applied sciences

MDPI

Article

Impact and a Novel Representation of Spatial Data Uncertainty in Debris Flow Susceptibility Analysis

Laurie Jayne Kurilla and Giandomenico Fubelli *

Department of Earth Sciences, University of Turin, 10124 Torino, Italy; laurie.kurilla@edu.unito.it
* Correspondence: giandomenico.fubelli@unito.it

Abstract: In a study of debris flow susceptibility on the European continent, an analysis of the impact between known location and a location accuracy offset for 99 debris flows demonstrates the impact of uncertainty in defining appropriate predisposing factors and consequent analysis for areas of susceptibility. The dominant predisposing environmental factors, as determined through Maximum Entropy modeling, are presented and analyzed with respect to the values found at debris flow event points versus a buffered distance of locational uncertainty around each point. Maximum Entropy susceptibility models are developed utilizing the original debris flow inventory of points, randomly generated points, and two models utilizing a subset of points with an uncertainty of 5 km, 1 km, and a model utilizing only points with a known location of "exact". The AUCs are 0.891, 0.893, 0.896, 0.921, and 0.93, respectively. The "exact" model, with the highest AUC, is ignored in final analyses due to the small number of points and localized distribution, and hence susceptibility results are likely non-representational of the continent. Each model is analyzed with respect to the AUC, highest contributing factors, factor classes, susceptibility impact, and comparisons of the susceptibility distributions and susceptibility value differences. Based on model comparisons, geographic extent, and the context of this study, the models utilizing points with a location uncertainty of less than or equal to 5 km best represent debris flow susceptibility for the continent of Europe. A novel representation of the uncertainty is expressed and included in a final susceptibility map, as an overlay of standard deviation and mean of susceptibility values for the two best models, providing additional insight for subsequent action.

Keywords: spatial uncertainty; debris flow; susceptibility

Citation: Kurilla, L.J.; Fubelli, G. Impact and a Novel Representation of Spatial Data Uncertainty in Debris Flow Susceptibility Analysis. *Appl. Sci.* **2022**, *12*, 6697. https://doi.org/10.3390/app12136697

Academic Editor: Adriano Ribolini

Received: 14 April 2022
Accepted: 27 June 2022
Published: 1 July 2022

1. Introduction

Debris flows, and landslides in general, are worldwide catastrophic phenomena [1–4]. Due to an expanding population and urbanization trends, the human and economic impact due to debris flow hazards necessitates broader geographic research. At least 14% of total casualties from natural hazards are due to slope failures, and ~49% of natural hazards are landslides [5]. Between 1988 and 2017 it is estimated that there were more than 56,000 deaths worldwide due to landslides, more than 4.8 million people affected injuriously and/or economically, and about 6 billion EUR/year for damages in industrialized countries [5–9]. Although it must be pointed out that in Europe, and likely the world, the number of landslides and their societal and economic impact are grossly underestimated [10]. With peak flow speeds reaching 10 m/s and volumes approaching 10^9 m^3, debris flows pose a significant hazard to structures and lives [11,12].

As the world population and urbanization grows in number and geographic coverage [13], we realize the need to extend our focus, research, and modeling to a continental scale. Localized field surveys to collect event inventories are not a practicable approach in continental hazard susceptibility modeling. Thus, debris flow susceptibility at this scale requires data-driven and statistical methodologies, which include continental remotely sensed and aggregated coverages of environmental factors that may influence susceptibility.

Susceptibility, in this context, is a qualitative assessment of potential areas of instability with respect to debris flows.

In a landslide susceptibility analysis, landslide event location accuracy is paramount yet often inaccurately known unless a direct field survey is conducted. Landslide inventories are often constructed based on mapping from aerial imagery, media reports, local governmental agencies, witness accounts, and field work by third party sources [5,14,15]. Uncertainties are inherent in all spatial data and at all scales [16]; however, when working at a continental scale and in the absence of direct field surveys, the uncertainties are inherently greater [17,18]. "Uncertainty exists widely in the natural world, and certainty is conditional and relative" [19]. When utilizing methods that overlay and correlate multiple datasets each with their own uncertainty, their derivative products, such as susceptibility maps, are prone to error propagation of an unknown magnitude [19]. It is not a matter of adding more or better data, but rather a "sobering reminder that uncertainty is an irreducible part of sufficiently complex knowledge" [20], and thus should be thought of as a natural component of, and addressed in, every project. The presence of data uncertainty does not preclude the use of the data, but rather necessitates a methodology for qualitatively or quantitively characterizing and conveying the level of uncertainty and modeling the associated uncertainty for the benefit of the end-users of the hazard model. "Unlike industrial and other products of material processes, knowledge products do not carry with them the evidence of their own inadequacy and the most critical aspects of the quality of its products are often only testable through their indirect and sometimes remote consequences" [20]. Our aim is to minimize the societal 'testing' of an inadequately understood susceptibility model.

Although there are many factors and attributes associated with debris flow analyses which are prone to uncertainty, such as debris flow location, type, volume/size, setting, predisposing factors, triggers, etc. [14,21,22], for simplicity, in this study, only the uncertainty associated with debris flow event location is investigated. The focus herein is the impact of uncertainty on the determination of principal environmental predisposing factors, factor classes, and resulting susceptibility analyses.

The debris flow predisposing environmental factors initially employed are those that are commonly associated with debris flows [23–30]; they are aridity, climate, depth to bedrock, elevation, fault density, landcover, landform, lithology, topsoil percent clay, precipitation, slope, soil drainage, soil type, and soil thickness. The study herein is focused on the European continent, as a case study. While all fourteen environmental factors were inputs to the Maximum Entropy (MaxEnt) models, only a subset of these debris flow predisposing factors are used to demonstrate the uncertainty issues. The factors further investigated are those with the highest gain (contribution) to debris flow susceptibility as defined by MaxEnt modelling. They are precipitation, fault density, and soil type.

2. Study Area

The study area is the continent of Europe. The mean elevation of Europe is 300 m above sea level, 31% of the continent is depositional plains, 30% erosional plains and plateau in sedimentary rocks, and mountain belts account for 25% [31]. The European continent is generally characterized by a temperate climate. Soils across the continent are diverse with twenty-three out of the total of thirty Reference Soil Groups of the world being present [32]. It is expected that the uncertainty and susceptibility impact principles discussed herein, and found to be true in Europe, may be true on any of the continents.

3. Materials and Methods

Determination of the most relevant debris flow conditioning factors, as well as the factor classes, is essential to susceptibility analyses. Global coverage of fourteen conditioning factor datasets (aridity, climate, precipitation, elevation, fault density, landform, lithology, depth to bedrock, slope, soil thickness, soil type, topsoil percent clay, soil drainage, and landcover) were acquired from various sources. A European continent boundary dataset was used to clip each factor and create a Europe-only coverage. The landside inventory

was sourced from NASA [15]. The inventory contains 99 debris flows and mudslides within the continent of Europe. Mudslides are included in this study as it is a common misnomer for debris flows. It is noteworthy that landslide classification itself may be a source of data uncertainty. The event inventory contains a "location accuracy" attribute for each event. The locational accuracies associated with the 99 debris flows in Europe range from an "exact" known location to 50 km (Table 1), plus six events that were identified with a location accuracy of "unknown". Figure 1 shows the distribution of this model training data depicted with locational uncertainty buffers for the 93 events with defined locational uncertainties.

Table 1. Locational uncertainty frequency distribution of global inventory: Europe debris flows (left), world debris flows (center), and world–all landslide types (right). The Europe debris flows are used for this study.

