Impact of spatial data uncertainty in debris flow susceptibility analysis

5   Laurie J. Kurilla[1] and Giandomenico Fubelli[1]

[1]University of Turin, Department of Earth Sciences

Correspondence to: Laurie J. Kurilla (Laurie.Kurilla@edu.unito.it)

10   **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
15   analyzed with respect to the values found at debris flow event points versus a buffered distance of locational uncertainty around each point.

Five 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
20   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 likely non-representational of the continent.

Each model is analyzed with respect to the AUC, highest contributing factors, factor classes, susceptibility impact, and
25   comparisons of the susceptibility distributions and susceptibility value differences.

Based on model comparisons, geographic extent and 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 of 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
30   and mean of susceptibility values for the two best models, providing additional insight for subsequent action.

1 Introduction

Debris flows, and landslides in general, are worldwide catastrophic phenomena (Brabb et al., 1999; Brighenti et al., 2013;
35 Campbell, 1974; Dowling and Santi, 2013). 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 (Froude and Petley, 2018). Between 1988-2017 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 euro/year for damages in industrialized countries (Costa, 1984; Highland
40 and Bobrowsky, 2008; Froude and Petley, 2018, 2019; Ispra, 2020). 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 (Gunther et al., 2007). With peak flow speeds reaching 10 m/s and volumes approaching 10^9 km^3, debris flows pose a significant hazard to structures and lives (Iverson, 1997; Nettleton et al., 2005).

45 As the world population and urbanization grows in number and geographic coverage (Ritchie and Roser, 2018), 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 which may influence susceptibility. Susceptibility, in this context, is a qualitative assessment of
50 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 (Malamud et al., 2004;
55 Kirschbaum et al., 2015; Froude and Petley, 2018).

Uncertainties are inherent in all spatial data and at all scales (Chrisman, 1989), however when working at a continental scale, and in the absence of direct field surveys, the uncertainties are inherently greater (Maffini et al., 1989; Openshaw, 1989). "Uncertainty exists widely in the natural world, and certainty is conditional and relative" (Shi, 2010). When utilizing methods
60 which 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 (Shi, 2010). 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" (Couclelis, 2003),

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

65   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." (Couclelis, 2003). Our aim is to minimize the societal 'testing' of an inadequately understood susceptibility model.

70

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. (Carrara et al., 1992; Ardizzone et al., 2002; Malamud et al., 2004), 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

75   factors, factor classes, and resulting susceptibility analyses.

The debris flow predisposing environmental factors initially employed are those that are commonly associated with debris flows (Lorente et al., 2002; Devkota et al., 2012; Grozavu and Patriche, 2013; Dou et al., 2015; Meten et al., 2015; Kornejady et al., 2017; Kirschbaum and Stanley, 2018; Nsengiyumva et al., 2018), they are aridity, climate, depth to bedrock, distance

80   to faults, distance to rivers, drainage, elevation, fault density, landcover, landform, lithology, topsoil % clay, precipitation, slope, soil type, and soil thickness. The study herein is focused on the European continent, with only a subset of these debris flow predisposing factors to demonstrate the issues. The factors further investigated are those with the highest gain (contribution) to debris flow susceptibility as defined by Maximum Entropy modelling. They are precipitation, fault density, and soil type. The landside inventory was sourced from NASA (Kirschbaum et al., 2015). This inventory contains 11,033

85   landslides of various types (Table 1), and locational uncertainties. It contains 194 debris flows and 2100 mud slides, herein collectively referred to as "debris flows", 99 are 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 locational accuracy of the global inventory ranges from "exact" to 250 kilometers. The locational accuracies associated

90   with the 99 debris flows in Europe range from an "exact" known location to 50 kilometers (Table 2), plus six events which were identified with a location accuracy of "unknown". Fig. 1 shows the distribution of the training data depicted with locational uncertainty buffers for the 93 events with defined locational uncertainties.

Landslide Type	Frequency
<null>	1
complex	232
creep	5
debris flow	194
earth flow	7
lahar	7
landslide	7648
mudslide	2100
other	68
riverbank collapse	37
rock fall	671
snow avalanche	15
topple	1
translational slide	9
unknown	38

Table 1 Landslide type categorization of global landslide inventory.

