MDPI

Article

Bias in the Estimation of Seismic Risk for Municipal Building Stocks Due to Limited Data

Anže Babič *, **Jure Žižmond** and **Matjaž Dolšek**

Faculty of Civil and Geodetic Engineering, University of Ljubljana, 1000 Ljubljana, Slovenia;
jzizmond@fgg.uni-lj.si (J.Ž.); mdolsek@fgg.uni-lj.si (M.D.)
* Correspondence: ababic@fgg.uni-lj.si

Abstract: This study investigated the effect of the building data knowledge level on seismic risk estimation for municipal building stocks, focusing on identifying the characteristics that influence loss estimation bias. Fifteen municipalities in two Slovenian regions were analysed using twelve building data knowledge levels, defined by combining different knowledge levels about building location and floor area. The knowledge levels ranged from those using data aggregated at the municipality level to those using building-specific data. The bias was quantified as the log residual between the expected annual losses estimated for the given knowledge level and the base-case level, characterised by building-specific data. The results indicate that loss estimation bias is affected by both the building location and floor-area knowledge levels. The data on building density distribution across the municipality and building-class-specific floor areas are sufficient for estimating loss with low bias with respect to the base-case level. The effect of potential data improvement on bias reduction can be assessed using building stock homogeneity and hazard variation indexes determined from readily available data. Further research is needed to explore loss estimation bias for building data knowledge levels not considered in this study and generalise the concepts to other regions and building classifications.

1. Introduction

Information about the seismic risk to building stocks at the level of administrative units such as municipalities supports decision making about strengthening the community against earthquakes. Because of the numerous buildings considered in such risk estimation, the buildings are typically described by a few essential characteristics needed to construct the building stock exposure model considered in risk assessment (e.g., [1–3]). These characteristics of buildings can be defined at different knowledge levels. Building-specific data, accounted for in some studies (e.g., [3,4]), represent a relatively high building stock knowledge level. However, the knowledge available about the building stock is usually less comprehensive. It is much more common that only aggregated building data are readily available for use in seismic risk assessment (e.g., [5]). In such cases, additional bias can be introduced into the seismic risk estimation.

The effect of the building data knowledge level on seismic risk assessment has been investigated in several studies. Bal et al. [6] studied the effect of exposure data and ground-motion field resolution on event loss for an idealised city and a region in Turkey. They found that a crude resolution level does not imply a significantly biased result. A similar study by Dabbeek et al. [7] explored the effect of exposure data resolution on the expected annual loss (*EAL*) at the national and sub-national levels. In their study, the average bias at the national level was estimated at 27% with respect to the base-case results. Kalakonas et al. [8] investigated the effects of various exposure model resolutions on risk estimation

for Guatemala. They concluded that a coarse resolution at an urban level should be refined to avoid biased results. The importance of the geographical scale used to represent the building exposure model was also highlighted by Douglas [9] and Ordaz et al. [10]. Bazzurro and Park [11] studied the effect of location data aggregation for a selected building type in California. They found that aggregating data affected the loss curve more than the *EAL*. Senouci et al. [12] explored different building aggregation options and their effects on the uncertainties of event loss estimates. Sanderson and Cox [13] compared the losses estimated using national and local building inventories and found that using the national inventory underpredicted the losses aggregated at the city level. Some studies have focused on the feasibility of developing fragility and vulnerability models using limited data. Basic fragility and vulnerability models can be established using information from freely available databases (e.g., [14,15]), and additional data obtained using an interview-based approach (e.g., [16]) can further improve such models (e.g., [17–20]). Methods for incorporating new data into risk estimation [21,22] and improving the accuracy of risk estimation based on a given level of data [23–26] have also been developed.

Previous studies have shown that the bias in seismic risk estimation can be reduced by upgrading the building data knowledge level. However, collecting additional data is a demanding process that requires resources and is thus not always justified. The decision about whether upgrading the building data knowledge level is worth the effort requires an understanding of the effect of new data on the accuracy of risk estimation. Moreover, new data can affect bias reduction differently for different administrative units. Therefore, an understanding of which characteristics of administrative units are indicative of the degree of bias reduction attainable with new data is necessary for additional data collection to be effective.

The current study addressed the problems described by investigating the effect of gradually upgrading the building data knowledge level on seismic risk estimation, with the ultimate goals of (1) identifying which gradual improvements to the knowledge level significantly contribute to bias reduction and (2) developing indexes based on readily available data that can be used to estimate in advance the degree of bias reduction achievable by introducing new data. Fifteen municipalities in two Slovenian regions were analysed using each municipality's *EAL* as the risk indicator. The building data knowledge levels were defined by combining different knowledge levels about building location (i.e., the location knowledge level) and floor area (i.e., the floor-area knowledge level). The location knowledge level affects the *EAL* because of the potential bias in estimating the seismic hazard. The floor-area knowledge level determines the accuracy of defining the floor area of each building, which affects the outcome of the consequence assessment for the selected risk indicator.

Section 2 describes the seismic risk estimation procedure used in this study. Section 3 presents the building stock data for the investigated municipalities. Section 4 presents the method for determining the bias in seismic risk estimation for different building data knowledge levels. The method considers twelve building data knowledge levels, with the base-case level defined by building-specific data. The bias results are presented in Section 5, and the municipal characteristics affecting the bias are identified in Section 6.

2. Estimation of Seismic Risk for Municipal Building Stocks

The seismic risk for municipal building stocks was quantified in this study using the *EAL* as a measure of the direct economic loss resulting from physical damage to buildings. The *EAL* at the municipality level was obtained by summing the *EAL*s of the individual buildings in the municipality. For the base-case knowledge level, the *EAL*s of individual buildings were estimated using the available building-specific data. However, for lower knowledge levels, building data were adjusted to be consistent with the assumptions at that knowledge level. The assumptions for each knowledge level considered in this study are presented in Section 4.

The general methodology used for the estimation of *EAL* at the building level comprises (1) hazard assessment, (2) damage assessment and (3) consequence (loss) assessment and can be divided into the following sub-steps:

- Step 1: Assessment of seismic hazard in terms of seismic hazard curves $H_i(im)$, each representing the mean annual frequency of exceedance of a given intensity level (*im*) at the location of the *i*th building.
- Step 2a: Assessment of seismic fragility curves for the designated damage states of the *i*th building.
- Step 2b: Estimation of the mean annual frequencies of exceedance of the designated damage states $\lambda_i(DS \geq ds_d)$, where *d* is the index of the damage state.
- Step 2c: Estimation of the mean annual frequencies of occurrence of the designated damage states $\lambda_i(DS = ds_d)$. For the most severe damage state, $\lambda_i(DS = ds_d)$ is equal to $\lambda_i(DS \geq ds_d)$. For other damage states, $\lambda_i(DS = ds_d)$ is equal to the difference between $\lambda_i(DS \geq ds_d)$ and $\lambda_i(DS \geq ds_{d+1})$.
- Step 3a: Estimation of the expected economic losses given the occurrences of the designated damage states $C_i(DS = ds_d)$.
- Step 3b: Calculation of *EAL* by summing the products of $\lambda_i(DS = ds_d)$ and $C_i(DS = ds_d)$ over all designated damage states.

In the current study, the first step was performed by using the peak ground acceleration (PGA) as the ground-motion intensity measure. To assess the seismic hazard curves, the municipalities were first divided into grid cells with an approximate size of 0.5×0.5 km. The centre of each cell represented the calculation point for the assessment of the hazard curve, which was then applied to all buildings within the cell. The hazard curves, $H_i(pga)$, were defined as follows:

$$H_i(pga) = \lambda_i(PGA > pga), \tag{1}$$

where $\lambda_i(PGA > pga)$ is the mean annual frequency of the PGA exceeding the designated value, *pga*, for the grid cell containing the *i*th building. The hazard curves were calculated for soil type A [27] using the OpenQuake engine [28,29] and the ESHM2020 model [30]. For other soil types, the PGA values from the hazard curves for soil type A were transformed based on the soil factors proposed in the draft of the new Eurocode 8 [31], as in a previous study [32].

In the second step (damage assessment), five damage states were considered, consistent with the HAZUS methodology [33]: no-to-minor damage (DS0), slight damage (DS1), moderate damage (DS2), extensive damage (DS3) and complete damage (DS4). A detailed description of damage associated with each damage state can be found elsewhere [33]. The mean annual frequency of exceedance of the *d*th damage state for the *i*th building, $\lambda_i(DS \geq ds_d)$, was calculated as follows:

$$\lambda_i(DS \geq ds_d) = \int_0^\infty P_i(DS \geq ds_d | PGA = pga) \left| \frac{d\lambda_i(PGA > pga)}{dPGA} \right| dPGA, \tag{2}$$

where $P_i(DS \geq ds_d | PGA = pga)$ is the seismic fragility function of the *i*th building for the *d*th damage state, defined as the conditional probability of exceeding the *d*th damage state given the PGA, and $\left| \frac{d\lambda_i(PGA > pga)}{dPGA} \right|$ is the absolute value of the derivative of the seismic hazard curve at the location of the *i*th building.

The fragility functions were defined at the building class level, as is typical in large-scale seismic risk assessment (e.g., [3,33,34]). Twenty building classes were defined based on the load-bearing structure material, construction period and number of storeys (Table 1), as in a previous study [3,4].

Table 1. Definition of building classes based on the material of the load-bearing structure, construction period and number of storeys.

Building Class	Load-Bearing Structure Material	Construction Period	Number of Storeys
1		≤1964	
2		1965–1981	1–3
3	Masonry	≥1982	
4	(brick, stone)	≤1964	
5		1965–1981	4 or more
6		≥1982	
7		≤1964	
8		1965–1981	1–3
9		≥1982	
10		≤1964	4 or more
11	Reinforced concrete	1965–1981	4–6
12		≥1982	4–6
13		1965–1981	7 or more
14		≥1982	7 or more
15		≤1964	
16		1965–1981	1–3
17		≥1982	
18	Unknown, other	≤1964	
19		1965–1981	4 or more
20		≥1982	

In terms of the load-bearing structure material, a distinction between masonry (brick or stone) and reinforced concrete buildings was made since the vast majority of buildings in Slovenia are made of these materials. All other load-bearing structure materials (i.e., steel and timber buildings and buildings with mixed or unknown materials of the load-bearing structure) were grouped in a third material class. It is important to highlight that buildings with steel and timber structures were not segregated into distinct classes due to their minimal representation in the building stock.

A distinction was made between periods of construction based on standards for earthquake-resistant design in Slovenia and the former Yugoslavia, which extended to the territory of today's Slovenia until 1991 (Table 1). Buildings constructed up to 1964 were designed with no or minimal consideration for lateral loads and represented the first construction period category. The first building code that explicitly dealt with seismic actions was effective from 1964 until 1981 and was issued after the Skopje earthquake. The buildings built in this period constituted the second construction period category. The third category included buildings constructed after the implementation of the second generation of earthquake-resistant design codes in 1982. These regulations remained in effect until 2008 when Eurocodes became compulsory in Slovenia. Nonetheless, no distinction was made between buildings constructed between 1982 and 2008 and those constructed after 2008 due to the relatively limited number of buildings from the latter period and the already substantial earthquake resistance of buildings from the 1982–2008 timeframe.

