for seismic surveillance in Italy. The procedure encompasses the production of three different locations with increasing refinement: the *AUTO* ones, the *REV* ones, and the *FINAL* ones. We combined the information pertaining to the local network (input data of the earthquake location) with the information pertaining to the quality of the location (output of the earthquake location). The main aim is to understand what controls the uncertainty of the locations and to also discuss the differences between the catalogues. Because the considered estimators are uncorrelated with the earthquake magnitude, data are analyzed under the assumption of independence of the size of the earthquakes. The spatial distribution of the network estimators depends on the general arrangement of the monitoring stations, whereas the uncertainty estimators are mainly controlled by other (local) factors. Furthermore, even the reliability of uncertainties themselves can be investigated if, as already discussed, the average epicentre shift is greater than the horizontal error. This is out of the scope of this paper,

which rather aims to obtain insights about their controlling factors and, in turn, address the optimization of the network and of the location procedures. Data indicate that, generally, the Italian Seismic Network is able to provide a uniform spatial response; small-scale variations of the quality of the location are not ascribable to the network distribution (nor to the size of the earthquakes), rather are ascribable to local causes. The stars in Fig. 7 indicate the areas where the performance of the locations can be improved. These areas partly correspond to the region of weakness of the Italian Seismic Network highlighted by Siino et al. (2020).

The calculation of an earthquake hypocentre is an inverse problem based on the arrival times of the seismic phases at the nodes of the seismic network and geophysical models of the Earth's crust. We suppose that the lower accuracy encountered in some areas (Fig. 7) is the systematic effect introduced by the models used to represent the medium in which the seismic waves propagate. Consequently, what is actually controlling the accuracy of locations is the ability to model the heterogeneous geodynamic setting of the study area. In summary, we can assume that the quality of the locations is controlled by the geological pattern of Italy, which is rather complex and cannot be accounted for in the basic 2D velocity model. This model inevitably suffers from a lack of accuracy, which, in turn, affects the quality of the locations at a local scale.

The three locations have increasing precision, whereas systematic errors can still affect their accuracy. This is clearly demonstrated comparing the values of the epicentre shift (and the earthquake magnitude difference) with the corresponding error in the estimate ERH (and ERM). The differences (i.e. accuracy) can be easily one order of magnitude greater than the errors (i.e. precision). The improvement in the first step (i.e. from *AUTO* to *REV*) is systematic and clearer than the second step (i.e. from *REV* to *FINAL*). Even the *AUTO* location itself can be considered a relatively good-quality location for seismic surveillance at the national scale because of the good performance reached by the automatic wave picking and inversion algorithms (Spallarossa et al. 2014; Scafidi et al. 2018; Huang and Wu 2019). The comparison shows that the weaker areas detected in the *AUTO* stage are consistent with the other stages (Fig. 7). At the same time, other weak areas can emerge into a single stage of

the location procedure. However, we can conclude that with its present-day configuration, the Italian Seismic Network is able to provide robust information independent of the size of the earthquake with a good coverage over the whole country.

In this framework, the priority for the homogenization of the quality of the locations would be the assessment of more reliable, and locally optimized, velocity models rather than increasing the number of nodes of the network (Latorre et al. 2016). Optimized velocity models can be assessed even with small and temporary networks and they would be especially desirable in off-shore regions (Molinari et al. 2015; D'Alessandro et al. 2016; Latorre et al. 2016; Marchetti et al. 2016; Frepoli et al. 2017), while a further contribution could be provided by seismic stations able to increase the signal-to-noise ratio (e.g. downhole stations; Kinnaert et al. 2018).

Funding Open access funding provided by Istituto Nazionale di Geofisica e Vulcanologia within the CRUI-CARE Agreement.

Data and resources All data used in this work were collected by the Istituto Nazionale di Geofisica e Vulcanologia (INGV), Osservatorio Nazionale Terremoti. Earthquake data and the “Bollettino Sismico Italiano” can be downloaded from the websites: http://cnt.rm.ingv.it/webservices_and_software and <http://terremoti.ingv.it/>.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Allen RV (1978) Automatic earthquake recognition and timing from single traces. *Bull Seismol Soc Am* 68(5):1521–1532
- Allen R (1982) Automatic phase pickers: their present use and future prospects. *Bull Seismol Soc Am* 72(6B):S225–S242