Locational Accuracy	Frequency Distribution	% of Total Events	% of Total minus "unknown"
exact	5	5.1	5.4
1 km	19	19.2	20.4
5 km	31	31.3	33.3
10 km	20	20.2	21.5
25 km	12	12.1	12.9
50 km	6	6.0	6.4
unknown	6	6.0	-

Table 2 Locational uncertainty associated with 99 European debris flows.

95

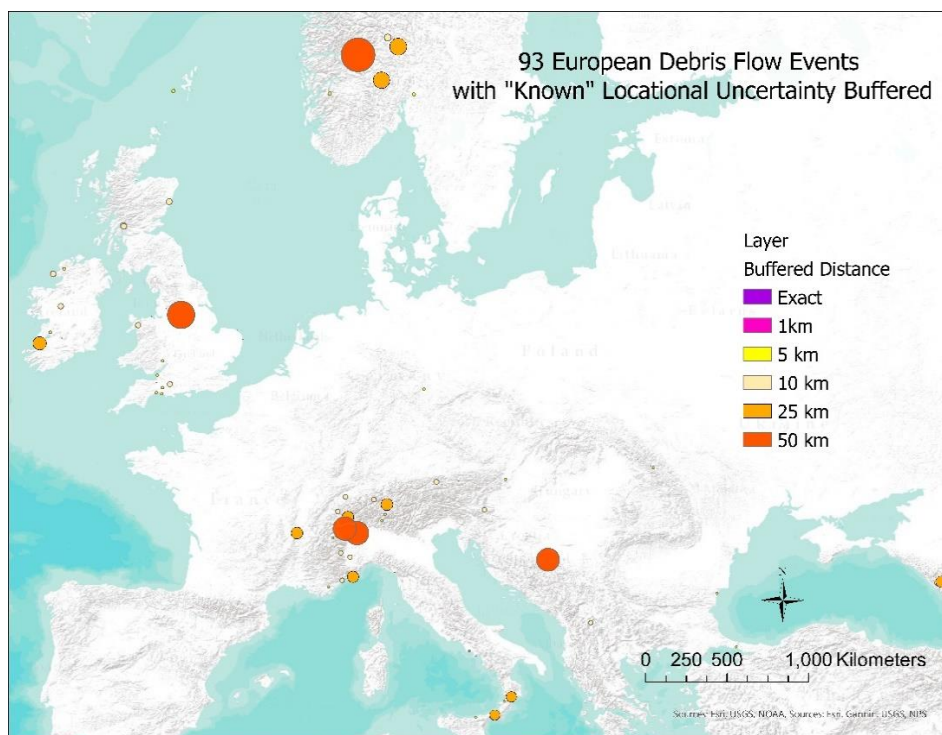



Figure 1 93 (of 99) European debris flow events with a buffered "known" locational uncertainty. Some larger buffers overlap and occlude nearby smaller buffers. Base map is from ArcGIS®, the intellectual property of Esri, used herein under license. Copyright © Esri.

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% (Bridges, 2012). 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 (Toth et al., 2008). 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 Data and Methodology

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 % 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. 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.

Maximum Entropy (MaxEnt) a "presence-only" machine learning algorithm (Phillips and Dudik, 2008) 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 is a widely used technique in biological species distribution modeling with recent and growing interest in its use for landslide susceptibility modelling due to its predictive success compared with other methodologies in "presence" only scenarios (Convertino et al., 2013; Park, 2014; Lombardo et al., 2016; Kornejady et al., 2017; Yuan et al., 2017; Gál et al., 2018). The MaxEnt model renders information for those debris flow predisposing factors that provide the greatest contribution to the susceptibility analysis. 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 (Phillips et al., 2021) for susceptibility

and environmental factor analyses, ArcGIS Pro 2.7 (Esri, 2020) was used for data processing, spatial calculations, statistical

130   analyses, 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. Copyright © Esri. All rights reserved.

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

135   model utilizes a training dataset created from 93 randomly generated points, one randomly located within each of the original 93 locational uncertainty buffers. This model is referred to as "93 Random". The third model utilizes only those original inventory event points with known locational uncertainties <= 5km (55 events) as the training data. The "exact" points are included. This is the "LTE 5km" model. The fourth model, "LTE 1km" utilizes only those original inventory event points with known locational uncertainties <= 1km (24 events), including "exact" points, as the training data. The fifth model,

140   "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, collected from numerous sources.