Locational Uncertainty Radius (km)	Frequency Distribution (Europe Debris Flows)	Percent of Total Debris Flow Events	Cum Percent	Frequency Distribution (World Debris Flows)	Percent of Total Debris Flow Events	Cum Percent	Frequency Distribution (World All Landslide Types)	Percent of Total Landslide Events	Cum Percent
exact	5	5.1	5.1	174	7.6	7.6	1386	12.6	12.6
1	19	19.2	24.3	620	27.0	34.6	2185	19.8	32.4
5	31	31.3	55.6	763	33.3	67.9	3178	28.8	61.2
10	20	20.2	75.8	277	12.1	79.9	1435	13.0	74.2
25	12	12.1	87.9	240	10.5	90.4	1470	13.3	87.5
50	6	6.1	93.9	125	5.4	95.9	794	7.2	94.7
100	0			9	0.4	96.3	25	0.2	94.9
250	0			4	0.2	96.4	16	0.1	95.1
"unknown"	6	6.1	99.9	82	3.6	100.0	546	4.9	100.0
Total	**99**			**2294**			**11,033**		

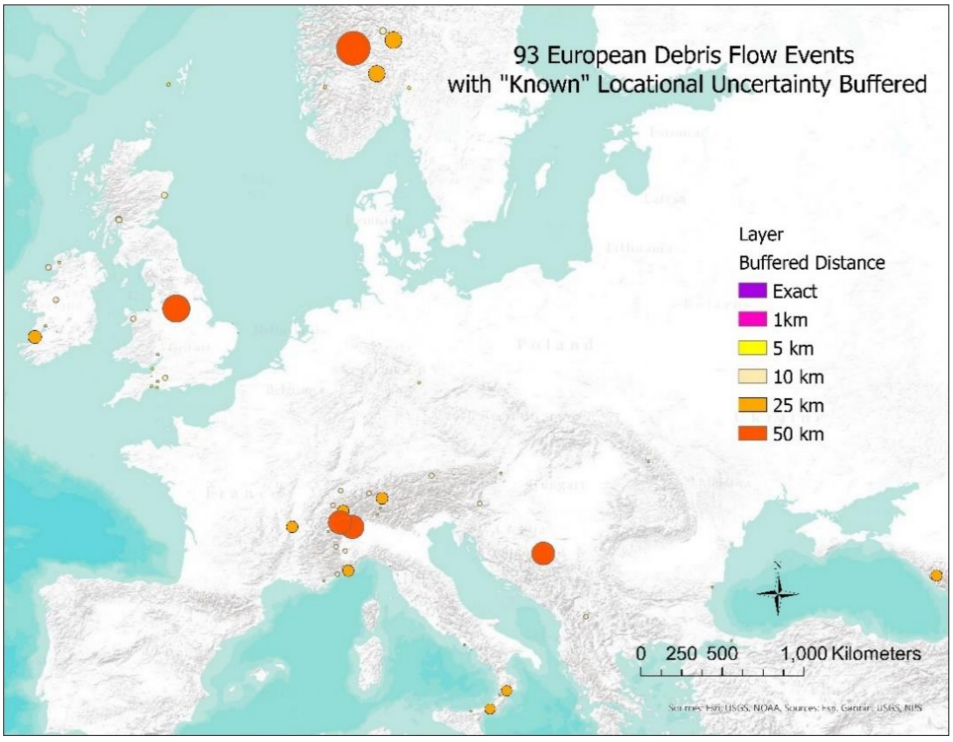

Figure 1. 93 (of 99) European debris flow events with a buffered "known" locational uncertainty. Some larger buffers overlap and occlude nearby smaller buffers. Base map is from ArcGIS®, the intellectual property of Esri, used herein under license. Copyright© Esri.

The predisposing factors were analyzed for Pearson's correlation coefficient (r) with Excel Correl function and found to have a weak to very weak correlation among all factors

(Table 2). Analysis of the predisposing factor contributions, from a Maximum Entropy model of the original 99 events, resulted in precipitation, fault density, and soil with the highest relative contributions at 42%, 27.6%, and 8.6%, respectively. Thus, for purposes of this study, the location uncertainty impact is analyzed and demonstrated for only these three factors. A fault density derivative dataset was developed with the ArcGIS "Line Density" tool and the active faults from the GEM Global Active Faults database [33], utilizing a search radius of 250 km and an output cell size of 25 sq km. The new dataset represents the fault length (km) per square kilometers of area.

Table 2. Matrix of environmental factor correlation coefficients.

	Climate Class	Distance to Fault	Drainage	Percent Clay	Lithology	Soil Type	Precip	Elev	Soil Thick	Aridity Index	Landform	Landcover	Slope	Fault Density	Depth to Bedrock
Climate Class	1	−0.180	0.107	−0.246	−0.051	−0.009	−0.109	−0.009	−0.036	−0.065	−0.029	0.041	0.026	−0.010	−0.052
Distance to Fault		1	−0.068	0.072	0.050	0.093	0.011	0.001	−0.012	0.015	0.038	−0.015	0.023	−0.016	−0.026
Drainage			1	−0.416	0.006	0.040	−0.135	−0.016	−0.001	−0.087	0.067	0.019	−0.041	−0.057	−0.017
Percent Clay				1	0.026	−0.013	0.020	−0.001	0.022	0.005	0.017	−0.025	0.024	−0.001	0.025
Lithology					1	0.087	−0.120	0.053	−0.003	−0.095	0.011	0.009	0.041	0.018	−0.052
Soil Type						1	−0.315	0.096	−0.010	−0.274	0.148	0.102	0.030	0.029	−0.107
Precip							1	−0.261	0.010	0.864	−0.436	−0.174	−0.058	−0.102	0.252
Elev								1	−0.213	−0.240	−0.070	−0.143	0.027	−0.067	−0.061
Soil Thick									1	−0.015	−0.052	0.241	0.036	0.026	0.007
Aridity Index										1	−0.416	−0.162	−0.058	−0.081	0.217
Landform											1	−0.008	−0.015	0.076	−0.096
Landcover												1	0.030	0.105	−0.047
Slope													1	0.043	−0.012
Fault Density														1	−0.016
Depth to Bedrock															1

Maximum Entropy (MaxEnt), a "presence-only" machine learning algorithm [34], is used due to the ambiguity of "absence" in this context and the dependence on landslide inventories that were not collected through manual field surveys and thus without verified locations. Absence does not necessarily mean that there are or were no debris flows in an area. It means we do not know and/or we do not have substantiating data sources or ability to conduct field surveys, particularly at a continental scale. MaxEnt utilizes the environmental data across the entire study area, rather than the area strictly associated with the presence data, which is an important factor when working with limited presence data [35]. The MaxEnt model output is a maximum likelihood estimate of relative probability of presence. MaxEnt is a widely used technique in biological species distribution modeling with recent and growing interest in its use for landslide susceptibility modeling due to its predictive success compared with other methodologies in "presence" only scenarios [27,36–40]. The MaxEnt model renders information for those debris flow predisposing factors that provide the greatest contribution to the susceptibility analysis. For this study, each MaxEnt model was run with five replications (cross-validation), logistic output format, with jackknife tests of variable importance. The MaxEnt models were developed with the event point data. Since MaxEnt requires the event data input be in point format, it alone does not lend itself to assessing the impact using debris flow locational accuracy buffered areas and associated 'what if' scenario representations. Therefore, in addition to MaxEnt software v. 3.4.4 [41] for susceptibility and environmental factor analyses, ArcGIS Pro 2.7 [42] was used for creating uncertainty buffers around each point based on its location accuracy (uncertainty), spatial calculations (pixel math for determining differences between models), statistical analyses (standard deviation, mean), and additional scenario representations. All maps presented herein include base maps from ArcGIS® and are the intellectual property of Esri and are used herein under license.

Five models were developed with MaxEnt. The first model was developed using the original inventory of 99 Europe debris flow events with varying locational uncertainties. This model will subsequently be referred to as "Original". The second model utilizes a training dataset created from 93 randomly generated points, randomly located within each of the original 93 locational uncertainty buffers. In ArcGIS, for each event buffer (regardless of location accuracy), the "Create Random Points" tool was used to generate a new point randomly located within each buffer. These new points were then used in MaxEnt to create a susceptibility model. This model is referred to as "93 Random". The third model utilizes only those original inventory event points with known locational uncertainties $\leq$ 5 km

(55 events) as the training data, including "exact" points. This is the "LTE 5km" model. The fourth model, "LTE 1km" utilizes only those original inventory event points with known locational uncertainties $\leq$ 1 km (24 events), including "exact" points. The fifth model, "Exact", utilizes only those points identified as having an "exact" known location (five events). All MaxEnt models include a test dataset composed of 2743 debris flow events in Europe, with no location accuracy attribute. The data were curated from approximately 100 governmental organizations throughout Europe. These data include landslide type and location among other attributes, but no information with regard to location accuracy. They are representative of typical landslide inventories. The "Original" model is the model against which all other models are compared, with the purpose of understanding the suitability of such an inventory of events with locational uncertainties, and the option for choosing an alternative set of events for the best susceptibility representation.

4. Results

4.1. Impact of Locational Uncertainty on Precipitation as a Predisposing Factor

According to the "Original" MaxEnt model, the monthly average precipitation predisposing factor provided the highest contribution (gain) and was the most significant factor in the jackknife test both in most significance as the only variable, and most significant negative impact when removed from the model. In a worldwide study of non-seismic landslide occurrences, ref. [5] found a strong correlation between the mean monthly precipitation and landslide events in four of five global regions studied. Table 3 shows the AUC (area under the curve) and the relative factor contribution (gain) for the "Original" MaxEnt model, and all factors input to the model.

Table 3. AUC and factor contribution derived from MaxEnt "Original" model.