In terms of the number of storeys, two or three classes per material and construction period were defined (Table 1), thus considering that the seismic performance is affected by the height of the building, mainly due to its direct effect on the building's vibration period. Buildings with up to three storeys, four to six storeys and seven storeys or more were classified as low-, medium- and high-rise buildings, respectively. This classification is similar to that used in the literature (e.g., [33,35,36]).

The data needed to classify buildings were obtained from the Slovenian Real Estate Register [37]. In addition to those data, building-specific parameters were obtained, including the buildings' coordinates and net floor areas, for use in the *EAL* calculation. These parameters, aggregated at the municipality level, are presented in Section 3.

For each building class, fragility functions were defined based on the fragility model presented by Babič et al. [4] in the form of a lognormal cumulative distribution function, which is typically assumed (e.g., [33,38]). However, a simpler definition of fragility functions was used than that presented by Babič et al. [4], who randomly simulated the fragility functions for a given building class to account for the uncertainty in the building class's median fragility and the variability in building fragility within the class. In this study, one fragility function per building class and damage state, defined as the average fragility function of those simulated in [4], was considered, thus disregarding the effect of the uncertainty in the fragility. There were two reasons for this simplification. Firstly, the study focused on the bias in *EAL* resulting from the building data knowledge level rather than the variability in *EAL* caused by the uncertainty in the fragility. Secondly, the bias in *EAL* was quantified by the log ratio of two *EAL* values (log residual) (see Section 4). Changes in the fragility functions have a lower impact on the *EAL* ratio than on individual *EAL* values because they partly cancel each other out. The medians and the lognormal standard deviations of the fragility function are presented in Table 2.

Table 2. Median (θ) and lognormal standard deviation (β) of the damage-state PGA for each building class and damage state considered in this study.

Building Class	θ_{DS1}	β_{DS1}	θ_{DS2}	β_{DS2}	θ_{DS3}	β_{DS3}	θ_{DS4}	β_{DS4}
1	0.13	0.44	0.20	0.47	0.38	0.50	0.71	0.52
2	0.18	0.45	0.27	0.48	0.53	0.51	0.98	0.52
3	0.19	0.44	0.30	0.47	0.58	0.50	1.07	0.52
4	0.12	0.46	0.17	0.51	0.34	0.55	0.62	0.58
5	0.17	0.45	0.24	0.51	0.46	0.55	0.85	0.57
6	0.21	0.45	0.29	0.50	0.55	0.54	1.03	0.57
7	0.10	0.44	0.18	0.47	0.39	0.50	0.80	0.52
8	0.14	0.44	0.25	0.47	0.52	0.50	1.07	0.52
9	0.17	0.44	0.30	0.47	0.64	0.50	1.31	0.52
10	0.13	0.47	0.22	0.54	0.49	0.60	0.95	0.63
11	0.19	0.46	0.31	0.53	0.69	0.59	1.34	0.62
12	0.22	0.45	0.36	0.53	0.81	0.58	1.56	0.61
13	0.16	0.48	0.32	0.59	0.76	0.67	1.47	0.72
14	0.19	0.48	0.38	0.58	0.89	0.67	1.72	0.71
15	0.11	0.62	0.15	0.65	0.33	0.68	0.76	0.69
16	0.14	0.63	0.21	0.65	0.44	0.68	1.03	0.70
17	0.20	0.63	0.28	0.66	0.61	0.69	1.42	0.70
18	0.14	0.64	0.23	0.70	0.51	0.75	0.92	0.78
19	0.21	0.64	0.35	0.70	0.77	0.75	1.41	0.78
20	0.37	0.64	0.61	0.70	1.35	0.75	2.46	0.78

In the third step in estimating *EAL* at the building level, the *EAL* was calculated for each building of the municipality's building stock as follows:

$$EAL_i = \sum_{d=1}^{4} \lambda_i(DS = ds_d)C_i(DS = ds_d). \tag{3}$$

The expected direct economic loss for the *i*th building in the case of the *d*th damage state, $C_i(DS = ds_d)$, was calculated as follows:

$$C_i(DS = ds_d) = A_i \cdot C_R \cdot c_d. \tag{4}$$

where A_i is the net floor area of the *i*th building, C_R is the average building replacement cost per m^2 of the net floor area, and c_d is the ratio between the repair costs for the *d*th damage state and C_R. The areas A_i were obtained from the Real Estate Register [37] (see Section 3). The costs of new construction and the demolition and removal of the damaged building were taken into account in estimating C_R. The cost of new construction was estimated to be EUR 1100 per m^2 of net floor area, based on the construction costs for 2020 obtained from a

Slovenian construction cost database [39]. The estimated value includes a tax rate of 9.5%. The cost of the demolition and removal of the damaged building was estimated based on previous studies [40–42], in which it varied between 8% and 15% of the new construction cost. On that basis, the assumed cost of reconstruction was EUR 1250 per m^2 of the net floor area, which is 13.5% more than the cost of a new building. The c_d ratios were defined based on the HAZUS methodology [33]. For damage states DS1–DS4, they are equal to 0.02, 0.1, 0.4 and 1.0, respectively.

Finally, based on the values of *EAL* estimated for each building (Equation (3)), the *EAL* at the municipality level was calculated as follows:

$$EAL = \sum EAL_i, \tag{5}$$

thus considering that the *EAL* for a portfolio of properties can be obtained by summing the *EAL*s estimated for each of the properties (e.g., [43]).

3. Municipalities Investigated and Their Seismic Hazard

The municipalities investigated are located in two Slovenian regions (Figure 1a). The first region (the Eastern region) is located in the northeastern part of Slovenia, close to the Austrian border. It contains nine municipalities, i.e., Tišina, Radenci, Gornja Radgona, Apače, Sveta Ana, Benedikt, Šentilj, Pesnica and Kungota (see Figure 1b). The second region (the Western region) is located in the western part of the country, near the Italian border. It contains six municipalities, i.e., Kanal, Brda, Nova Gorica, Šempeter–Vrtojba, Renče–Vogrsko and Miren–Kostanjevica (Figure 1c). The two regions are approximately 200 km apart.

The knowledge level of the data on the building stock in Slovenia is relatively good since quite a significant amount of building characteristics data is stored in the Real Estate Register [37]. Some of the basic building characteristics aggregated at the municipality level are presented below. Please note that the data from [37] refer to building units, which can either be entire buildings or parts of buildings. As the distinction between these two categories is not evidenced in [37], it is also omitted in this study, and all building units are termed 'buildings'. Table 3 shows the number of buildings in a single municipality, the municipality area, the total net floor area of all buildings in the municipality and the number of buildings per square kilometre.

Table 3. Number of buildings, total floor area, municipality area and number of buildings per square kilometre of each municipality.

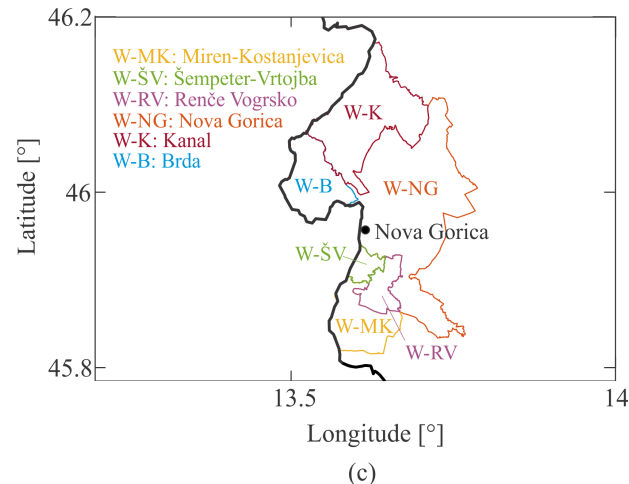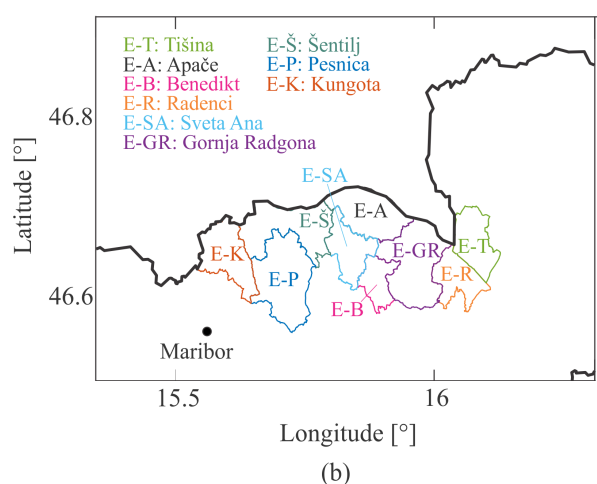|	Municipality	Municipality Label	Buildings	Net Floor Area of Buildings (m^2)	Municipality Area (km^2)	Buildings per km^2
Eastern region	Benedikt	E-B	775	174,167	24.1	32.2
	Radenci	E-R	1653	456,914	34.1	48.5
	Apače	E-A	1219	247,729	53.5	22.8
	Gornja Radgona	E-GR	2542	762,810	74.6	34.1
	Tišina	E-T	1274	254,765	38.8	32.8
	Sveta Ana	E-SA	715	157,652	37.2	19.2
	Šentilj	E-Š	2550	657,876	65.0	39.2
	Pesnica	E-P	2441	540,243	75.8	32.2
	Kungota	E-K	1669	365,958	49.0	34.1
Western region	Nova Gorica	W-NG	7451	2,903,638	279.5	26.7
	Miren–Kostanjevica	W-MK	1708	440,786	62.8	27.2
	Renče–Vogrsko	W-RV	1478	408,520	29.5	50.1
	Šempeter–Vrtojba	W-ŠV	1878	780,942	14.9	126.0
	Brda	W-B	1971	527,752	72.1	27.3
	Kanal	W-K	1638	549,967	146.5	11.2