The "Original" model is the model against which all other models are compared, with the purpose of understanding the

145   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 Impact Analysis

4.1 Impact of locational uncertainty on precipitation as a predisposing factor

150   According to the "Original" MaxEnt model, the monthly average precipitation predisposing factor provided the highest percent 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, Froude and Petley (2018) 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 factor percent contribution (gain) for the

155   "Original" MaxEnt model, and all factors input to the model.

160

165   Table 4 describes the precipitation value range found by an overlay of the 93 "known" locational uncertainty buffered areas
and the precipitation factor layer. Two points, events 560 and 6381, each with a 50 km location uncertainty, are selected to
view the data and impact at a larger scale. Table 5 and Fig. 2 depict precipitation values and range of values found within their
buffered areas.

170

Variable/Model	"Original" Percent contribution (99 points)
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 % clay	0.1
depth to bedrock	0
aridity	0.5

Table 3. AUC and factor percent contribution for "Original" for MaxEnt susceptibility model.

175

Maximum number of different precipitation values associated within all buffered events	46
Maximum value spread in all buffered events	92 mm
Average number of different precipitation values associated with buffered events	5.8
Average value spread in buffered events	13.6 mm

Table 4 Range and number of varying precipitation values associated with all 93 debris flow event buffered areas.

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

Table 5 Example impact of uncertain location: buffered area values vs point value on associated precipitation factor class determination, for two sample events, 560 and 6381.

180

185

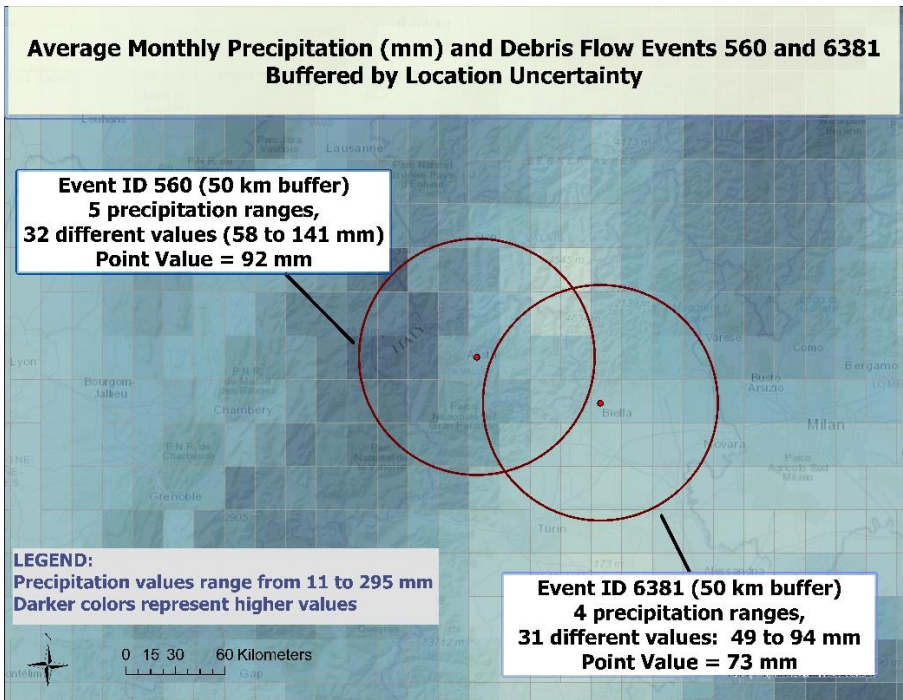



Figure 2. Map overlay of monthly average precipitation (mm) raster and debris flow event locational uncertainty buffers, highlighting event 560 and 6381, both with 50 km locational uncertainty. Each color cell represents an associated precipitation (mm) at that cell. Base map is from ArcGIS®, the intellectual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

190

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 2nd highest percent contribution (gain), at 27.6%, and was the 2nd most significant factor in the jackknife test with most significance as the only variable. Fig.

195 3(a) depicts the fault density in western Europe with debris flow locational uncertainty buffers. Table 6 and Fig. 3(b) depict the point fault density value versus the range of values within sample events 560 and 6381.