- Amato A, Mele F (2008) Performance of the ingv national seismic network from 1997 to 2007. *Ann Geophys* 51(2/3):417–431
- Baddeley A, Turner R (2005) Spatstat: an r package for analyzing spatial point patterns. *J Stat Softw* 12(6):1–42. <https://doi.org/10.18637/jss.v012.i06>
- Cattaneo M, Frapiccini M, Ladina C, Marzorati S, Monachesi G (2017) A mixed automatic-manual seismic catalog for central-eastern Italy: analysis of homogeneity. *Ann Geophys*
- Chiarabba C, Jovane L, DiStefano R (2005) A new view of Italian seismicity using 20 years of instrumental recordings. *Tectonophysics* 395(3–4):251–268
- Chiarabba C, De Gori P, Mele FM (2015) Recent seismicity of Italy: active tectonics of the central mediterranean region and seismicity rate changes after the mw 6.3 l'aquila earthquake. *Tectonophysics* 638:82–93
- Chiodi M, Adelfio G (2014) etasflp: estimation of an etas model. mixed flp (forward likelihood predictive) and ml estimation of non-parametric and parametric components of the etas model for earthquake description. R package version 10
- D'Alessandro A, Ruppert NA (2012) Evaluation of location performance and magnitude of completeness of the Alaska regional seismic network by the snes method. *Bull Seismol Soc Am* 102(5):2098–2115
- D'Alessandro A, Stickney M (2012) Montana seismic network performance: an evaluation through the snes method. *Bull Seismol Soc Am* 102(1):73–87
- D'Alessandro A, Papanastassiou D, Baskoutas I (2011) Hellenic unified seismological network: an evaluation of its performance through snes method. *Geophys J Int* 185(3):1417–1430
- D'Alessandro A, Dăneț A, Grecu B (2012) Location performance and detection magnitude threshold of the romanian national seismic network. *Pure Appl Geophys* 169(12):2149–2164
- D'Alessandro A, Badal J, D'Anna G, Papanastassiou D, Baskoutas I, Özel NM (2013) Location performance and detection threshold of the Spanish national seismic network. *Pure Appl Geophys* 170(11):1859–1880
- D'Alessandro A, Mangano G, D'Anna G, Scudero S (2016) Evidence for serpentinization of the ionian upper mantle from simultaneous inversion of p-and s-wave arrival times. *J Geodyn* 102:115–120
- Frepoli A, Cimini G, De Gori P, De Luca G, Marchetti A, Monna S, Montuori C, Pagliuca N (2017) Seismic sequences and swarms in the Latium-Abruzzo-Molise Apennines (central Italy): new observations and analysis from a dense monitoring of the recent activity. *Tectonophysics* 712:312–329
- Gulia L (2010) Detection of quarry and mine blast contamination in European regional catalogues. *Nat Hazards* 53(2):229–249
- Huang TC, Wu YM (2019) A robust algorithm for automatic p-wave arrival-time picking based on the local extrema scalogram robust algorithm for automatic p-wave arrival-time picking based on the local extrema scalogram. *Bull Seismol Soc Am* 109(1):413–423
- Johnson CE, Bittenbinder A, Bogaert B, Dietz L, Kohler W (1995) Earthworm: a flexible approach to seismic network processing. *Iris News* 14(2):1–4
- Kinnaert X, Gaucher E, Kohl T, Achauer U (2018) Contribution of the surface and down-hole seismic networks to the location of earthquakes at the Soultz-sous-Forêts geothermal site (France). *Pure Appl Geophys* 175(3):757–772
- Latorre D, Mirabella F, Chiaraluce L, Trippetta F, Lomax A (2016) Assessment of earthquake locations in 3-d deterministic velocity models: a case study from the altotiberina near fault observatory (Italy). *J Geophys Res: Solid Earth* 121(11):8113–8135
- Lomax A, Virieux J, Volant P, Berge-Thierry C. (2000) Probabilistic earthquake location in 3d and layered models. In: *Advances in seismic event location*. Springer, pp 101–134
- Lu GY, Wong DW (2008) An adaptive inverse-distance weighting spatial interpolation technique. *Comput Geosci* 34(9):1044–1055
- Mahani AB, Kao H, Walker D, Johnson J, Salas C (2016) Performance evaluation of the regional seismograph network in northeast British Columbia, Canada, for monitoring of induced seismicity. *Seismol Res Lett* 87(3):648–660
- Marchetti A, Barba S, Cucci L, Pirro M (2004) Performances of the Italian seismic network, 1985–2002: the hidden thing. [arXiv:0405032](https://arxiv.org/abs/0405032)
- Marchetti A, Ciaccio MG, Nardi A, Bono A, Mele FM, Margheriti L, Rossi A, Battelli P, Melorio C, Castello B et al (2016) The Italian seismic bulletin: strategies, revised pickings and locations of the central Italy seismic sequence. *Ann Geophys* 59(5):1–7. <https://doi.org/10.4401/ag-7169>
- Mele F, Arcoraci L, Battelli P, Berardi M, Castellano C, Lozzi G, Marchetti A, Nardi A, Pirro M, Rossi A (2010a) Bollettino sismico Italiano 2008. *Quad Geofis* 85:1–48. <https://www.earth-prints.org/handle/2122/6159>
- Mele F, Bono A, Lauciani V, Mandiello A, Marcocci C, Pintore S, Quintiliani M, Scognamiglio L, Mazza S (2010b) Tuning an earthworm phase picker: some considerations on the pick_ew parameters
- Michele M, Latorre D, Emolo A (2019) An empirical formula to classify the quality of earthquake locations. *Bull Seismol Soc Am* 109(6):2755–2761
- Michellini A, Margheriti L, Cattaneo M, Cecere G, D'Anna G, Delladio A, Moretti M, Pintore S, Amato A, Basili A et al (2016) The Italian national seismic network and the earthquake and tsunami monitoring and surveillance systems. *Adv Geosci* 43:31–38. <https://doi.org/10.5194/adgeo-43-31-2016>
- Molinari I, Argani A, Morelli A, Basini P (2015) Development and testing of a 3d seismic velocity model of the po plain sedimentary basin, Italy. *Bull Seismol Soc Am* 105(2A):753–764
- Möllhoff M, Bean CJ, Baptie BJ (2019) Sn-cast: seismic network capability assessment software tool for regional networks-examples from ireland. *J Seismol* 23(3):493–504
- Peng C, Yang J, Wang W, Zheng Y, Jiang X, Han J, Tian B, Xu Z (2018) The Namche Barwa Temporary Seismic Network (NBTSN) and its application in monitoring the 18 November 2017 m 6.9 Mainling, Tibet, China, earthquake. *Seismol Res Lett* 89(5):1730–1740
- R Development Core Team (2005) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. <http://www.R-project.org>, ISBN 3-900051-07-0