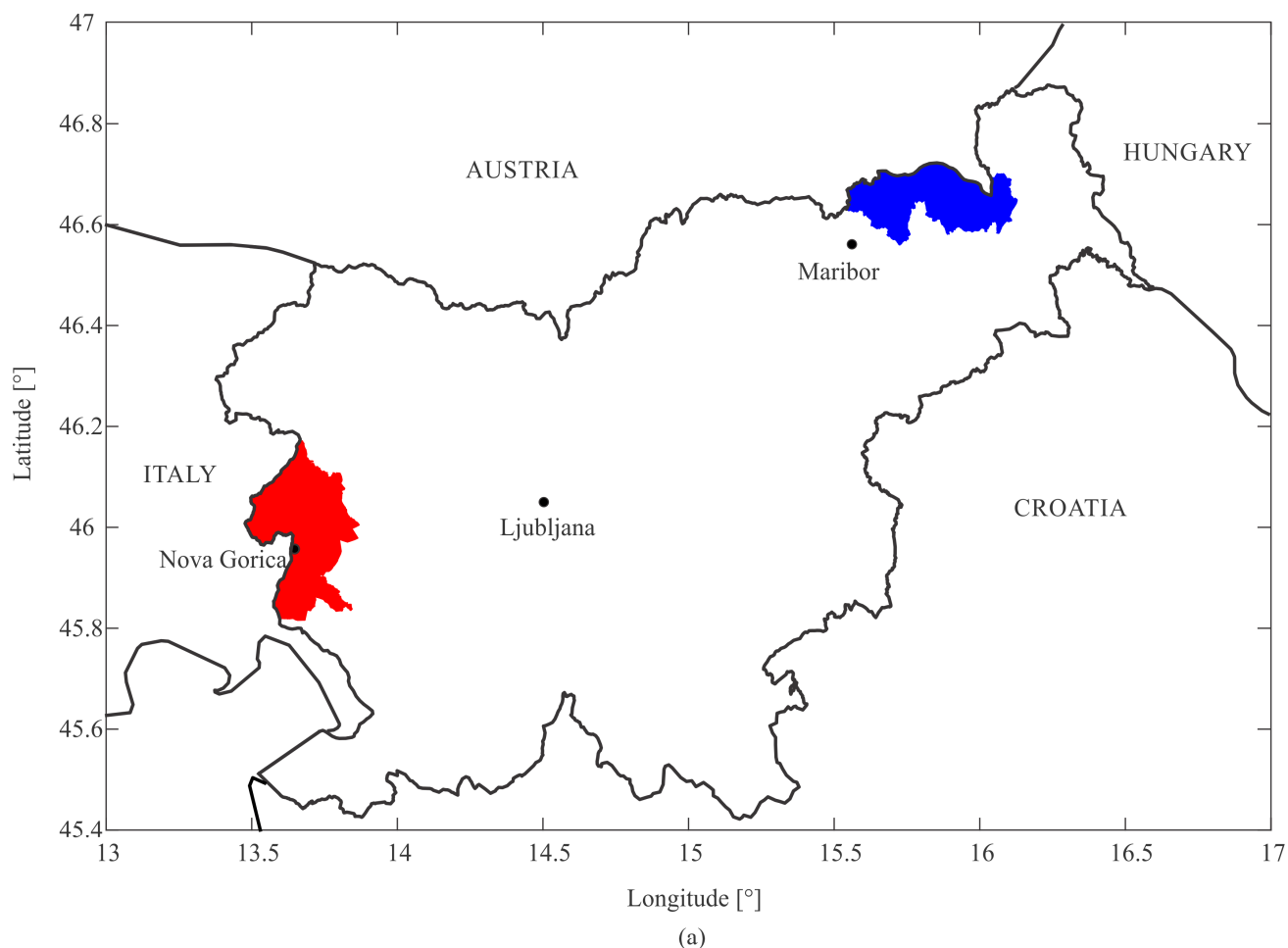**Figure 1.** (**a**) Locations of the Western (red) and Eastern (blue) regions and municipalities in the (**b**) Eastern and (**c**) Western region.

In the Western region, the largest municipality is Nova Gorica, with an area of 279.5 km^2. Nova Gorica also has the largest number of buildings (7451) and the largest total net floor area of buildings (2,903,638 m^2). However, the buildings are not evenly distributed across the municipality; most are situated in an urban area around the city of Nova Gorica, while the rest of the municipality is mostly rural. The smallest municipality in terms of area is Šempeter pri Gorici, but it has more buildings than, for example, the municipality of Kanal. Šempeter pri Gorici also has the highest number of buildings per

square kilometre (126). In terms of the total net floor area, Renče–Vogrsko is the smallest, with 400,000 m². It also has the lowest number of buildings.

In the Eastern region, the areas of the municipalities and the number of buildings per square kilometre are more uniform than in the Western group. The largest municipality is Gornja Radgona, with 76.6 km² of total area, followed by Šentilj with 65.0 km². These municipalities have almost the same number of buildings (i.e., about 2550), with Gornja Radgona having approximately 16% more total net floor area. The municipalities with the fewest buildings are Benedikt and Apače, with only 775 and 715 buildings, respectively.

Based on the Real Estate Register data, the buildings in the municipalities were further classified into the building classes defined in Section 2. Table 4 reports the percentage of buildings in each building class relative to the total number of buildings in the municipality and the percentage of the total net floor area of each building class relative to the total net floor area in the municipality. In the Western region, most buildings are in building class 1 (low-rise masonry buildings built before 1964). In contrast, in the Eastern region, most buildings are in building class 3 (low-rise masonry buildings built after 1982). Based on the net floor area percentages, the municipalities with the highest numbers of high-rise buildings are Nova Gorica, Šempeter–Vrtojba and Kanal in the Western region and Radenci, Šentilj and Gornja Radgona in the Eastern region.

Table 4. Percentage of buildings in each building class relative to the total number of buildings in the municipality and the percentage of the total net floor area of each building class relative to the total net floor area in the municipality. The green and red colours indicate higher and lower percentages, respectively.

		Building Class																			
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
E-B	No. Buildings	15	14	38	0	0	1	0	0	1	0	0	0	0	0	9	5	18	0	0	0
	Floor area	10	13	40	0	0	5	0	0	2	0	0	0	0	0	6	4	19	0	0	1
E-R	No. Buildings	20	29	28	0	0	0	1	2	2	0	1	0	0	0	3	8	6	0	0	0
	Floor area	13	21	22	1	1	1	3	10	5	1	3	3	4	0	2	5	6	0	0	0
E-A	No. Buildings	25	22	39	0	0	0	0	0	0	0	0	0	0	0	6	2	5	0	0	0
	Floor area	20	21	41	0	0	0	0	1	0	0	0	0	0	0	5	5	7	0	0	0
E-GR	No. Buildings	23	25	33	1	0	0	1	0	2	0	0	0	0	0	4	3	6	0	0	0
	Floor area	15	18	25	2	2	2	1	1	10	0	1	2	0	0	7	4	6	1	1	1
E-T	No. Buildings	28	34	28	0	0	0	0	0	2	0	0	0	0	0	2	2	4	0	0	0
	Floor area	25	35	29	0	1	0	0	0	2	0	0	0	0	0	2	1	3	0	0	0
E-SA	No. Buildings	14	15	34	0	0	0	0	0	2	0	0	0	0	0	11	5	16	0	0	0
	Floor area	11	14	39	0	0	1	0	2	4	0	0	0	0	0	7	4	17	0	0	0
E-Š	No. Buildings	13	13	29	0	0	0	2	6	7	0	0	0	0	0	9	6	13	0	0	0
	Floor area	12	11	23	1	1	1	2	10	9	0	5	1	1	0	6	5	13	0	0	0
E-P	No. Buildings	16	23	34	0	0	0	0	1	2	0	0	0	0	0	9	3	11	0	0	0
	Floor area	15	20	35	0	0	0	0	3	2	0	0	2	0	0	8	3	11	0	0	1
E-K	No. Buildings	17	17	39	0	0	0	0	1	2	0	0	0	0	0	9	3	12	0	0	0
	Floor area	15	16	36	0	0	1	0	1	4	0	0	0	0	0	8	3	15	0	0	0
W-NG	No. Buildings	32	18	22	1	0	0	1	3	4	0	1	0	0	0	8	3	6	1	0	0
	Floor area	18	11	15	2	1	1	2	5	9	1	4	5	3	4	6	2	5	2	2	1
W-MK	No. Buildings	37	11	23	0	0	0	0	1	3	0	0	0	0	0	15	2	6	0	0	0
	Floor area	30	11	25	0	0	0	0	3	5	0	0	1	0	0	12	4	8	1	0	1
W-RV	No. Buildings	26	17	25	0	0	0	1	1	2	0	0	0	0	0	19	2	7	0	0	0
	Floor area	21	16	26	1	0	1	1	5	5	0	0	0	0	0	15	2	6	0	0	1
W-ŠV	No. Buildings	26	20	21	0	0	0	1	4	5	0	0	1	0	0	13	2	6	0	0	0
	Floor area	19	13	14	2	1	0	3	9	10	0	1	5	0	3	9	2	8	0	0	1
W-B	No. Buildings	32	19	24	0	0	0	0	1	3	0	0	0	0	0	10	5	6	0	0	0
	Floor area	26	19	25	1	0	0	2	1	5	1	0	1	0	0	9	4	5	1	0	0
W-K	No. Buildings	48	10	11	1	0	0	3	6	5	0	0	0	0	0	7	4	4	1	0	0
	Floor area	31	7	8	2	2	0	4	10	6	5	1	2	1	0	8	5	4	2	0	0

To demonstrate the difference in the seismic hazard between the two regions, the PGA values for the centres of municipalities, soil type A and return periods of 475 and 2475 years are presented in Table 5. The PGA values were obtained from the hazard curves derived using the OpenQuake engine [28,29] and ESHM2020 model [30] (see Section 2). Based on Table 5, it is evident that the seismic hazard in the Western region is higher than in the Eastern region. The PGAs for the 475-year return period do not vary significantly within a

given region. However, a slightly larger variation can be observed for the return period of 2475 years. For the Eastern region, the PGA for the return period of 475 years ranges between 0.07 g and 0.08 g, while that for the return period of 2475 years is between 0.16 g and 0.18 g. The range of PGAs in the Western region is between 0.14 g and 0.17 g for the return period of 475 years and between 0.32 g and 0.43 g for the return period of 2475 years.

Table 5. PGA on soil type A according to Eurocode 8 for return periods of 475 and 2475 years.

	Municipality	PGA for 475-Year Return Period (g)	PGA for 2475-Year Return Period (g)
Eastern region	Benedikt	0.08	0.18
	Radenci	0.08	0.18
	Apače	0.07	0.16
	Gornja Radgona	0.08	0.18
	Tišina	0.08	0.17
	Sveta Ana	0.08	0.17
	Šentilj	0.08	0.16
	Pesnica	0.08	0.18
	Kungota	0.08	0.17
Western region	Nova Gorica	0.16	0.43
	Miren–Kostanjevica	0.14	0.32
	Renče–Vogrsko	0.15	0.37
	Šempeter–Vrtojba	0.15	0.34
	Brda	0.16	0.35
	Kanal	0.17	0.42

4. Methodology for the Estimation of Bias in Loss Estimation for Different Knowledge Levels of Building Data

The loss estimation bias for a municipal building stock was calculated in terms of log residuals $\delta_{EAL,KL}$:

$$\delta_{EAL,KL} = \ln\left(\frac{EAL_{KL}}{EAL_{base}}\right), \qquad (6)$$

where EAL_{KL} is the municipality's expected annual loss for the building data knowledge level (KL), and EAL_{base} is the municipality's expected annual loss for the base-case KL. A positive and a negative $\delta_{EAL,KL}$ indicate overestimation and underestimation of the EAL, respectively. Such a measure of bias was selected because it allows a more straightforward comparison of negative and positive errors than the relative difference $(EAL_{KL} - EAL_{base})/EAL_{base}$, which is bounded by -100% on the negative side and unbounded on the positive side. However, for low values of $\delta_{EAL,KL}$ (up to 0.1), the following applies:

$$\delta_{EAL,KL} \approx \frac{EAL_{KL} - EAL_{base}}{EAL_{base}}. \qquad (7)$$

Both EAL_{KL} and EAL_{base} are calculated as described in Section 2 (Equation (5)). The base-case KL corresponds to the complete data, as described in Section 3. However, other KLs represent a variation in the base-case KL defined to be consistent with the assumptions on the data limitations at those levels. Twelve KLs were defined, including the base-case KL. KLs were defined as a combination of (1) the location knowledge level (KL_L) and (2) the floor-area knowledge level (KL_{FA}) (Table 6). KL_L indicates the accuracy of the definition of buildings' locations within a municipality. By lowering KL_L, the seismic hazard assigned to the municipality's building stock is less accurate, which introduces bias in the estimation of the EAL. However, because building-specific characteristics that affect the building's fragility are location-dependent, KL_L also affects the EAL through the damage assessment. In contrast, KL_{FA} indicates the precision of defining the net floor areas of buildings within a given building class in a municipality. The term 'net' is omitted hereinafter for brevity. The floor area of a building is the building's main characteristic affecting the outcome of the

consequence assessment. Therefore, reducing KL_{FA} also introduces bias in the estimation of *EAL* at the building level and, consequently, at the municipality level.