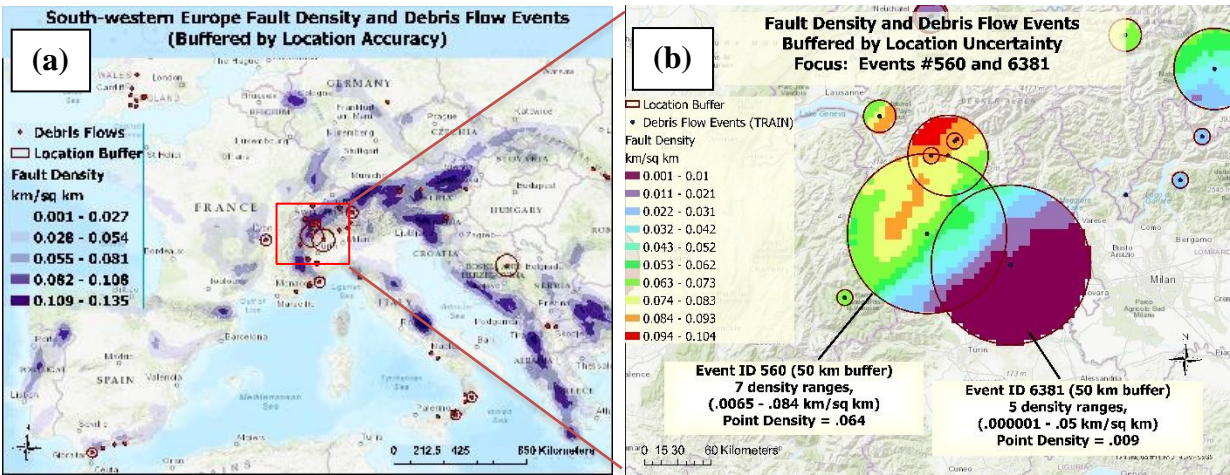



Figure 3 Fault density (km/sq. km) overlay and debris flow event locational uncertainty buffers. (a) Fault densities in western Europe. (b) Fault density (Natural Break (Jenks) classification – 10 breaks) and debris flow event locational uncertainty buffers, highlighting events 560 and 6381, both with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

Event ID	560	6381
Fault Density point value (km/sq km)	.0065-.0084	.000001-.05
Locational Uncertainty Buffer	50 km	50 km
Number of different Natural Breaks (Jenks) categories within buffer	7	6
Fault density range of values within buffer	.0050-.0840	0-.059

Table 6. Example impact of uncertain location buffered area values vs point value of associated fault density factor class determination, for sample events 560 and 6381.

4.3 Impact of locational uncertainty on soil as a predisposing factor

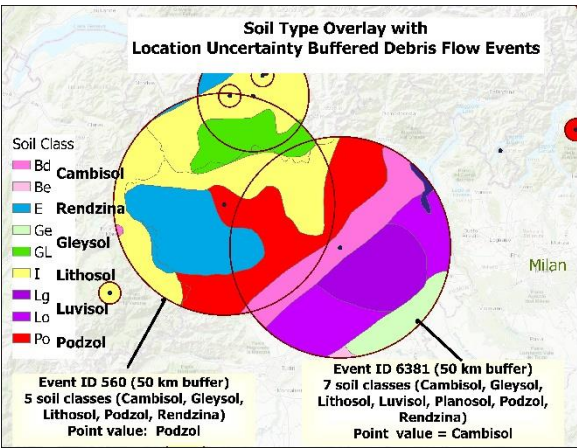
The environmental layer with the 3rd highest percent contribution in the "Original" MaxEnt model is soil type with a gain of 8.6% and tied with fault density as the 2nd most significant factor in the jackknife test with most significance as the only variable. 35% of the 93 buffered events had from two to seven different soil types. Fig. 4 depicts the soil types within each buffered event in a partial view of western Europe, highlighting sample events 560 and 6381, Table 7 presents their point value soil types versus the range of values within their buffered locations.

Event ID	560	6381
Soil type point value	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

Table 7. Example impact of location uncertainty (buffered area) values vs point values in soil class determination, sample events #560 and 6831.

Figure 4. Events #560 and 6381 buffered by location
215 **uncertainty, with soil overlay within each buffer**. Base map
is from ArcGIS®, the intellectual property of Esri, used herein
under license. Copyright © Esri.

220 **4.4 Impact of factor uncertainties on susceptibility results**

Susceptibility maps were generated by MaxEnt for the models, "Original", "93 Random", "LTE 5km", and "LTE 1km", Fig.
5A, 5B, 5C, 5D, respectively.