Variable/Model	"Original" Factor Contribution to Model
AUC	0.891
precipitation	42.0
fault density	27.6
soil type	8.6
landcover	4.7
climate	4.5
lithology	2.1
soil thickness	2.4
landform	4.9
elevation	0.2
drainage	2.3
topsoil percent clay	0.1
depth to bedrock	0.0
aridity	0.5

Two events, 560 and 6381, each with a 50 km location uncertainty, are selected to view the precipitation factor classes within each buffer, at a larger scale. Table 4 and Figure 2 depict precipitation values and range of values found within their buffered areas.

Table 4. Impact of location uncertainty: buffered area values vs. point value of precipitation factor class, for two sample events, 560 and 6381.

Event ID	560	6381
Precipitation point value	92 mm	73 mm
Locational uncertainty buffer	50 km	50 km
Number of different precipitation factor classes within buffer	32	31
Precipitation range of values within buffer	58–141 mm	49–94 mm

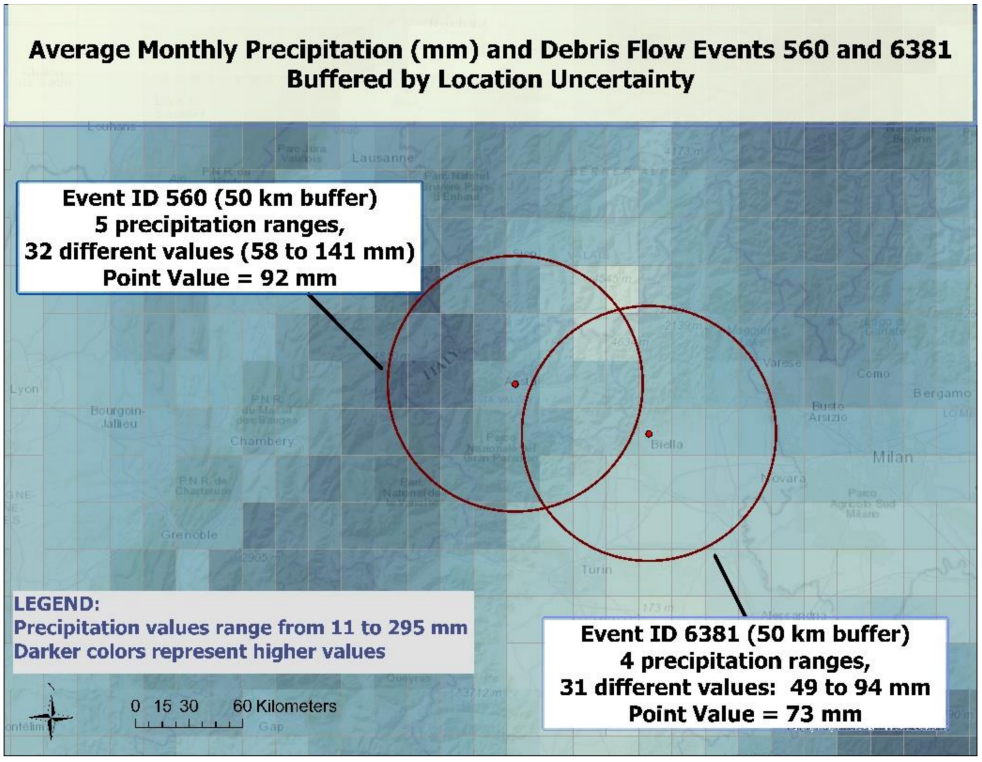

Figure 2. Map overlay of monthly average precipitation (mm) raster and debris flow event locational uncertainty buffers, highlighting event 560 and 6381, both with 50 km locational uncertainty. Each color cell represents an associated precipitation factor class value (mm). Base map is from ArcGIS®, the intellectual property of Esri, used herein under license. Copyright© Esri.

4.2. Impact of Locational Uncertainty on Fault Density as a Predisposing Factor

According to the "Original" MaxEnt model, the fault density environmental layer provided the second highest contribution (gain), at 27.6%, and was the second most significant factor in the jackknife test with most significance as the only variable. Figure 3a depicts the fault density in western Europe with debris flow locational uncertainty buffers. Table 5 and Figure 3b depict the point fault density value versus the range of values within the buffered areas, for sample events 560 and 6381.

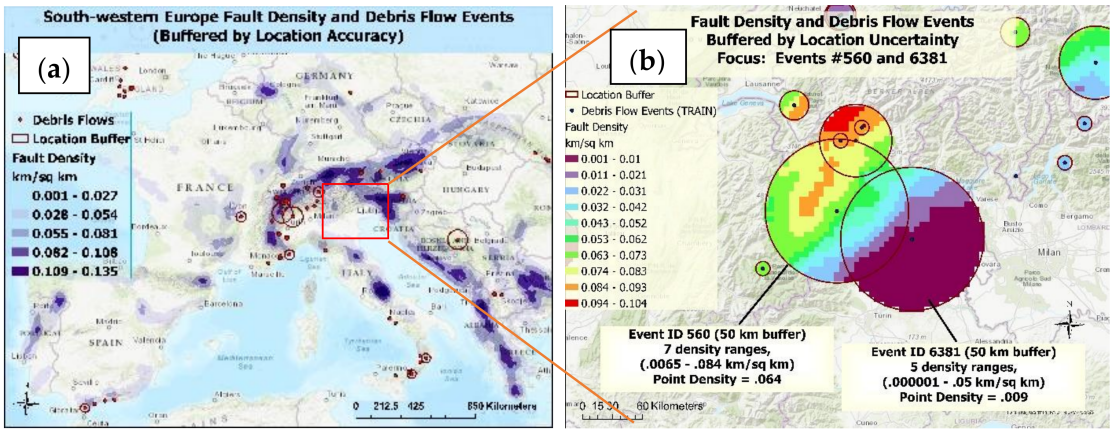

Figure 3. Fault density (km/sq km) overlay and debris flow event locational uncertainty buffers. (**a**) Fault densities in western Europe. (**b**) Fault density (Natural Break (Jenks) classification—10 breaks) and debris flow event locational uncertainty buffers, highlighting events 560 and 6381, both with 50 km location uncertainties. Base maps and imagery are from ArcGIS®, the intellectual property of Esri, used herein under license. Copyright© Esri.

Table 5. Impact of location uncertainty: buffered area values vs. point value of fault density factor class, for events 560 and 6381.

Event ID	560	6381
Single fault density class at point (km/sq km)	0.0065–0.0084	0.000001–0.05
Locational Uncertainty Buffer	50 km	50 km
Range of multiple fault density classes within buffer (km/sq km)	0.0050–0.0840	0.0–0.059

4.3. Impact of Locational Uncertainty on Soil as a Predisposing Factor

The environmental layer with the third highest contribution in the "Original" MaxEnt model is soil type with a gain of 8.6% and tied with fault density as the second most significant factor in the jackknife test with most significance as the only variable. A total of 35% of the 93 buffered events had from two to seven different soil types. Figure 4 depicts the soil types within each buffered event in a partial view of western Europe, highlighting events 560 and 6381, Table 6 presents their point value soil types versus the range of values found within their buffered locations.

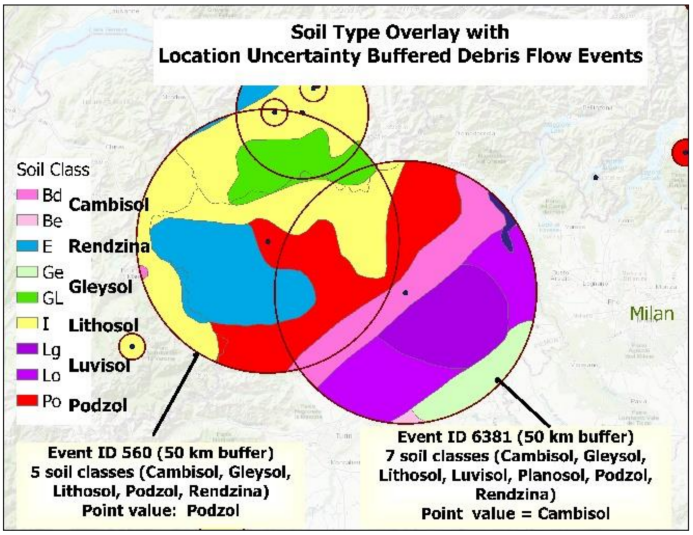

Figure 4. Events 560 and 6381 buffered by location uncertainty, with soil overlay within each buffer. Background of soil types, outside buffers, was filter out for better display. Base map is from ArcGIS®, the intellectual property of Esri, used herein under license. Copyright© Esri.