Table 6. Definitions of building data knowledge levels considered.

Floor-Area Knowledge Level	Location Knowledge Level	KL_{FA} = 'a' Floor Area of Each Building Equal to the Average Floor Area in the Municipality	KL_{FA} = 'b' Floor Area of Each Building Equal to the Average Floor Area within the Building Class in the Municipality	KL_{FA} = 'c' Actual Floor Area of Each Building Considered
KL_L = '1'	All buildings lumped in one point; hazard equal to that in the municipality's centre	KL = '1a'	KL = '1b'	KL = '1c'
KL_L = '2'	All buildings lumped in one point; hazard equal to the municipality's average	KL = '2a'	KL = '2b'	KL = '2c'
KL_L = '3'	Buildings' locations are randomly selected consistent with the building density distribution across the municipality; hazard determined for defined locations	KL = '3a'	KL = '3b'	KL = '3c'
KL_L = '4'	Actual location of each building considered; hazard determined for defined locations	KL = '4a'	KL = '4b'	KL = '4c'

Four location knowledge levels (KL_L = '1', '2', '3' or '4') and three floor-area knowledge levels (KL_{FA} = 'a', 'b' or 'c') were considered. The highest (base-case) KL combines KL_L = '4' and KL_{FA} = 'c' and is therefore denoted as '4c'. This KL indicates building-specific data, both in terms of the buildings' locations and floor areas. Such a KL is available in Slovenia, as explained in Section 3.

At the lowest KL ('1a'), no distinction was made between the locations or floor areas of different buildings in a municipality. Instead, the seismic hazard of all buildings in a municipality was considered equal to that in the centre of the municipality, taking into account the municipality's soil conditions (KL_L = '1'). Therefore, the hazard curve applied to each building in a municipality was changed to the hazard curve in the grid point closest to the municipality's centre and adjusted for the average soil conditions. Moreover, as no distinction was considered between the floor areas of different buildings, the floor areas of all buildings in a municipality were changed to the average floor area in the municipality (KL_{FA} = 'a'). Such a KL was selected as the lowest one because it is based on very limited data while still representing a realistic scenario. It combines the treatment of the seismic hazard used in a previous study of seismic risk [5] and the availability of floor-area data observed for Austria [44].

The first intermediate location knowledge level (KL_L = '2') was defined similarly as KL_L = '1' in that no distinction was made between the locations of buildings in a municipality. However, in contrast to KL_L = '1', the hazard of buildings at KL_L = '2' was considered equal to the municipality's average hazard, assuming uniform building density distribution across the municipality. Therefore, the hazard curve applied to each building

in a municipality was changed to the average of the hazard curves from all grid points in the municipality and adjusted for the average soil conditions.

At the second intermediate location knowledge level (KL_L = '3'), the variation in buildings' locations in a municipality was considered. The locations were selected consistently with the building density distribution across the municipality. Thus, they were the same as in the case of KL_L = '4'. However, level KL_L = '3' differed from KL_L = '4' in that it did not allow consideration of the actual building characteristics at a given location. Therefore, the locations of buildings in the municipality were randomly permuted in a set of simulations, and the average *EAL* over all simulations was considered as EAL_{KL}.

Moreover, the intermediate floor-area knowledge level (KL_{FA} = 'b') indicates that the distinction is made between the building classes but not between the buildings within a given building class. Therefore, the floor area of each building was changed to the average floor area in its building class.

The availability of data can influence the selection of *KL*. For example, suppose the coordinates of the municipality's borders are not fully known, but those of the municipality's centre are known. In that case, KL_L = '1' can be selected because it does not require calculation of the hazard over the entire municipality's area. If the municipality's borders are also known, the average hazard over the municipality area can be evaluated, and KL_L = '2' can be applied. If the building density distribution is known, from remote sensing imagery, for example, KL_L = '3' can be used. Finally, if data on the characteristics of buildings at each specific location are available, the highest location knowledge level considered in this study, KL_L = '4', can be selected. Similarly, if there are no data on the floor areas typical for different building types, the average floor area in the municipality may be assigned to each building, which corresponds to KL_{FA} = 'a'. If the floor areas typical of the building types considered can be estimated, KL_{FA} can be increased to 'b'. Finally, if the variation in floor areas within the building classes is known, KL_{FA} = 'c' can be used. Increasing *KL* can be quite simple in some cases (e.g., identifying the municipality's borders) but more demanding in others (e.g., determining the floor-area variation in the municipality). However, a *KL* increase typically requires a certain amount of effort, which is not necessarily justified given the resources available for the analysis.

It should be noted that the selection of *KL* can also represent the analyst's decision to simplify the analysis despite the available data. For example, the selection of an appropriate exposure resolution to obtain low-bias results, given that data are available, was explored in [7]. It should also be noted that the list of *KL*s considered herein is not exclusive. Additional *KL*s could be defined between the lowest and highest *KL*. For example, the location and floor-area data could be defined at the census tract level, which is an intermediate level between the building-specific and municipality levels. In addition, other *KL*s could be defined using less accurate data than those representative of *KL* = '1a' or more accurate data than those representative of *KL* = '4c'. For example, the number of buildings in different building classes could be considered unavailable, which would introduce additional uncertainty in the application of the fragility model. Consideration of such options is beyond the scope of this study.

5. Bias in Seismic Risk for the Selected Building Data Knowledge Levels

The log residuals $\delta_{EAL,KL}$ (Equation (6)) selected as the measure of bias were estimated for each *KL*. At the lowest *KL* ('1a'; Figure 2), the highest $\delta_{EAL,KL}$ was calculated for the municipality of Benedikt (+0.17) in the Eastern region, while the lowest $\delta_{EAL,KL}$ corresponded to the municipality of Miren–Kostanjevica in the Western region (−0.17). Interestingly, most municipalities in the Western region had a negative $\delta_{EAL,KL}$, while a majority of the municipalities in the Eastern region had a positive $\delta_{EAL,KL}$. Accordingly, the mean $\delta_{EAL,KL}$ calculated at the regional scale (rightmost dot in Figure 2) was positive for the Eastern region and negative for the Western region. The bias was relatively small in both cases, as the absolute values of $\delta_{EAL,KL}$ were between 0.05 and 0.06.
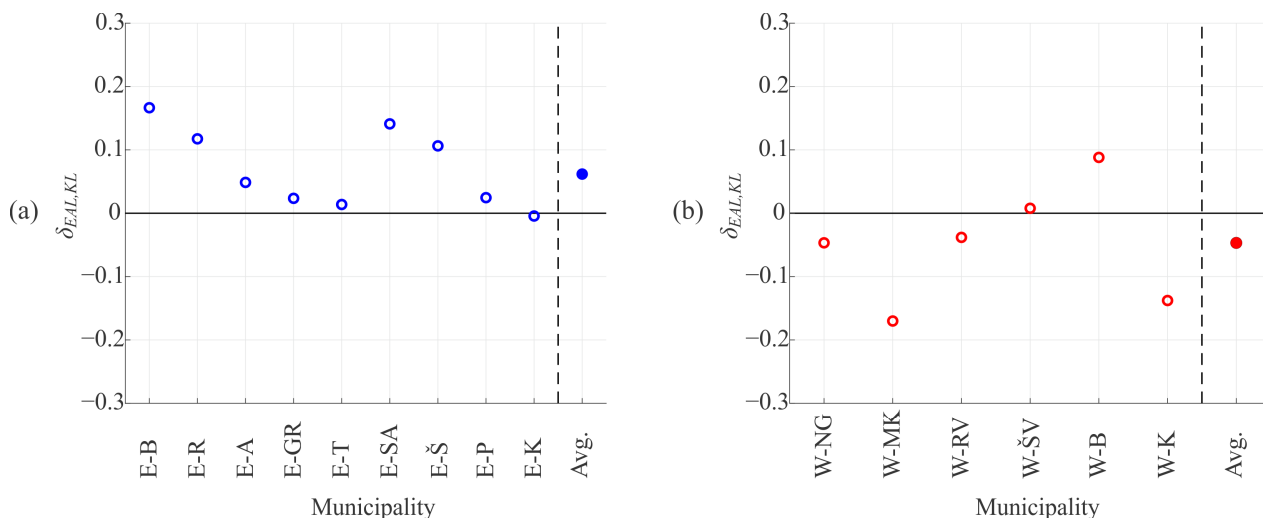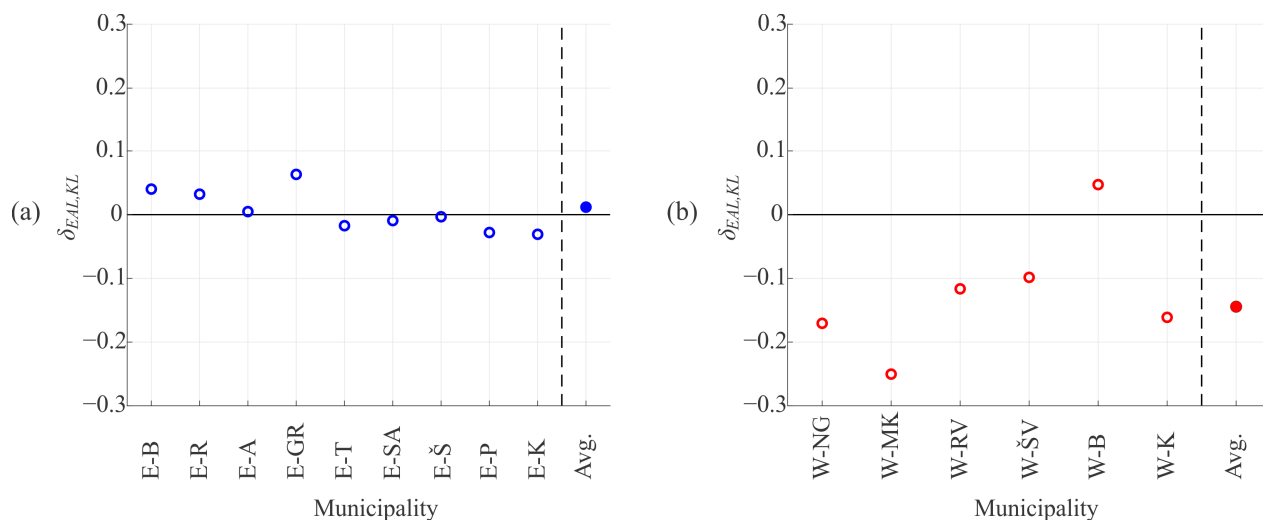
Figure 2. Log residuals $\delta_{EAL,KL}$ for KL = '1a' for municipalities in the (**a**) Eastern region and (**b**) Western region.