225

230

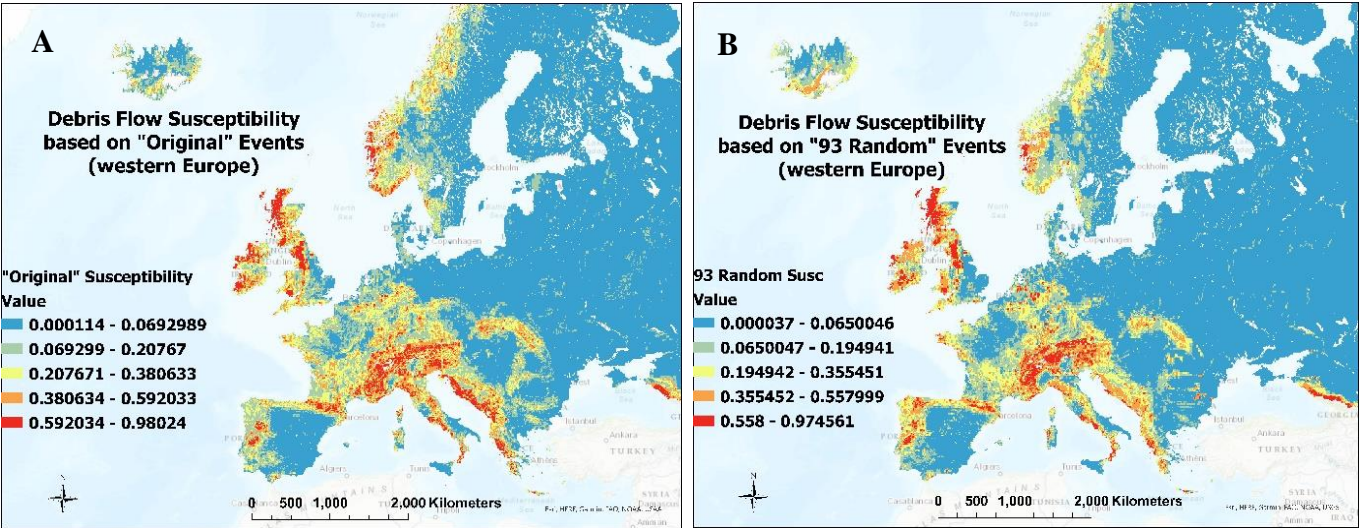
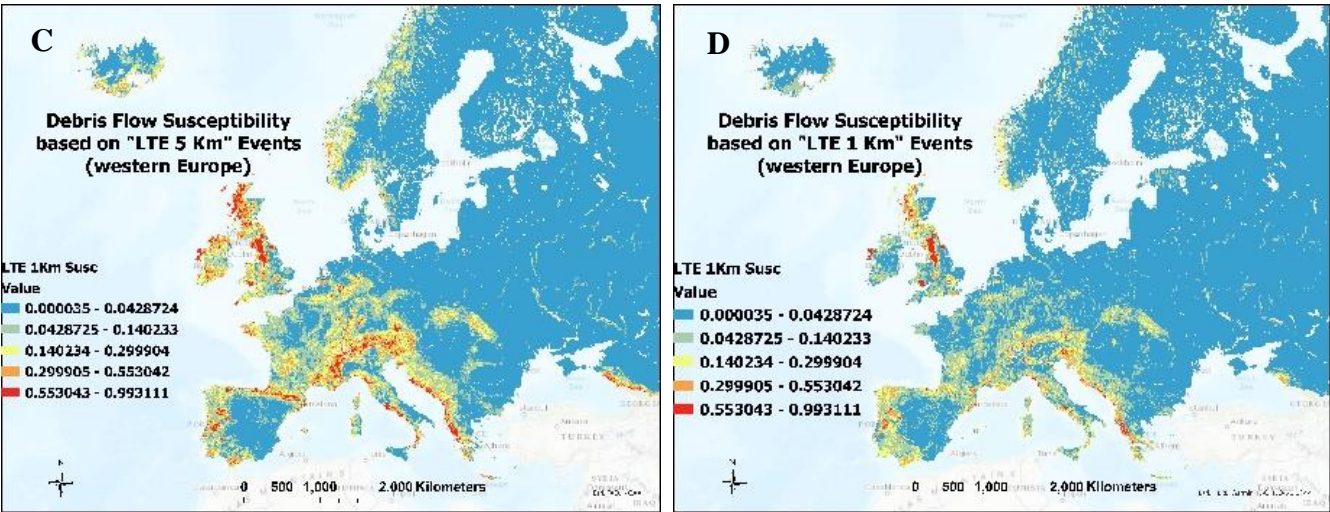



Figure 5. Debris flow susceptibility maps. A. "Original" B. "93 Random" C. "LTE 5Km" D. "LTE 1Km". Warmer colors represent higher susceptibilities. Base maps are from ArcGIS®, the intellectual property of Esri, used herein under license. Copyright © Esri.