Table 6. Impact of location uncertainty (buffered area) values vs. point values in soil class determination, sample events 560 and 6831.

Event ID	560	6381
Soil class at point	Podzol	Cambisol
Locational Uncertainty Buffer	50 km	50 km
Number of different soil classes within buffer	5	7
Soil classes within buffer	Cambisol, Gleysol, Lithosol, Podzol, Rendzina	Cambisol, Gleysol, Lithosol, Luvisol, Planosol, Podzol, Rendzina

4.4. Impact of Factor Uncertainties on Susceptibility Results

Numerous debris flow susceptibility classifications are found within the location uncertainty buffers in the "Original" model, as highlighted for events 560 and 6381, Figure 5. All susceptibility classifications, from Very Low to Very High, are found within both

of the 50 km sample buffered areas. Even within a 5 km area of these points, the susceptibility ranges from Low to Very High (event 560) and Low to Moderate (event 6381), Table 7. A susceptibility classification for these, and other events with location uncertainties, is inconclusive.

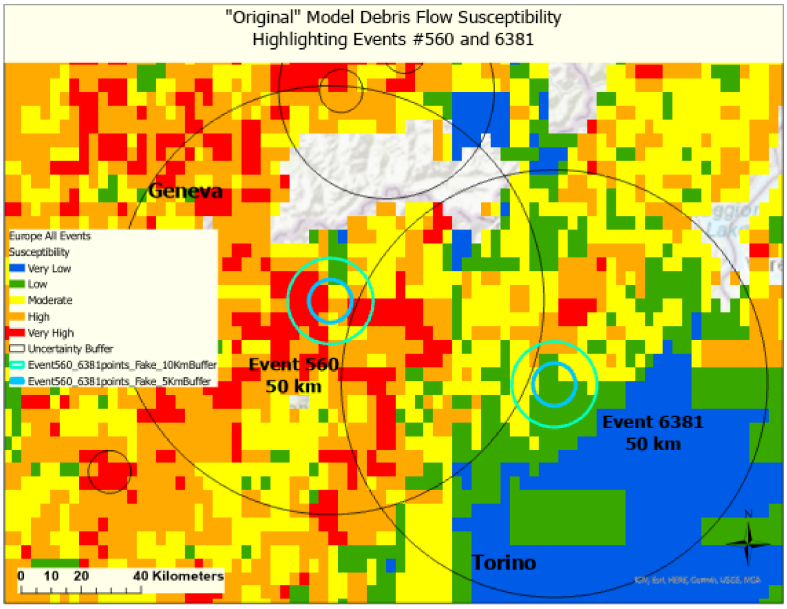

Figure 5. "Original" susceptibility model highlighting numerous susceptibility classifications within 5 to 50 km buffers for events 560 and 6381.

Table 7. "Original" susceptibility results for events 560 and 6381, within location uncertainty buffers from 5 to 50 km.

Uncertainty buffer	Event 560	Event 6381
50 km	Very Low to Very High	Very Low to Very High
10 km	Low to Very High	Low to High
5 km	Low to Very High	Low, Moderate

Debris flow susceptibility results are shown for the "Original", "93 Random", "LTE 5km", and "LTE 1km" models, Figure 6A–D, respectively, with the AUC and factor contributions for each model presented in Table 8.

Table 8. AUC and factor contributions (gain) for each of the five MaxEnt models.

Variable/Model	"Original" Factor Contribution (99 Events)	"93 Random" Factor Contribution (93 Events)	"LTE 5km" Factor Contribution (55 Events)	"LTE 1km" Factor Contribution (24 Events)	"EXACT" Factor Contribution (5 Events)
AUC	0.891	0.893	0.896	0.921	0.93
precipitation	42	37.4	30.2	11.9	3
fault density	27.6	29.7	24.8	10.3	0.1
soil type	8.6	10.5	13.6	25.8	75.3
landcover	4.7	7.9	8.9	18.8	2.3
climate	4.5	2.8	8.1	11.1	1.4
lithology	2.1	1.6	4.1	3.1	0.5
soil thickness	2.4	1.8	3.3	4	13.3
landform	4.9	3.2	3.1	10.4	2.2
elevation	0.2	1.6	2.1	1.1	1.5
drainage	2.3	2.9	1.6	1.5	0.2
topsoil percent clay	0.1	0.2	0.1	1.5	0.2
depth to bedrock	0.0	0.1	0.1	0.4	0.0
aridity	0.5	0.3	0.1	0.1	0.0

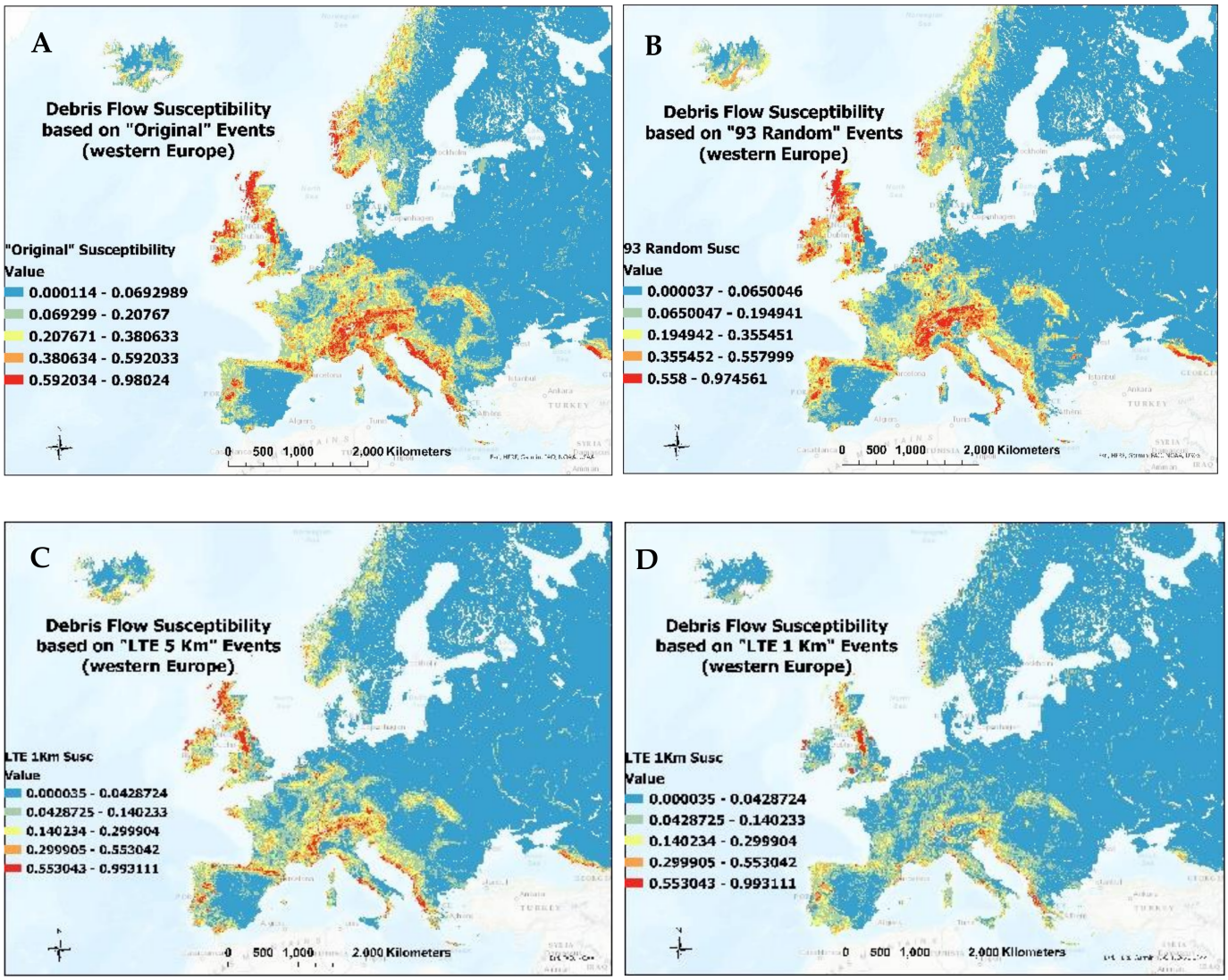

Figure 6. Debris flow susceptibility maps. (**A**). "Original", (**B**). "93 Random", (**C**). "LTE 5km", (**D**). "LTE 1km". Warmer colors represent higher susceptibilities. Base maps are from ArcGIS®, the intellectual property of Esri, used herein under license. Copyright© Esri.