By increasing the building data knowledge level to KL = '1b', the $\delta_{EAL,KL}$ values of all municipalities in the Western region shifted by a similar amount (Figure 3b). This shift was negative (-0.09 on average), thus intensifying the error compared to KL = '1a'. The largest bias was again observed for the municipality of Miren–Kostanjevica (-0.25). Such a result may be surprising because it implies that by improving the data, the bias increases. However, it can be explained by the fact that, for this particular region, the bias due to low KL_L was negative, while the bias due to low KL_{FA} was positive. By increasing KL_{FA} from 'a' to 'b', the positive part of the bias decreased, leading to a lower reduction in the negative bias and, consequently, to an increase in the absolute values of $\delta_{EAL,KL}$. Conversely, in the case of the Eastern region (Figure 3a), an increase in the building data knowledge level to KL = '1b' diminished the overall bias so that the maximum municipality $\delta_{EAL,KL}$ was only $+0.06$, and the average $\delta_{EAL,KL}$ in the region was practically negligible.
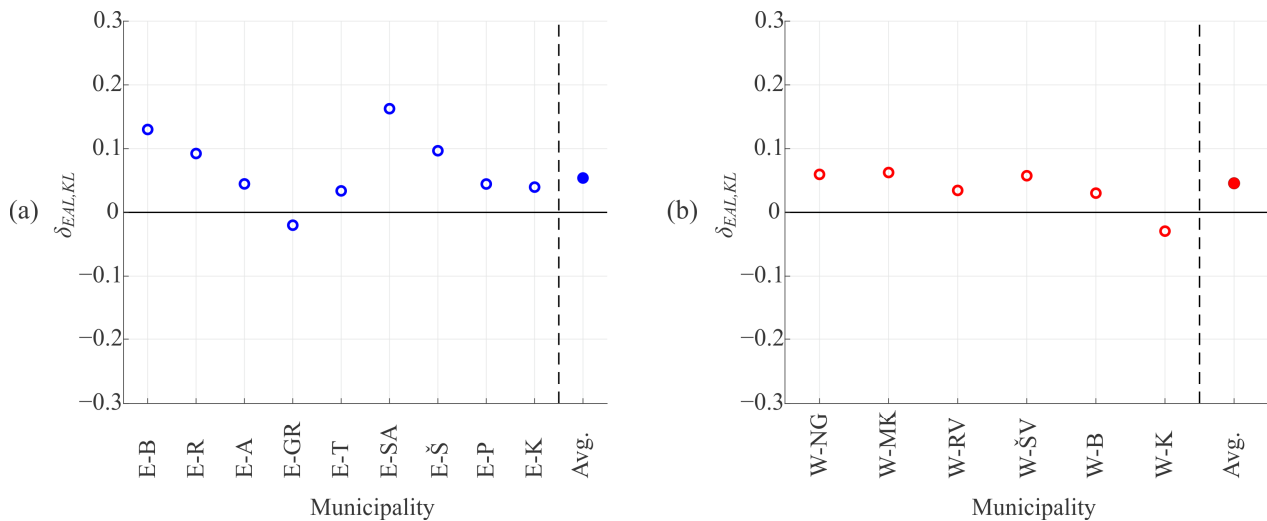

Figure 3. Log residuals $\delta_{EAL,KL}$ for KL = '1b' for municipalities in the (**a**) Eastern region and (**b**) Western region.

An opposite outcome was observed when the floor-area knowledge level was maintained at KL_{FA} = 'a' and the location knowledge level was increased to KL_L = '3' (Figure 4). In this case, the $\delta_{EAL,KL}$ values for the municipalities in the Western region generally in-

creased compared to the values obtained at KL = '1a'. However, their absolute values decreased and were all lower than 0.07, which means that improving KL_L had a significantly better effect on bias reduction than improving KL_{FA}. On the other hand, $\delta_{EAL,KL}$ for the municipalities in the Eastern region did not drastically change when KL increased from '1a' to '3a' (Figure 4a), and the bias reduction was not as pronounced as when KL increased to '1b'. This implies that KL_{FA} was more crucial for this particular region than KL_L.

Figure 4. Log residuals $\delta_{EAL,KL}$ for KL = '3a' for municipalities in the (**a**) Eastern region and (**b**) Western region.

For other KLs, only the minimum, maximum and average $\delta_{EAL,KL}$ values for a region are presented for brevity (Figure 5). Some of the observations mentioned above related to KL = '1a', '1b' and '3a' can be further generalised. For the Western region (Figure 5b), an improvement in KL_{FA} from 'a' to 'b' generally decreased $\delta_{EAL,KL}$. However, decreasing $\delta_{EAL,KL}$ did not clearly reduce the bias (the absolute value of $\delta_{EAL,KL}$ did not change notably or was even further increased), except for KL_L = '4'. When KL_{FA} improved further to 'c', a small reduction in bias was again observed for KL_L = '4', while this reduction was negligible for other location knowledge levels. In fact, $\delta_{EAL,KL}$ was exactly the same for KL = '1b' and '1c' and KL = '2b' and '2c'. This is to be expected because all buildings were assigned the same hazard curve at location knowledge levels KL_L = '1' and '2'. Therefore, dividing the floor area within a building class evenly (KL_{FA} = 'b') and based on the actual data (KL_{FA} = c) resulted in the same building class's EAL. In contrast, for the Eastern region (Figure 5a), increasing KL_{FA} from 'a' to 'b' clearly improved the accuracy of the risk estimation, as it resulted in a smaller range of $\delta_{EAL,KL}$, centred closer to 0. However, when the level of this type of knowledge improved further to level 'c', the beneficial effect was again less significant, for the same reasons as in the case of the Western region.

With regard to the location knowledge, the most significant bias reduction was achieved by increasing KL_L from '2' to '3' (Figure 5), i.e., by including data on the building density distribution across the municipality. However, the effect of these data was much more pronounced for the municipalities in the Western region than for those in the Eastern region. Further improving KL_L to '4' had a small bias reduction effect, especially for the Western region, while improving this type of knowledge level from '1' to '2' had a negligible effect for both regions.

The results indicate that the effect of including different data types on the bias in seismic risk is municipality-dependent. To improve the understanding of this dependency, the municipal characteristics that affect the bias in EAL were identified, as presented in the following section.
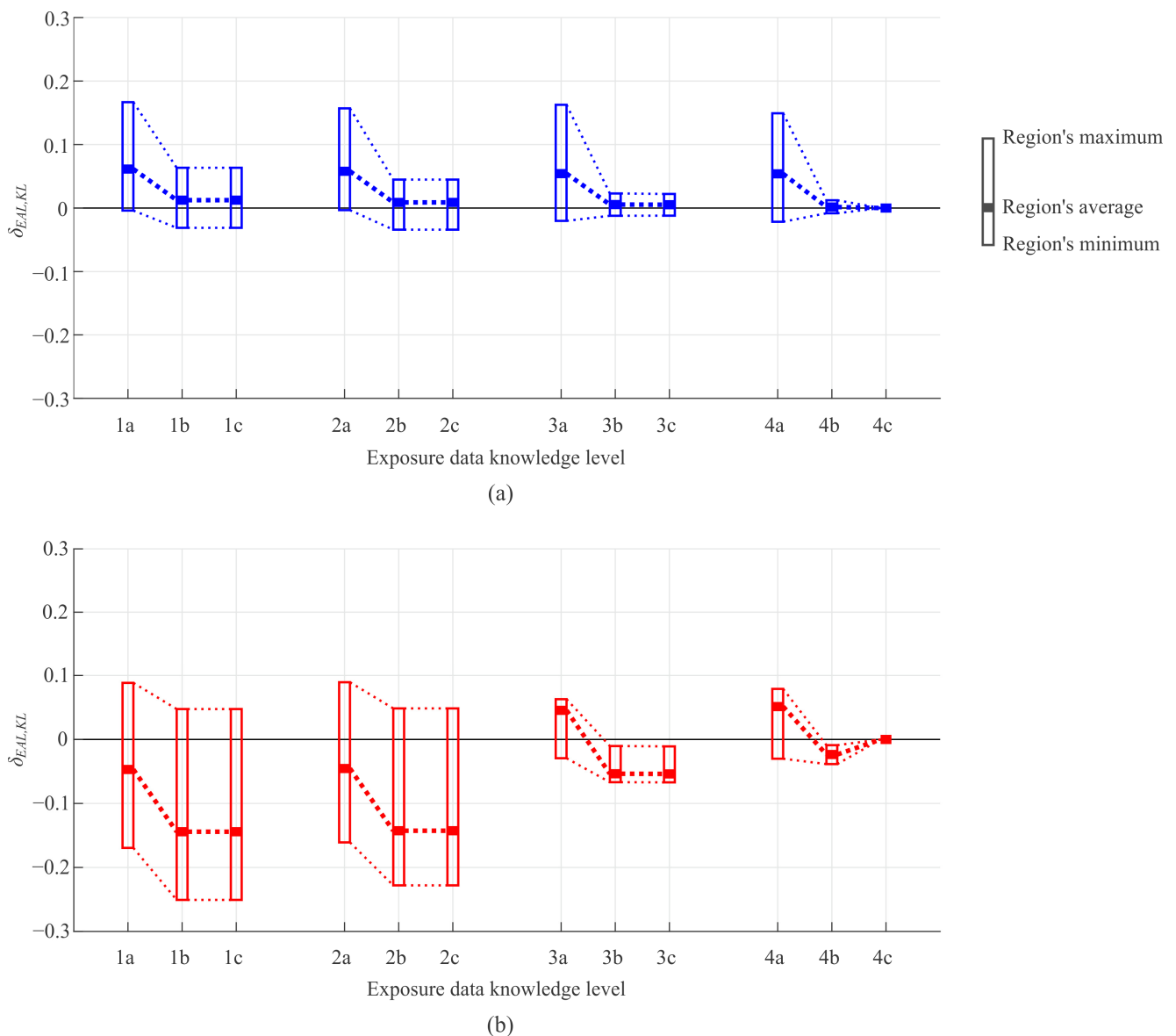
Figure 5. Regional minimum, maximum and average $\delta_{EAL,KL}$ for all *KL*s considered for the (**a**) Eastern region and (**b**) Western region.

6. Identification of Municipal Characteristics Affecting the Bias in Seismic Risk

Identifying the characteristics of the municipalities affecting the bias reduction can improve the understanding of why some municipalities are more prone to loss estimation bias in the case of incomplete data. It can also allow the analyst to predict whether adding data of a given type can reduce the bias if these characteristics are already available at the given building data knowledge level.