235 There is close agreement in AUC and the most significant factor contributions (precipitation, fault density, soil type) in the first three models ("Original", "93 Random", and "LTE 5 Km'), Table 8. "LTE 1 Km" and "Exact" are significantly different than the others with respect to both the AUC and factors with highest percent contribution. 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

240 of such events (five) and their localized distribution, Fig. 6, hence the "Exact" model is not further considered.

245

250

Variable/Model	"Original" Percent contribution (99 events)	"93 Random" Percent contribution (93 events)	"LTE 5 Km" Percent contribution (55 events)	"LTE 1 Km" Percent contribution (24 events)	"EXACT" Percent 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 % clay	0.1	0.2	0.1	1.5	0.2
depth to bedrock	0	0.1	0.1	0.4	0
aridity	0.5	0.3	0.1	0.1	0

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

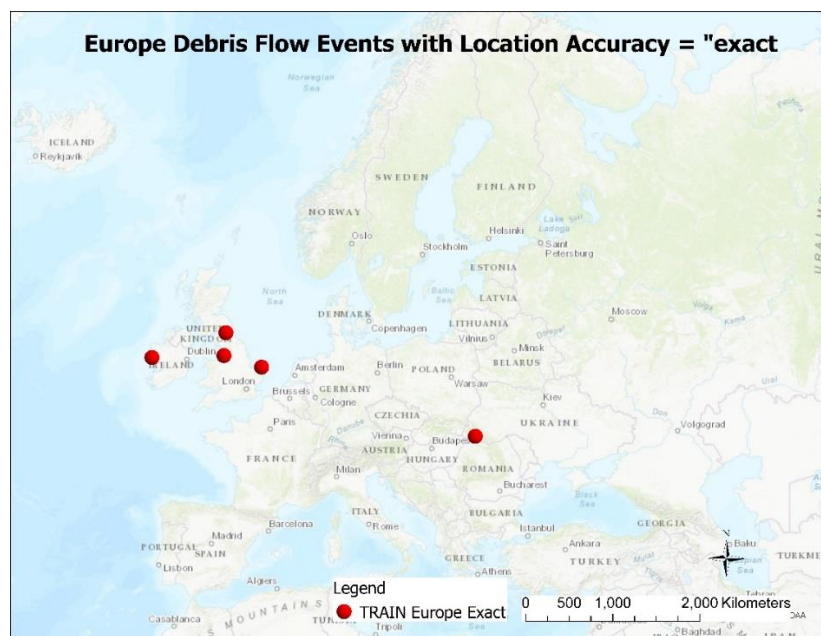



Figure 6. Distribution of the five European debris flow events with a location accuracy of "exact".
Base map is from ArcGIS®, the intellectual property of Esri, used herein under license. Copyright © Esri.

255

For precipitation, soil type, soil thickness, topsoil % 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
260 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.

Factor /Model	"Original"	"93 Random"	"LTE 5 km"	"LTE 1 Km"
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 %clay	*5-18*	*20*	*20*	*0-22*
depth to bedrock (cm)	0	114	0	0-2000
aridity (dimensionless index)	~1300 (Arid)	~1300 (Arid)	>= 60000 (Humid)	~15000 (Humid)

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

265

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", Fig. 7A, 7B, and 7C, respectively. The results of each map are displayed with a ten-break Natural Break (Jenks) classification.