The "Random" model was developed merely as a view on the integrity of the "Original" model, in comparison. So, it is noteworthy that the "Original" model does not perform as well as a model built on randomly generated locations. Although the "Exact" model has verified locations and the highest AUC, results cannot be used to represent the European continent due to the statistically small number of such events (five) and their localized distribution, Figure 7, hence the "Exact" model is not further considered.

For precipitation, soil type, soil thickness, topsoil percent clay, and landform factors, there is concurrence in the factor classes associated with susceptibility among the four models, within a reasonable margin. There are significant differences for fault density, landcover, climate, elevation, drainage, depth to bedrock, and aridity. In most of these cases there is reasonable agreement in three of the four models, Table 9.

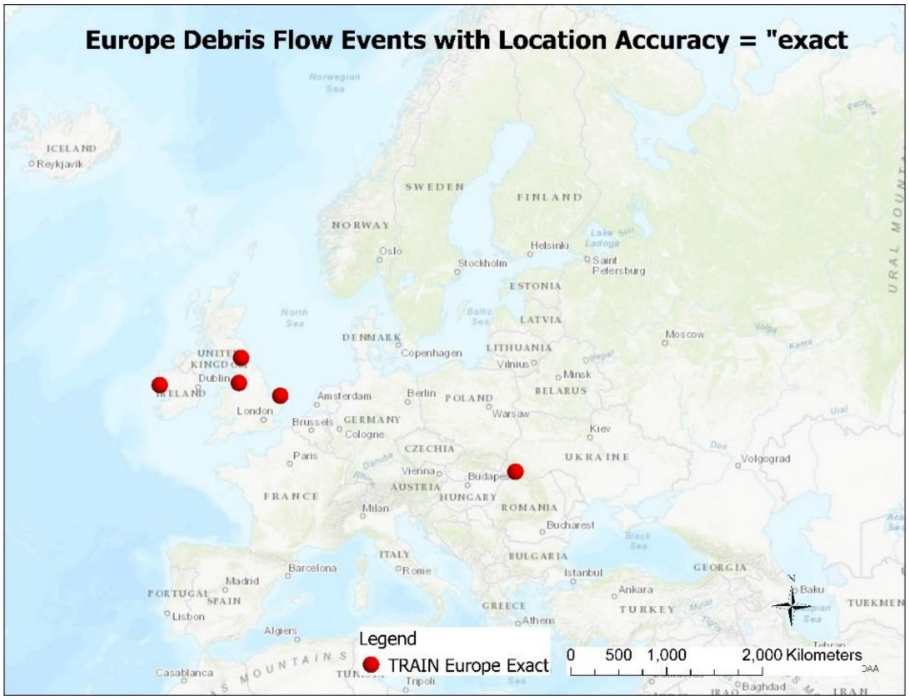

Figure 7. Distribution of the five European debris flow events with a location accuracy of "exact". Base map is from ArcGIS®, the intellectual property of Esri, used herein under license. Copyright© Esri.

Table 9. Comparison of factor classes with highest significance to susceptibility in four models. *Italicized* factors are those with closest agreement.

Factor/Model	"Original"	"93 Random"	"LTE 5km"	"LTE 1km"
precipitation (monthly average mm 1970–2000)	*275–300*	*300–325*	*260–280*	*275–300*
fault density (km/sq km)	0.02–0.14	0.01–0.02	0.13–0.14	0.01–0.14
soil type	*Gleysol*	*Fluvisol, Gleysol*	*Fluvisol, Gleysol*	*Gleysol*
landcover	urban	urban	urban	sparse vegetation
Climate (Köppen-Geiger)	Dfa—hot summer humid continental climate	BSk—semi arid steppe	Csa—Mediterranean hot summer climate and Cfc—subpolar oceanic climate	Csa—Mediterranean hot summer climate
lithology	unconsolidated sedimentary (su)	intermediate volcanic (vi)	unconsolidated sedimentary (su)	basic plutonic (pb)
soil thickness (m)	*0–2.5*	*0–1*	*0–2*	*0–2.5*
landform	*Plains on sedimentary lithology*	*Humid plains on sedimentary lithology*	*Plains on sedimentary lithology*	*Plains in alpine system*
elevation (m)	3250–3500	3200–3500	0–500	3250–5000
drainage	"Very poor"	"Imperfectly"	"Imperfectly"	"Moderately well"
topsoil percentclay	*5–18*	*20*	*20*	*0–22*
depth to bedrock (cm)	0	114	0	0–2000
aridity (dimensionless index)	~1300 (Arid)	~1300 (Arid)	≥60,000 (Humid)	~15,000 (Humid)

Next, three pairwise comparison maps were produced with pixel differencing of the susceptibility values. The first comparison is the "Original" model minus "93 Random", the second comparison is the "Original" model minus "LTE 5km", and the third is "Original" minus "LTE 1km", Figure 8A–C, respectively. The results of each map are displayed with a ten-break Natural Break (Jenks) classification. Figure 8D provides an enlarged view of the susceptibility differences for the two sample events 560 and 6381.

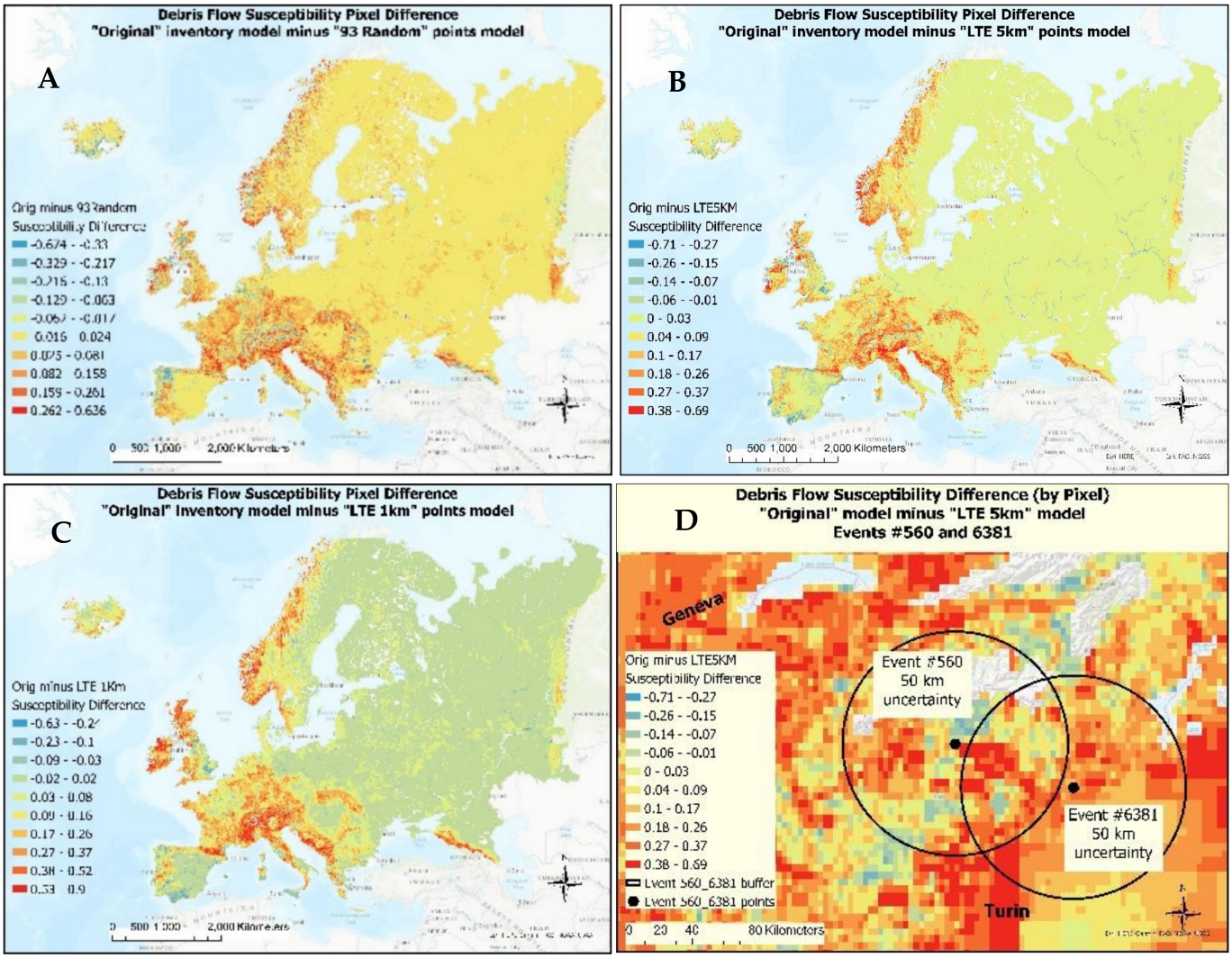

Figure 8. Model susceptibility differences. (**A**). Pixel difference between "Original" and "93 Random" models. (**B**). Pixel difference between "Original" and "LTE 5km" models. (**C**). Pixel difference between "Original" and "LTE 1km" models. Warmer colors represent areas of higher susceptibility values in "Original" model. (**D**). Map 9B zoomed into area around the Italian Alps for visual enhancement of susceptibility disparities. Base maps are from ArcGIS®, the intellectual property of Esri, used herein under license. Copyright© Esri.