The municipal characteristics affecting the bias at the selected building data knowledge levels were identified in two steps. In the first step, the data types that contribute most to the bias reduction were determined. The results presented in the previous section indicate that the bias at the regional scale can be most reduced by including the data on building density distribution across the municipality (transition from KL_L = '2' to KL_L = '3') and building-class-specific floor areas (transition from KL_{FA} = 'a' to KL_{FA} = 'b'). However, to obtain a clearer indication of the effect of different data types, the absolute differences in $\delta_{EAL,KL}$ caused by a gradual increase from knowledge level *KL* to knowledge level *KL'*, i.e., $|\delta_{EAL,KL'} - \delta_{EAL,KL}|$, were also calculated at the municipality level (Figure 6).
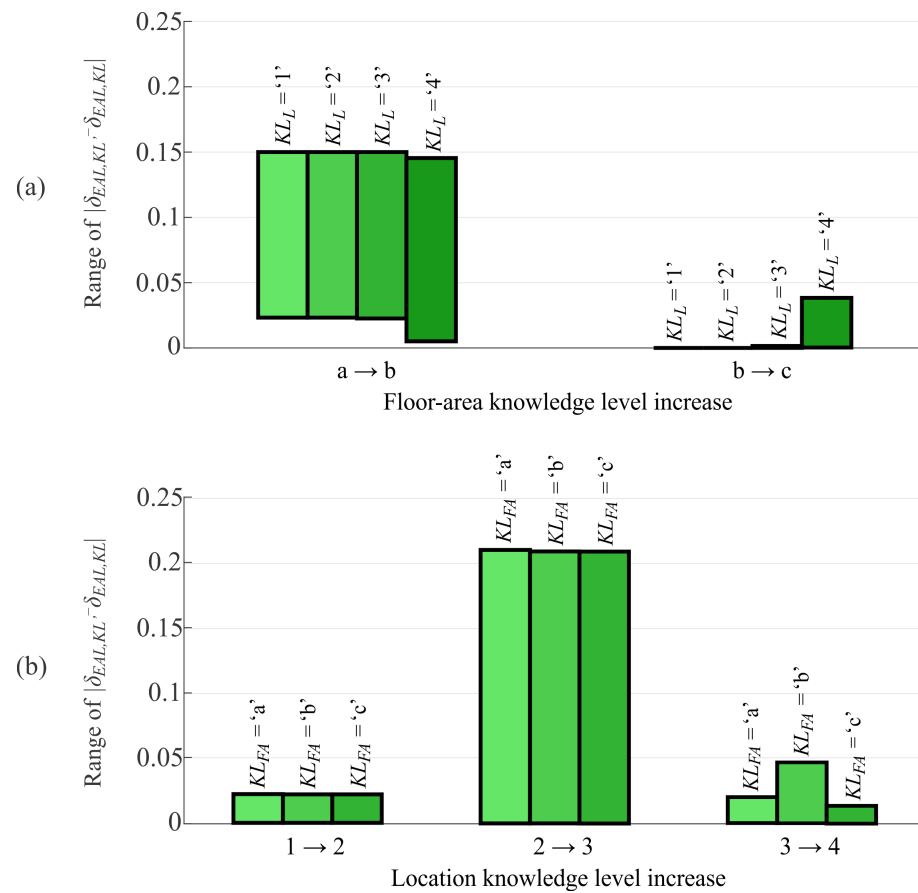
Figure 6. Ranges of $|\delta_{EAL,KL'} - \delta_{EAL,KL}|$ (absolute differences in $\delta_{EAL,KL}$) at the municipality level caused by gradual increases in (**a**) KL_{FA} and (**b**) KL_L.

Consistent with the observations at the regional scale, it was found that the absolute difference in $\delta_{EAL,KL}$ at the municipality level was the largest when increasing KL_L from '2' to '3' and KL_{FA} from 'a' to 'b' (Figure 6). All other gradual increases in KL had a much lower effect on the bias. Including the data on building density distribution across the municipality in the analysis (KL_L from '2' to '3') changed $\delta_{EAL,KL}$ by as much as 0.21, regardless of the floor-area knowledge level, while adding the data on building-class-specific floor areas (KL_{FA} from 'a' to 'b') resulted in the absolute difference of $\delta_{EAL,KL}$ being as high as 0.15, with a small variation across the location knowledge levels. For both knowledge-level increases, the lowest absolute difference of $\delta_{EAL,KL}$ was close to 0, which implies that including new data affects the bias reduction differently for different municipalities.

In the second step of the identification of the essential municipal characteristics related to the bias in seismic risk, the variations in the absolute difference of $\delta_{EAL,KL}$ for the two most significant increases in KL were analysed further to identify the indicators that correlate well with the bias reduction. Several simple indicators (e.g., municipality's area, number of buildings and average building density) and more advanced indicators were tested. The indicators that were found to best explain the bias reduction for the two most significant increases in KL are presented in the following.

The effect of increasing KL_{FA} from 'a' to 'b' on the bias can be explained by the municipality's building stock homogeneity index I_{BSH}, calculated as follows:

$$I_{BSH} = \sum_{i=1}^{n_c} \left(\frac{N_{b,i}}{N_b}\right)^2, \qquad (8)$$

where $N_{b,i}$ is the number of buildings in the *i*th building class in the municipality, N_b is the total number of buildings in the municipality and n_c is the number of building classes in the municipality. The value of I_{BSH} is between zero and one. It equals one if the municipality has only one building class and approaches zero if the buildings are distributed into more building classes. Moreover, for a given n_c, I_{BSH} is equal to $1/n_c$ if the building class distribution is uniform, while it takes a larger value if most buildings are in only a few building classes.

I_{BSH} correlated reasonably well with the absolute difference in $\delta_{EAL,KL}$ caused by the increase from KL_{FA} = 'a' to KL_{FA} = 'b', i.e., $|\delta_{EAL,KL'} - \delta_{EAL,KL}|_{a\rightarrow b}$ (Figure 7). Regardless of the location knowledge level, the correlation coefficient ρ was about -0.65, and the coefficient of determination R^2 was approximately 0.43. The higher the I_{BSH} is, the lower the benefit is of adding the data on building-class-specific floor areas in the analysis. In other words, the higher the I_{BSH} is, the more likely it is that KL_{FA} = 'a' results in a low-bias estimate of *EAL*. The reason for this is that higher values of I_{BSH} indicate that the municipality's building stock is less fragmented and that, consequently, the average floor area over all buildings in the municipality can better describe the average floor area in the building class with the predominant contribution to the *EAL*. For example, almost all buildings in the municipality of Tišina (E-T) are masonry with a maximum of three storeys (Table 4). Therefore, the building stock is relatively homogenous, which is reflected by I_{BSH} = 0.275 (Figure 7), and the average floor area over all buildings in the municipality (200 m²) is close to the average floor areas of the three building classes that contribute most to the *EAL* (182, 202 and 211 m²). Consequently, providing additional data on building-class-specific floor areas did not reduce the bias significantly (by approximately 3% for all location knowledge levels). For other municipalities, I_{BSH} ranged from 0.16 to 0.27. In the case of I_{BSH} greater than 0.25, using floor-area knowledge level KL_{FA} = 'a' could produce relatively accurate results (Figure 7). It should be noted that the absolute values of I_{BSH} depend on the building classification. For a different number of classes, different I_{BSH} values would indicate the sufficiency of KL_{FA} = 'a'.

However, the effect of increasing KL_L from '2' to '3' on the loss estimation bias was found to be explainable by the hazard variation index I_{HV}, calculated as follows:

$$I_{HV} = \frac{\sigma_{EAL_i,2a,HV}}{\overline{EAL_{i,2a}}}, \tag{9}$$

where $\overline{EAL_{i,2a}}$ is the average building-specific *EAL* in the municipality calculated for KL = '2a', and $\sigma_{EAL_i,2a,HV}$ is defined as follows:

$$\sigma_{EAL_i,\,2a,HV} = \sqrt{\overline{\sigma_{EAL,\,2a,rand}} - \sigma_{EAL,\,2a}}. \tag{10}$$

In Equation (10), $\sigma_{EAL,\,2a}$ is the standard deviation of building-specific *EAL*s in the municipality calculated for KL = '2a'. Because buildings are assigned the same hazard curve and the same floor area at this KL, $\sigma_{EAL,2a}$ reflects only the variability in the buildings' fragility. However, $\overline{\sigma_{EAL,\,2a,rand}}$ is the average standard deviation of building-specific *EAL*s in the municipality with randomly distributed building stock. It is calculated as the average of the standard deviations $\sigma_{EAL,2a,rand}$ obtained in multiple simulations. In each simulation, each building in the municipality is assigned a random location within the municipality (disregarding the building density distribution across the municipality), while the fragility functions and floor area remain the same as in the calculation of $\sigma_{EAL,2a}$. Building-specific *EAL*s are then determined, and $\sigma_{EAL,2a,rand}$ is calculated as their standard deviation. In this study, 50 such simulations were performed. The calculated $\overline{\sigma_{EAL,2a,rand}}$ reflects the variability in the buildings' fragility and the variability in the hazard across the municipality. Assuming that these two variabilities are uncorrelated and that building-specific *EAL*s in the municipality are normally distributed, $\sigma_{EAL_i,\,2a,HV}$ in Equation (10) represents the variation in building-specific *EAL*s resulting only from the municipality's hazard variability.

The index I_{HV} can thus be understood as a coefficient of variation of building-specific *EAL*s resulting only from the hazard variability.
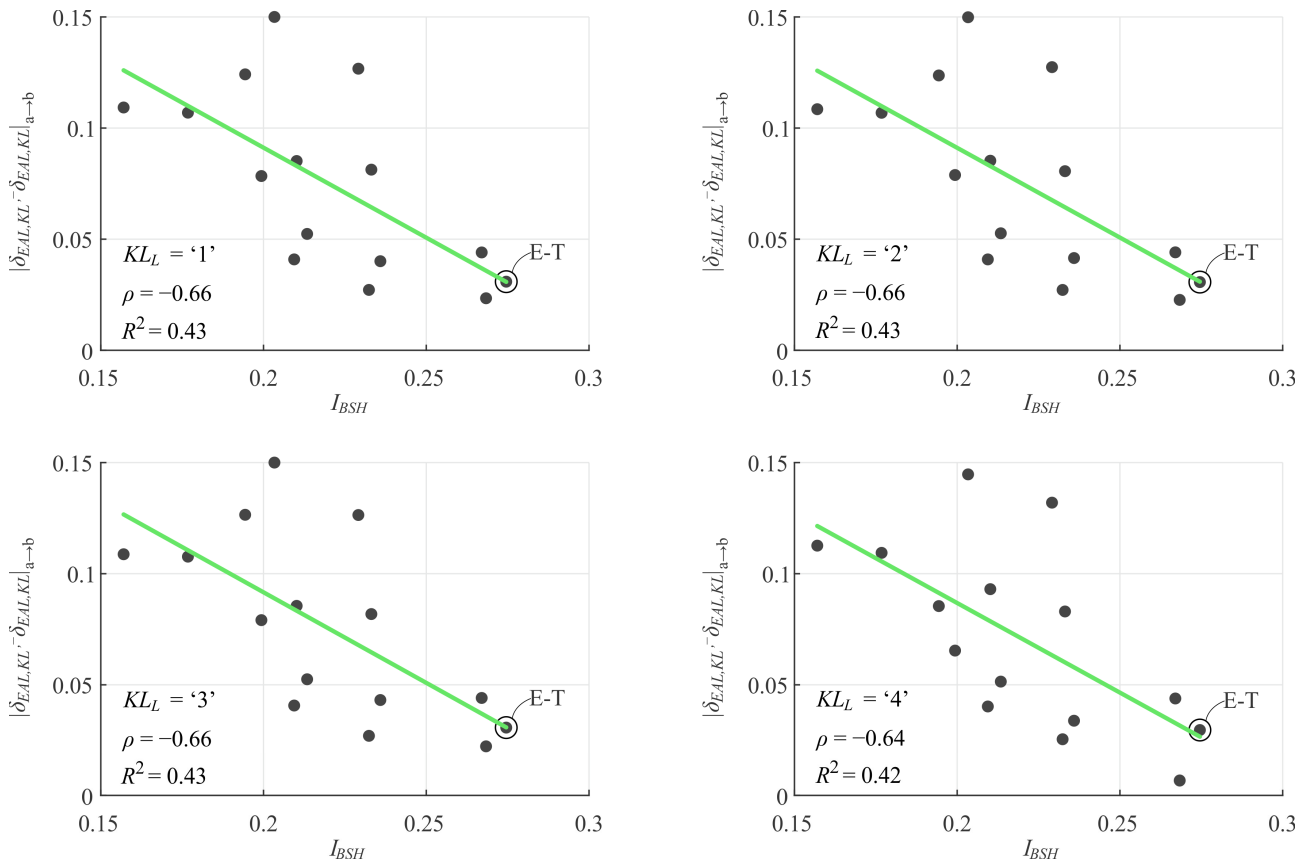

Figure 7. Correlation between I_{BSH} (building stock homogeneity index) and $|\delta_{EAL,KL'} - \delta_{EAL,KL}|_{a \to b}$ (the municipality's absolute difference in $\delta_{EAL,KL}$ caused by the increase from KL_{FA} = 'a' to KL_{FA} = 'b') for all location knowledge levels considered.