- Saunders I, Kijko A, Fourie C (2016) Statistical evaluation of seismic event location accuracy by the south African national seismograph network over four decades. *South Afr J Geol* 119(1):291–304
- Scafidi D, Viganò A, Ferretti G, Spallarossa D (2018) Robust picking and accurate location with rsni-picker2: real-time automatic monitoring of earthquakes and nontectonic events. *Seismol Res Lett* 89(4):1478–1487
- Schorlemmer D, Mele F, Marzocchi W (2010) A completeness analysis of the national seismic network of Italy. *J Geophys Res: Solid Earth* 115(B04308). <https://doi.org/10.1029/2008JB006097>
- Schumann R, Owen D, Asher-Bolinder S (1988) Factors affecting soil-gas radon concentrations at a single site in the semiarid western us. In: *Proceedings of the 1988 EPA symposium on radon and radon reduction technology*, p 2. Publication, Citeseer
- Scott D (1992) Density estimation: theory, practice and visualization. *The Curse of Dimensionality and Dimension Reduction*, pp 195–217
- Shepard D. (1968) A two-dimensional interpolation function for irregularly-spaced data. In: *Proceedings of the 1968 23rd ACM national conference*, pp 517–524
- Siino M, Scudero S, Greco L, D'Alessandro A (2020) Spatial analysis for an evaluation of monitoring networks: examples from the Italian seismic and accelerometric network. *J Seismol*, accepted <https://doi.org/10.1007/s10950-020-09937-0>
- Spallarossa D, Ferretti G, Scafidi D, Turino C, Pasta M (2014) Performance of the rsni-picker. *Seismol Res Lett* 85(6):1243–1254
- Tarantola A, Valette B (1982) Generalized nonlinear inverse problems solved using the least squares criterion. *Rev Geophys* 20(2):219–232
- Turino C, Morasca P, Ferretti G, Scafidi D, Spallarossa D (2010) Reliability of the automatic procedures for locating earthquakes in southwestern Alps and northern Apennines (Italy). *J Seismol* 14(2):393–411

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.