5. Discussion

Identifying specific factor classes that have a dominant association with past debris flow events is essential in preparing debris flow susceptibility maps across an area of study [28] at any scale. The impact of locational uncertainty in accurately selecting those predisposing factor classes varies depending on the scale at which one is performing the analysis, the scale of lateral change of the factor itself, the extent of location uncertainty of the debris flow events being used, and the risk associated with an incorrect model [21,22]. A total of 88.8% of this study's Europe debris flow inventory has a locational uncertainty

from 1 to 50 km, with 6% defined as "unknown". A total of 82.5% of the complete inventory (all continents, all landslide types) have a locational uncertainty from 1 to 250 km, with 4.9% unknown. Noteworthy differences arise in factor and factor class determinations and susceptibility results among five MaxEnt models, each using data with different location uncertainties. Within their buffered areas, 35% of the original inventory of debris flow events were found to have from two to seven different soil types, an average of 5.8 different precipitation values, and fault densities with high values from 1.75 to 4.2 times the low values within a buffered event. The "Original" model with the most uncertainty, and the "Exact" model with no uncertainty, provide end limits on a spectrum for model comparisons. Of the five models, "Original" model has the largest Very High susceptibility classification area (Figure 6A and Table 9), five times larger than the "Exact" model, and the lowest AUC (0.891). The "93 Random" model has a slightly higher AUC (0.893) than the "Original" model, and an area of Very High susceptibility three times the size of the "Exact" model. The better performance of a random model adds to the questionability of the original dataset as satisfactory input to susceptibility modeling. "LTE 1km" (24 points with good distribution) has the smallest Very High susceptibility area (Figure 6D and Table 9), and the second highest AUC (0.921). The "LTE 5km" model, with 55 points well distributed across the continent, has an intermediate AUC (0.896) and an intermediate areal coverage of Very High susceptibility (Figure 6C and Table 10), two times larger than the "Exact" model.

Table 10. Percent area of land in the Very High susceptibility classification, for each of five models.

Model	# Events	Percent Area in Very High Susceptibility	AUC
"Original"	99	0.5	0.891
"93 Random"	93	0.3	0.893
"LTE 5km"	55	0.2	0.896
"LTE 1km"	24	0.1	0.921
"Exact"	5	0.1	0.930

Due to the small number of points in the "Exact" model (five) and localized distribution, this model is not further considered as a viable model.

Commonly, 2D model uncertainty visualization methods rely on gray scale gradations; hue, saturation, intensity coloration; fill grain density; and crispness versus fuzziness of classes, to convey model uncertainty [43,44]. These representations do not lend themselves to complex displays at large geographic extents, such as continental debris flow susceptibility. As an enhanced view on susceptibility model choices and with an attempt to better convey the 'uncertainties' to model users, mean and standard deviation overlay maps were produced based on a combination of the two best models, "LTE 5km" and "LTE 1km" (Figure 9), and with the "Original" and "LTE 5km" models (Figure 10). The former may favor the error of omission, while the latter may favor the error of commission. The better model for further investigation is dependent on the intended use. These result sets provide the end user with a novel model representation by providing the ability to select areas for further detailed study based on areas with a high mean and low standard deviation, from two models with the least location uncertainty, high AUC, and a sufficient and well-distributed sampling size. The purpose of a study, the use of a model, and the risks associated with the results, will guide a determination as to whether an error of commission (maximizing susceptibility area) or error of omission (minimizing susceptibility area) is favored.

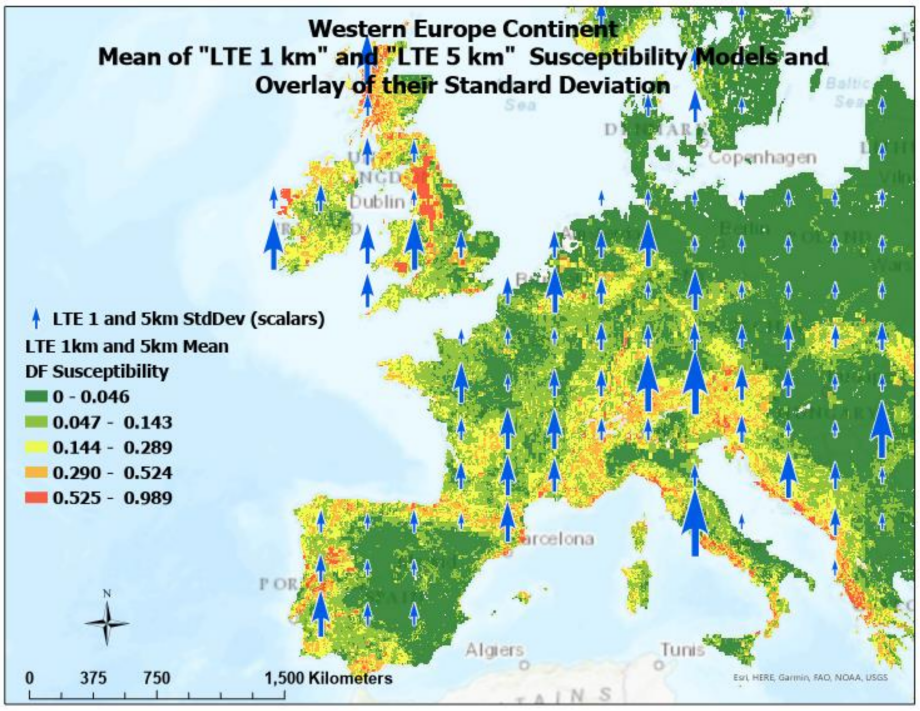

Figure 9. Overlay of susceptibility standard deviation and mean for two combined models ("LTE 5km" and "LTE 1km") zoomed into western Europe for detail. Standard deviation is represented by a scalar (blue arrows), small arrows = low values, etc. Warmers colors represent higher mean susceptibility. Base map is from ArcGIS®, the intellectual property of Esri, used herein under license. Copyright© Esri.

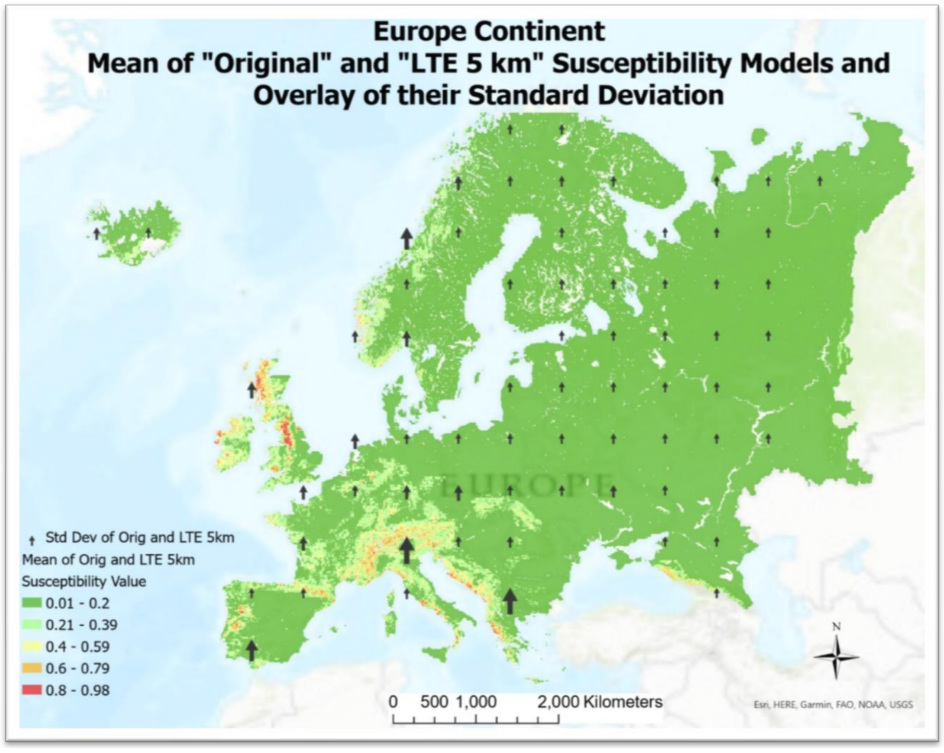

Figure 10. Overlay of susceptibility standard deviation and mean of two combined models ("Original" and "LTE 5km"). Standard deviation is represented by a scalar (black arrows), smaller arrows = lower values, etc. Warmers colors represent higher mean susceptibility. Base map is from ArcGIS®, the intellectual property of Esri, used herein under license. Copyright© Esri.