I_{HV} is positively correlated with the absolute difference in $\delta_{EAL,KL}$ caused by a knowledge level increase from KL_L = '2' to KL_L = '3', i.e., $|\delta_{EAL,KL'} - \delta_{EAL,KL}|_{2 \to 3}$ (Figure 8). Regardless of the floor-area knowledge level, ρ was about +0.69, and R^2 was approximately 0.47. This positive correlation implies that an increase in I_{HV} intensifies the potential for the hazard variation to alter the municipality's *EAL* when data on building density distribution across the municipality are included. For example, for the municipality of Miren–Kostanjevica (W-MK), the average and standard deviation of building-specific *EAL*s calculated for *KL* = '2a' ($\overline{EAL_{i,2a}}$ and $\sigma_{EAL,2a}$, respectively) were EUR 137 and EUR 69, respectively. When the buildings were moved from the municipality's centre to random locations in the municipality, the standard deviation, on average, increased notably ($\overline{\sigma_{EAL,2a,rand}}$ = EUR 85), leading to $\sigma_{EAL_i, 2a,HV}$ = EUR 49 (Equation (10)) and I_{HV} = 0.36 (Equation (9)). Such a high I_{HV} indicates that the municipality's *EAL* could change significantly if the buildings were moved to their actual location. This was also indicated by $|\delta_{EAL,KL'} - \delta_{EAL,KL}|_{2 \to 3}$ being equal to 0.21 (Figure 8). For other municipalities, I_{HV} was between 0.06 and 0.42. The results suggest that for I_{HV} less than 0.1, the hazard variation in the municipality does not have significant potential to cause bias in the estimation of seismic risk.
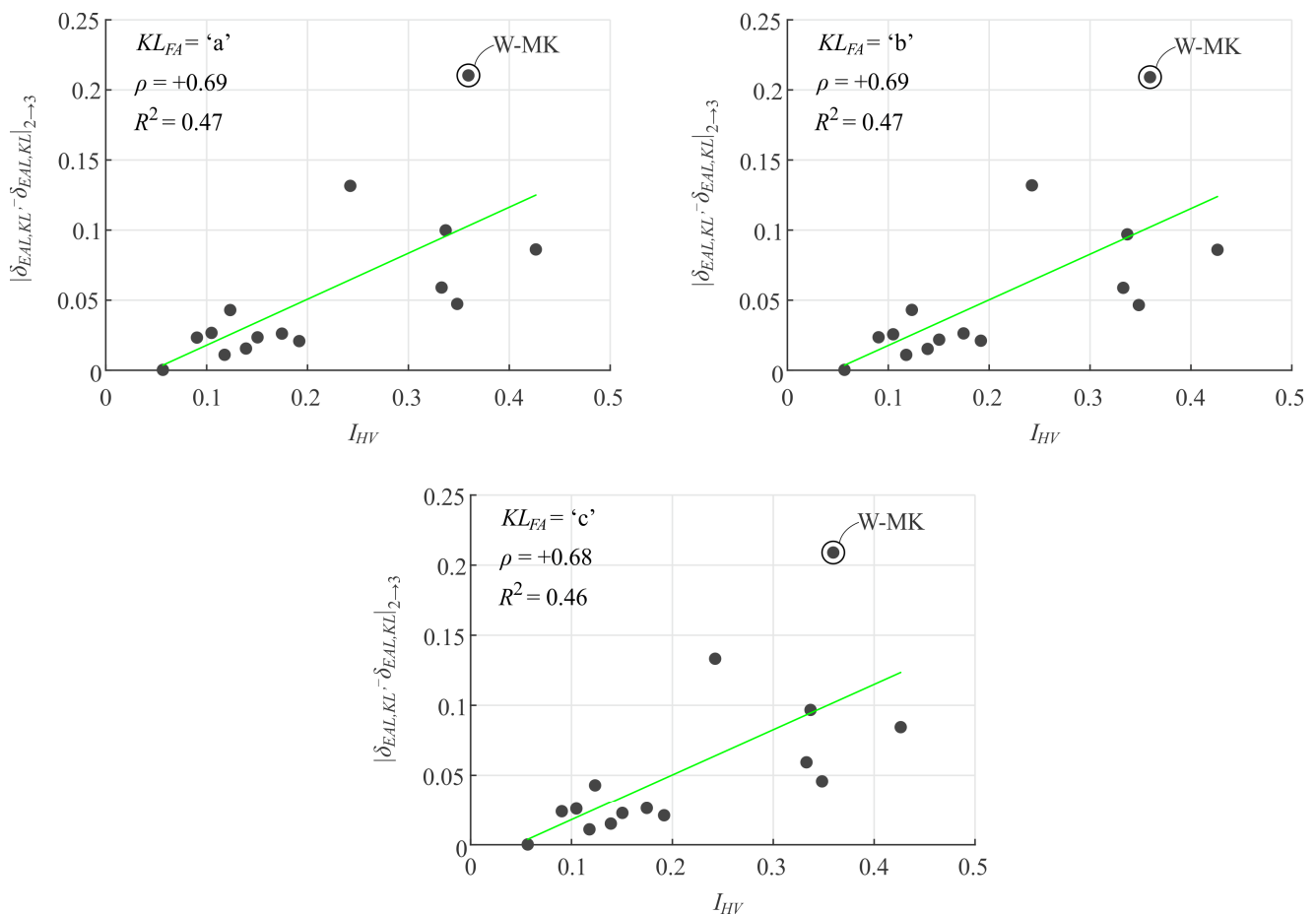
Figure 8. Correlation between I_{HV} (the municipality's hazard variation index) and $|\delta_{EAL,KL'} - \delta_{EAL,KL}|_{2\to3}$ (the municipality's absolute difference in $\delta_{EAL,KL}$ caused by the increase from $KL_L =$ '2' to $KL_L =$ '3') for all floor-area knowledge levels considered.

The indexes I_{BSH} and I_{HV} can therefore be used to estimate whether additional data would significantly reduce the bias in seismic risk estimation. It is worth emphasising that I_{BSH} can be calculated at all KLs considered in this study, while I_{HV} can be obtained at $KL =$ '2a' or higher. Therefore, its application requires at least data on the municipality's borders and average floor area over all buildings in the municipalities.

7. Conclusions

The effect of the availability of building data on loss estimation was analysed for fifteen municipalities in two Slovenian regions. Twelve knowledge levels of building data were introduced as a combination of the location knowledge level and floor-area knowledge level. The bias for a given building data knowledge level was defined as the log residual between the municipality's expected annual loss (*EAL*) estimated for that knowledge level and the municipality's *EAL* estimated for the base-case knowledge level.

The log residuals of *EAL* ranged from -0.25 to $+0.16$ at the lowest knowledge levels. The bias was affected by both the location and floor-area knowledge levels. It was found that introducing more data can reduce the bias but not in general. Among the gradual knowledge level improvements considered in this study, those corresponding to the inclusion of data on the building density distribution across the municipality and data on the building-class-specific floor areas proved to be the most beneficial. Additional data on building-specific locations and floor areas reduced the bias only slightly.

The importance of a given type of data was found to be dependent on the municipality. However, the results suggest that for a given municipality, the reduction in bias caused

by introducing new data can be estimated in advance based on readily available data. Introducing building-class-specific floor area data is anticipated to mitigate bias more effectively in municipalities with a lower building stock homogeneity index (I_{BSH}), which quantifies the fragmentation of buildings into building classes. For I_{BSH} values greater than 0.25, using the average floor area for all buildings in the municipality was sufficient to obtain a relatively accurate estimate of *EAL*. However, including data on the building density distribution across the municipality is expected to yield greater advantages for municipalities with a higher hazard variation index (I_{HV}), which measures the potential for the hazard variation in the municipality to intensify bias in the estimation of seismic risk. The findings imply that for I_{HV} values less than 0.1, the bias remains low even if assigning the average hazard in the municipality to each building. Both indexes were introduced in this study and require a low knowledge level of readily available data.

The presented findings can be beneficial in making decisions on collecting additional data when the current data knowledge level is low, especially in the case of limited resources. In such situations, it is suggested to focus primarily on obtaining data types that proved to have a higher impact on bias reduction. Moreover, it is suggested that the prioritisation of municipalities for enhancing their data knowledge levels be guided by indexes such as I_{BSH} and I_{HV}.

The findings refer to the selected building data knowledge levels and building classification. Additional research is needed to identify the sources of loss estimation bias for other knowledge levels, which may be below, above or between the knowledge levels considered in this study. Further research is also needed to generalise the concepts introduced in this study to other regions and building classifications.