6. Conclusions

The central hypothesis of this study is the principle that locational uncertainties may be present in any landslide inventory whose data is gathered by means other than direct field survey or detailed remote sensing identification tools with a known significance level of accuracy and ground control. This study confirms that the existence of locational uncertainties has a significant impact on the resulting debris flow susceptibility analyses. It is imperative that landslide inventories include a location accuracy (uncertainty) attribute, as is true in the NASA inventory used for this study. How to approach a landslide analysis based on data without an exact known location or measure of uncertainty is beyond the scope of this study, but one which warrants further research.

In this study, a MaxEnt debris flow susceptibility model was developed utilizing the full inventory of data that has varying levels of location uncertainty. The resulting predisposing factors with the highest contribution to the model (precipitation, fault density, soil class), were investigated to identify and demonstrate the scope of impact of location uncertainty on factor class determinations and ultimately on debris flow susceptibility classification. The resulting susceptibility map demonstrated the significance of the problem of multiple susceptibility classifications within the buffered areas of uncertainty, whereby it is not possible to determine a single susceptibility for a given debris flow event, beyond those with a known "exact" location. In search for an alternative to using the full inventory, additional models were developed based on subsets of event data with differing location uncertainties.

When uncertainty is known, it is suggested to share this information with the end user in a manner that will facilitate the use of the model for subsequent more detailed analyses. "It is critical to identify, assess, and quantify uncertainty in spatial data and analysis because not accounting for uncertainty can lead to overconfidence in the conclusions" [45]. The validity and limitations of a spatial model are best understood graphically yet representing uncertainty "remains a persistent and relevant challenge" in the spatial information sciences "particularly as it applies to decision making and analysis" [46].

Therefore, a novel visual representation was developed, which allows model users to identify and use susceptibility areas of least uncertainty impact, according to their intended use.

Additional sources of data uncertainty may be inherent in other attributes of the inventory, as well as uncertainties associated both with the spatial assignment and attributes of the environmental factors, and the modeling processes. This research exclusively demonstrates the importance of event location accuracy and the impact of its uncertainty. Location uncertainty is not the only uncertainty associated with landslide susceptibility modeling but it is a fundamental one. To adequately highlight the impact, it is treated singularly, ignoring other potential uncertainty variables. Modeling all possible uncertainties is outside the scope of this study, and we believe it is unlikely that additional uncertainties would lessen the impact of locational uncertainty.

The cause, nature, and handling of mapping errors (uncertainties) has been the subject of extensive research [17–19,47–50]. Uncertainty and uncertainty handling is context dependent. The combination of uncertainties in data, analytical methods, and the overlay processes, may be additive, multiplicative, and non-linear [51]. If a susceptibility map, at any scale, is to be used to direct further research or mitigation efforts, a confidence level associated with the results is desirable [52]. This confidence will be dependent upon many facets of the model, one of which is the reliability of the location of the events used to train the model.

Increasingly, dataset producers are including accuracy (uncertainty) attributions, such as the 'location accuracy' attribute in the NASA landslide inventory [15], used for this study, and the 'epistemic quality' and 'activity confidence' attributes of the GEM Active Faults data [34]. Utilizing these attributes, when available, provides the researcher with insight as to how or if to utilize the data and/or the uncertainty, based on the context of their project. This study does not propose how to handle the use of events whose locations are not known

with certainty, but rather provides a cautionary example as to the potential limitations of the validity of the results. However, based on this study, events with unknown accuracy may be addressed by creating buffers of various (arbitrary) distances around each of the event points during modeling, which can provide an awareness of the potential scope of result variances.

According to Goodchild et al. 1993 as cited in [53] there are three options for handling uncertainty: (1) omit all reference to it; (2) attach a descriptor to the output; and (3) show samples from the range of possible maps. The latter approach was chosen in this study, that is, to present different models of the susceptibility and choose that which is most reasonable for the project at hand.

Author Contributions: L.J.K. designed the project, conducted the data collection and preparation, analyses of results, model validations, and prepared the draft manuscript. G.F. provided advice and feedback in the process. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: The authors wish to acknowledge Francesco Dramis, Universita' Degli Studi Roma Tre, for review of this manuscript and suggestions.

Conflicts of Interest: The authors declare that they have no conflict of interest.

References

1. Brabb, E.E.; Colgan, J.P.; Best, T.C. *Map Showing Inventory and Regional Susceptibility for Holocene Debris Flows and Related Fast-Moving Landslides in the Conterminous United States: Raster Data*; USGS: Reston, VA, USA, 1999. [CrossRef]
2. Brighenti, R.; Segalini, A.; Ferrero, A.M. Debris flow hazard mitigation: A simplified analytical model for the design of flexible barriers. *Comput. Geotech.* **2013**, *54*, 1–15. [CrossRef]
3. Campbell, R.H. Debris flows originating from soil slips during rainstorms in Southern California. *J. Eng. Geol.* **1974**, *7*, 339–349. [CrossRef]
4. Dowling, C.A.; Santi, P.M. Debris flows and their toll on human life: A global analysis of debris-flow fatalities from 1950 to 2011. *Nat. Hazards* **2013**, *71*, 203–227. [CrossRef]
5. Froude, M.J.; Petley, D.N. Global fatal landslide occurrence from 2004 to 2016. *Nat. Hazards Earth Syst. Sci.* **2018**, *18*, 2161–2181. [CrossRef]
6. Costa, J.E. Physical Geomorphology of Debris Flows. In *Developments and Applications of Geomorphology*; Costa, J.E., Fleisher, P.J., Eds.; Springer: Berlin/Heidelberg, Germany, 1984; pp. 268–317. [CrossRef]
7. Highland, L.M.; Bobrowsky, P. *The Landslide Handbook—A Guide to Understanding Landslides*; US Geological Survey: Reston, VA, USA, 2008.
8. Froude, M.J.; Petley, D.N. *Guide to the Global Fatal Landslide Database on ArcGIS Online*; Esri: Redlands, CA, USA, 2019.
9. ISPRA. *World Landslides Forum: Each Year, Landslide Cause Damage to 6 Billion Euros*; ISPRA: Roma, Italy, 2020.
10. Gunther, A.; Reichenbach, P.; Guzzetti, F.; Richter, A. Criteria for the identification of landslide risk areas in Europe: The Tier 1 approach. In *Guidelines for Mapping Areas at Risk of Landslides in Europe*; Joint Research Center, Institute for Environment and Sustainability, European Commission: Luxembourg, 2007; pp. 37–40. [CrossRef]
11. Iverson, R.M. The Physics of Debris Flows. *Am. Geophys. Union* **1997**, *35*, 245–296. [CrossRef]
12. Nettleton, I.M.; Martin, S.; Hencher, S.; Moore, R. Debris Flow Types and Mechanisms. In *Scottish Road Network Landslide Study Scotland*; Transport Research Laboratory: Crowthorne, UK, 2005.
13. Ritchie, H.; Roser, M. Urbanization. 2018. Available online: https://ourworldindata.org/urbanization (accessed on 25 April 2021).
14. Malamud, B.D.; Turcotte, D.L.; Guzzetti, F.; Reichenbach, P. Landslide inventories and their statistical properties. *Earth Surf. Processes Landf.* **2004**, *29*, 687–711. [CrossRef]
15. Kirschbaum, D.; Stanley, T.; Zhou, Y.P. Spatial and temporal analysis of a global landslide catalog. *Geomorphology* **2015**, *249*, 4–15. [CrossRef]
16. Chrisman, N. Modeling error in overlaid categorical maps. In *The Accuracy of Spatial Databases*; Goodchild, M., Gopal, S., Eds.; Taylor & Francis: Philadelphia, PA, USA, 1989; pp. 21–43.
17. Maffini, G.; Arno, M.; Bitterlich, W. Observations and comments on the generation and treatment of error in digital GIS data. In *Accuracy of Spatial Databases*; Goodchild, M., Gopal, S., Eds.; Taylor & Francis: Philadelphia, PA, USA, 1989; pp. 55–68.
18. Openshaw, S. Learning to live with errors in spatial databases. In *Accuracy of Spatial Databases*; Goodchild, M., Gopal, S., Eds.; Taylor & Francis: Philadelphia, PA, USA, 1989; pp. 263–276.
19. Shi, W. *Principles of Modeling Uncertainties in Spatial Data and Spatial Analyses*; CRC Press: Boca Raton, FL, USA, 2010.
20. Couclelis, H. The Certainty of Uncertainty: GIS and the Limits of Geographic Knowledge. *Trans. GIS* **2003**, *7*, 165–175. [CrossRef]
21. Carrara, A.; Cardinali, M.; Guzzetti, F. Uncertainty in assessing landslide hazard and risk. *ITC J.* **1992**, *2*, 172–183.