References

1. ICPD—Presidency of the Council of Ministers Italian Civil Protection Department. *National Risk Assessment. Overview of the Potential Major Disasters in Italy: Seismic, Volcanic, Tsunami, Hydro-Geological/Hydraulic and Extreme Weather, Droughts and Forest Fire Risks*; ICPD: Rome, Italy, 2018.
2. Crowley, H.; Rodrigues, D.; Silva, V.; Despotaki, V.; Martins, L.; Romão, X.; Castro, J.M.; Pereira, N.; Pomonis, A.; Lemoine, A.; et al. The European seismic risk model 2020 (ESRM 2020). In Proceedings of the ICONHIC 2019—2nd International Conference on Natural Hazards & Infrastructure, Chania, Greece, 23–26 June 2019.
3. Dolšek, M.; Žižmond, J.; Babič, A.; Lazar Sinković, N.; Jamšek, A.; Gams, M.; Isaković, T. *Seismic Stress Test of Building Stock in the Republic of Slovenia (2020–2050)*; University of Ljubljana, Faculty of Civil and Geodetic Engineering, Institute of Structural Engineering, Earthquake Engineering and Construction IT: Ljubljana, Slovenija, 2020. (In Slovenian)
4. Babič, A.; Dolšek, M.; Žižmond, J. Simulating Historical Earthquakes in Existing Cities for Fostering Design of Resilient and Sustainable Communities: The Ljubljana Case. *Sustainability* **2021**, *13*, 7624. [CrossRef]
5. Polese, M.; Tocchi, G.; Dolsek, M.; Babič, A.; Faravelli, M.; Quaroni, D.; Borzi, B.; Prota, A. Seismic risk assessment in transboundary areas: The case study on the border between Italy and Slovenia. *Procedia Struct. Integr.* **2023**, *44*, 123–130. [CrossRef]
6. Bal, I.E.; Bommer, J.J.; Stafford, P.J.; Crowley, H.; Pinho, R. The influence of geographical resolution of urban exposure data in an earthquake loss model for Istanbul. *Earthq. Spectra* **2010**, *26*, 619–634. [CrossRef]
7. Dabbeek, J.; Crowley, H.; Silva, V.; Weatherill, G.; Paul, N.; Nievas, C.I. Impact of exposure spatial resolution on seismic loss estimates in regional portfolios. *Bull. Earthq. Eng.* **2021**, *19*, 5819–5841. [CrossRef]
8. Kalakonas, P.; Silva, V.; Mouyiannou, A.; Rao, A. Exploring the impact of epistemic uncertainty on a regional probabilistic seismic risk assessment model. *Nat. Hazards* **2020**, *104*, 997–1020. [CrossRef]

9. Douglas, J. Physical vulnerability modelling in natural hazard risk assessment. *Nat. Hazards Earth Syst. Sci.* **2007**, *7*, 283–288. [CrossRef]

10. Ordaz, M.; Salgado-Gálvez, M.A.; Huerta, B.; Rodríguez, J.C.; Avelar, C. Considering the impacts of simultaneous perils: The challenges of integrating earthquake and tsunamigenic risk. *Disaster Prev. Manag. Int. J.* **2019**, *28*, 823–837. [CrossRef]

11. Bazzurro, P.; Park, J. The effects of portfolio manipulation on earthquake portfolio loss estimates. In Proceedings of the 10th International Conference on Applications of Statistics and Probability in Civil Engineering, Tokyo, Japan, 31 July–3 August 2007; Volume 31.

12. Senouci, A.; Bard, P.Y.; Beck, E.; Farsi, M.N.; Cartier, S. Mapping seismic vulnerability at urban scale: Discussion on relevant cartography representations and smoothing for urban planning purposes on the Oran case study. *Soil Dyn. Earthq. Eng.* **2018**, *115*, 545–563. [CrossRef]

13. Sanderson, D.; Cox, D. Comparison of national and local building inventories for damage and loss modeling of seismic and tsunami hazards: From parcel-to city-scale. *Int. J. Disaster Risk Reduct.* **2023**, *93*, 103755. [CrossRef]

14. Hansapinyo, C.; Latcharote, P.; Limkatanyu, S. Seismic building damage prediction from GIS-based building data using artificial intelligence system. *Front. Built Environ.* **2020**, *6*, 576919. [CrossRef]

15. Leggieri, V.; Mastrodonato, G.; Uva, G. GIS multisource data for the seismic vulnerability assessment of buildings at the urban scale. *Buildings* **2022**, *12*, 523. [CrossRef]

16. Zuccaro, G.; Dolce, M.; De Gregorio, D.; Speranza, E.; Moroni, C. La scheda CARTIS per la caratterizzazione tipologico-strutturale dei comparti urbani costituiti da edifici ordinari. Valutazione dell'esposizione in analisi di rischio sismico. In Proceedings of the GNGTS 2015, Trieste, Italy, 17 November 2015.

17. Brando, G.; Cianchino, G.; Rapone, D.; Spacone, E.; Biondi, S. A CARTIS-based method for the rapid seismic vulnerability assessment of minor Italian historical centres. *Int. J. Disaster Risk Reduct.* **2021**, *63*, 102478. [CrossRef]

18. Vettore, M.; Donà, M.; Carpanese, P.; Follador, V.; da Porto, F.; Valluzzi, M.R. A multilevel procedure at urban scale to assess the vulnerability and the exposure of residential masonry buildings: The case study of Pordenone, Northeast Italy. *Heritage* **2020**, *3*, 1433–1468. [CrossRef]

19. Ruggieri, S.; Calò, M.; Cardellicchio, A.; Uva, G. Analytical-mechanical based framework for seismic overall fragility analysis of existing RC buildings in town compartments. *Bull. Earthq. Eng.* **2022**, *20*, 8179–8216. [CrossRef]

20. Ruggieri, S.; Liguori, F.S.; Leggieri, V.; Bilotta, A.; Madeo, A.; Casolo, S.; Uva, G. An archetype-based automated procedure to derive global-local seismic fragility of masonry building aggregates: META-FORMA-XL. *Int. J. Disaster Risk Reduct.* **2023**, *95*, 103903. [CrossRef]

21. Pittore, M.; Haas, M.; Silva, V. Variable resolution probabilistic modeling of residential exposure and vulnerability for risk applications. *Earthq. Spectra* **2020**, *36*, 321–344. [CrossRef]

22. Nievas, C.I.; Pilz, M.; Prehn, K.; Schorlemmer, D.; Weatherill, G.; Cotton, F. Calculating earthquake damage building by building: The case of the city of Cologne, Germany. *Bull. Earthq. Eng.* **2022**, *20*, 1519–1565. [CrossRef]

23. Gomez-Zapata, J.C.; Brinckmann, N.; Harig, S.; Zafrir, R.; Pittore, M.; Cotton, F.; Babeyko, A. Variable-resolution building exposure modelling for earthquake and tsunami scenario-based risk assessment: An application case in Lima, Peru. *Nat. Hazards Earth Syst. Sci.* **2021**, *21*, 3599–3628. [CrossRef]

24. Scheingraber, C.; Käser, M.A. The impact of portfolio location uncertainty on probabilistic seismic risk analysis. *Risk Anal.* **2019**, *39*, 695–712. [CrossRef] [PubMed]

25. Menteşe, E.Y.; Cremen, G.; Gentile, R.; Galasso, C.; Filippi, M.E.; McCloskey, J. Future exposure modelling for risk-informed decision making in urban planning. *Int. J. Disaster Risk Reduct.* **2023**, *90*, 103651. [CrossRef]

26. Kalakonas, P.; Silva, V. Seismic vulnerability modelling of building portfolios using artificial neural networks. *Earthq. Eng. Struct. Dyn.* **2022**, *51*, 310–327. [CrossRef]

27. *CEN. EN1998-1*; Eurocode 8: Design of Structures for Earthquake Resistance–Part 1: General Rules; Seismic Actions and Rules for Buildings. European Committee for Standardisation: Brussels, Belgium, 2004.

28. Silva, V.; Crowley, H.; Pagani, M.; Monelli, D.; Pinho, R. Development of the OpenQuake engine, the Global Earthquake Model's open-source software for seismic risk assessment. *Nat. Hazards* **2014**, *72*, 1409–1427. [CrossRef]

29. GEM. *The OpenQuake-Engine User Manual. Global Earthquake Model (GEM) OpenQuake Manual for Engine Version 3.17.2*; Global Earthquake Model Foundation: Pavia, Italy, 2022.

30. Danciu, L.; Nandan, S.; Reyes, C.; Basili, R.; Weatherill, G.; Beauval, C.; Rovida, A.; Vilanova, S.; Sesetyan, K.; Bard, P.Y.; et al. *The 2020 Update of the European Seismic Hazard Model—ESHM20: Model Overview*; EFEHR Technical Report 001, v1.0.0; ETH Zurich: Zurich, Switzerland, 2021. [CrossRef]

31. CEN. *prEN_1998-1-1_2021_ENQ: Eurocode 8—Design of Structures for Earthquake Resistance—Part 1-1: General Rules and Seismic Action (Version of prEN 1998-1-1:2021 Submitted to TC 250 for Preparation to ENQ Stage), N1141*; CEN European Committee for Standardisation: Brussels, Belgium, 2021.

32. Žižmond, J.; Dolšek, M. The risk-targeted verification model of earthquake-induced sliding displacement of low-rise thermally base-insulated buildings. *Earthq. Eng. Struct. Dyn.* **2022**; *in print*. [CrossRef]

33. *HAZUS MH 2.1, Technical Manual, Multi-Hazard Loss Assessment Methodology*; Federal Emergency Management Agency: Washington, DC, USA, 2015.

34. Polese, M.; Gaetani d'Aragona, M.; Prota, A. Simplified approach for building inventory and seismic damage assessment at the territorial scale: An application for a town in southern Italy. *Soil Dyn. Earthq.* **2019**, *121*, 405–420. [CrossRef]

35. Kappos, A.J.; Panagopoulos, G.; Panagiotopoulos, C.; Penelis, G. A hybrid method for the vulnerability assessment of R/C and URM buildings. *Bull. Earthq. Eng.* **2006**, *4*, 391–413. [CrossRef]

36. Pitilakis, K.; Argyroudis, S.; Kakderi, K.; Argyroudi, A. *SYNER-G Project: Deliverable D8.20: Systemic Seismic Vulnerability and Risk Analysis for Buildings, Lifeline Networks and Infrastructures Safety Gain*; Aristotle University of Thessaloniki & Commission of the European Communities—Directorate General Joint Research Centre: Ispra, Italy, 2013.

37. GURS. Surveying and Mapping Authority of the Republic of Slovenia. Real Estate Register. Available online: https://eprostor. gov.si/imps/srv/api/records/26252870-5100-4408-a3e0-54ea80eb3612 (accessed on 21 May 2021).

38. Schäfer, D.; Scherer, R.; Pietsch, M.; Wenzel, H. *SYNER-G Project: Deliverable D7.1: Systemic Seismic Vulnerability and Risk Analysis for Buildings, Lifeline Networks and Infrastructures Safety Gain*; SYNER-G D7.1—Functional Fragility Curve Archive; Vienna Consulting Engineers: Vienna, Austria, 2011.

39. PEG—Gradbeni Portal—Projektanske Ocene Investicij. Ljubljana. Available online: http://www.peg-online.net/ocene-investicij (accessed on 10 February 2020).

40. Porter, K.A.; Kiremidjian, A.S. *Assembly-Based Vulnerability of Buildings and Its Uses in Seismic Performance Evaluation and Risk Management Decision-Making*; Stanford University, The John A. Blume Earthquake Engineering Center, Department of Civil and Environmental Engineering: Stanford, CA, USA, 2001.

41. Bradley, B.A. Structure-Specific Probabilistic Seismic Risk Assessment. Ph.D. Thesis, University of Canterbury, Department of Civil and Natural Resources Engineering, Christchurch, New Zealand, 2009.

42. Snoj, J.; Dolšek, M. Pushover-based seismic risk assessment and loss estimation of masonry buildings. *Earthq. Eng. Struct. Dyn.* **2020**, *49*, 567–588. [CrossRef]

43. Wesson, R.L.; Perkins, D.M.; Luco, N.; Karaca, E. Direct calculation of the probability distribution for earthquake losses to a portfolio. *Earthq. Spectra* **2009**, *25*, 687–706. [CrossRef]

44. BORIS. Deliverable 5.1: Seismic Risk, Flood Risk and Multi-Risk Assessment at Pilot Cross-Border Sites. Available online: http://www.borisproject.eu/wp-content/uploads/2022/12/BORIS-Deliverable-D5.1-Seismic-risk-flood-risk-and-multi-risk-assessment-at-pilot-cross-border-sites_final.pdf (accessed on 27 July 2023).