22. Ardizzone, F.; Cardinali, M.; Carrara, A.; Guzzetti, F.; Reichenbach, P. Impact of mapping errors on the reliability of landslide hazard maps. *Nat. Hazards Earth Syst. Sci.* **2002**, *2*, 3–14. [CrossRef]
23. Lorente, A.; García-Ruiz, J.M.; Beguería, S.; Arnáez, J. Factors Explaining the Spatial Distribution of Hillslope Debris Flows. *Mt. Res. Dev.* **2002**, *22*, 32–39. [CrossRef]
24. Devkota, K.C.; Regmi, A.D.; Pourghasemi, H.R.; Yoshida, K.; Pradhan, B.; Ryu, I.C.; Dhital, M.R.; Althuwaynee, O.F. Landslide susceptibility mapping using certainty factor, index of entropy and logistic regression models in GIS and their comparison at Mugling–Narayanghat road section in Nepal Himalaya. *Nat. Hazards* **2012**, *65*, 135–165. [CrossRef]
25. Grozavu, A.; Patriche, C.V. Landslide Susceptibility Assessment: GIS Application to a Complex Mountainus Environment. In *The Carpathians: Integrating Nature and Society towards Sustainability*; Springer: Berlin/Heidelberg, Germany, 2013. [CrossRef]
26. Dou, J.; Tien Bui, D.; Yunus, A.P.; Jia, K.; Song, X.; Revhaug, I.; Xia, H.; Zhu, Z. Optimization of Causative Factors for Landslide Susceptibility Evaluation Using Remote Sensing and GIS Data in Parts of Niigata, Japan. *PLoS ONE* **2015**, *10*, e0133262. [CrossRef] [PubMed]
27. Meten, M.; PrakashBhandary, N.; Yatabe, R. Effect of Landslide Factor Combinations on the Prediction Accuracy of Landslide Susceptibility Maps in the Blue Nile Gorge of Central Ethiopia. *Geoenvironmental Disasters* **2015**, *2*, 9. [CrossRef]
28. Kornejady, A.; Ownegh, M.; Bahremand, A. Landslide susceptibility assessment using maximum entropy model with two different data sampling methods. *Catena* **2017**, *152*, 144–162. [CrossRef]
29. Kirschbaum, D.; Stanley, T. Satellite-Based Assessment of Rainfall-Triggered Landslide Hazard for Situational Awareness. *Earth's Future* **2018**, *6*, 505–523. [CrossRef]
30. Nsengiyumva, J.B.; Luo, G.; Nahayo, L.; Huang, X.; Cai, P. Landslide Susceptibility Assessment Using Spatial Multi-Criteria Evaluation Model in Rwanda. *Int. J. Environ. Res. Public Health* **2018**, *15*, 243. [CrossRef]
31. Bridges, E.M. *World Geomorphology*; Cambridge University Press: Cambridge, UK, 2012. [CrossRef]
32. Toth, G.; Montanarell, L.; Stolbovoy, F.; Bodis, K.; Jones, A.; Panagos, P.; Liedekerke, M. *Soils of the European Union*; European Commissiono Joint Research Centre Institute for Environment and Sustainability: Roma, Italy, 2008; pp. 1018–5593. [CrossRef]
33. Styron, R.; Pagani, M. The GEM Global Active Faults Database. *Earthq. Spectra* **2020**, *36*, 160–180. [CrossRef]
34. Phillips, S.; Dudik, M. Modeling of species distributions with Maxent: New extensions and a comprehensive evaluation. *Ecography* **2008**, *31*, 161–175. [CrossRef]
35. Phillips, S.J.; Elith, J. On estimating probability of presence from use-availability or presence-background data. *Ecology* **2013**, *94*, 1409–1419. [CrossRef]
36. Convertino, M.; Troccoli, A.; Catani, F. Detecting fingerprints of landslide drivers: A MaxEnt model. *J. Geophys. Res. Earth Surf.* **2013**, *118*, 1367–1386. [CrossRef]
37. Park, N.-W. Using maximum entropy modeling for landslide susceptibility mapping with multiple geoenvironmental data sets. *Environ. Earth Sci.* **2014**, *73*, 937–949. [CrossRef]
38. Lombardo, L.; Fubelli, G.; Amato, G.; Bonasera, M. Presence-only approach to assess landslide triggering-thickness susceptibility: A test for the Mili catchment (north-eastern Sicily, Italy). *Nat. Hazards J. Int. Soc. Prev. Mitig. Nat. Hazards* **2016**, *84*, 565–588. [CrossRef]
39. Yuan, S.; Huang, G.; Xiong, H.; Gong, Q.; Wang, J.; Chen, J. Maximum Entropy-Based Model of High-Threat Landslide Disaster Distribution in Zhaoqing, China. *J. Risk Anal. Crisis Response* **2017**, *7*, 108–126. [CrossRef]
40. GÁL, A.; Poszet, S.L.; Kerekes, A.H. Landslide susceptibility assessment using the maximum entropy model in a sector of the Cluj–Napoca Municipality, Romania. *Rev. Geomorfol.* **2018**, *20*, 130–146. [CrossRef]
41. Phillips, S.; Dudik, M.; Schapire, R.E. Maxent Software for Modeling Species Niches and Distributions (Version 3.4.1) [Code]. 2021. Available online: http://biodiversityinformatics.amnh.org/open_source/maxent/ (accessed on 20 April 2021).
42. Esri, I. *ArcGIS Pro (Version 2.7)*; ESRI, Inc.: Redlands, CA, USA, 2020.
43. Hengl, T. Visualisation of uncertainty using the HSI colour model: Computations with colours. In Proceedings of the Seventh International Conference on GeoComputation, Southampton, UK, 8–10 September 2003.
44. Ślusarski, M.; Jurkiewicz, M. Visualisation of Spatial Data Uncertainty. A Case Study of a Database of Topographic Objects. *ISPRS Int. J. Geo-Inf.* **2019**, *9*, 16. [CrossRef]
45. Li, L. Spatial data uncertainty. In *The Geographic Information Science & Technology Body of Knowledge*, 4th Quarter 2017 ed.; Wilson, J.P., Ed.; Association of American Geographers: Washington, DC, USA, 2017.
46. MacEachren, A.M.; Robinson, A.; Hopper, S.; Gardner, S.; Murray, R.; Gahegan, M.; Hetzler, E. Visualizing Geospatial Information Uncertainty: What We Know and What We Need to Know. *Cartogr. Geogr. Inf. Sci.* **2005**, *32*, 139–160. [CrossRef]
47. Hunter, G.J.; Goodchild, M. Communicating uncertainty in spatial databases. *Trans. GIS* **1996**, *1*, 13–24. [CrossRef]
48. Hunter, G.J. Managing uncertainty in GIS. *Geogr. Inf. Syst.* **1999**, *2*, 633–641.
49. Wechsler, S.P. Digital Elevation Model (DEM) Uncertainty: Evaluation and Effect on Topographic Parameters. In Proceedings of the ESRI User Conference, San Diego, CA, USA, 26–30 July 1999.
50. Zufle, A.; Trajcevski, G.; Pfoser, D.; Renz, M.; Rice, M.T.; Leslie, T.; Delamater, P.; Emrich, T. Handling Uncertainty in Geo-Spatial Data. In Proceedings of the 2017 IEEE 33rd International Conference on Data Engineering (ICDE), San Diego, CA, USA, 19–22 April 2017. [CrossRef]
51. Veregin, H. Error modeling for the map overlay operation. In *The Accuracy of Spatial Databases*; Goodchild, M., Gopal, S., Eds.; Taylor & Francis: Philadelphia, PA, USA, 1989; pp. 3–18.

52. Soma, A. Landslide Susceptibility Map Using Certainty Factor for Mitigation in Mountainous Area of Ujung-Loe Watershed South Sulawesi Indonesia. *For. Soc.* **2018**, *2*, 79–91.
53. Hunter, G.J.; Goodchild, M.; Robey, M. A Toolbox for Assessing Uncertainty in Spatial Databases. In Proceedings of the 22nd Annual Conference of the Australasian Urban and Regional Information Systems Association, Sydney, NSW, Australia, 21–25 November 1994.