270

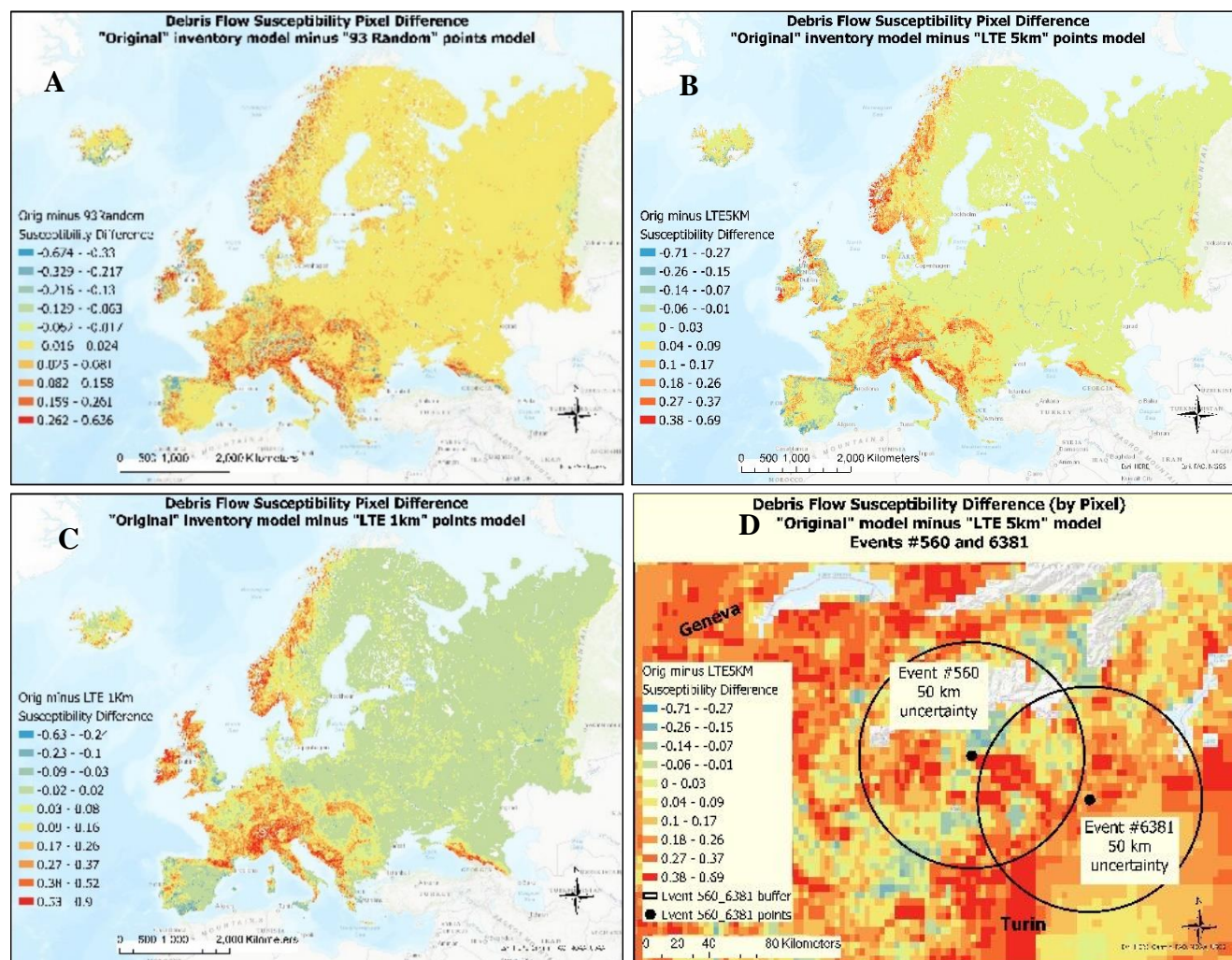



Figure 7. Model susceptibility differences.   A. Pixel difference between "Original" and "93 Random" models. B. Pixel difference between "Original" and "LTE 5Km" models. C. Pixel difference between "Original" and "LTE 1Km" models. Warmer colors represent areas of higher susceptibility values in "Original" 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

Determining the most relevant debris flow factors and factor classes are essential to producing reliable debris flow susceptibility maps, at any scale. However, due to event location uncertainties, noteworthy differences arise in factor and factor class determinations and susceptibility results among five MaxEnt models. Within their buffered areas, 35% of the

events had 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.

285

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 "high" susceptibility area (Fig. 5A and Table 10), five times larger than the "Exact" model, and the lowest AUC (0.891). "LTE 1Km (24 points with good distribution) has the smallest "high" susceptibility area (Fig. 5D and Table 10), with the same percentage coverage as the "Exact" model, and the

290   2nd 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 high susceptibility (Fig. 5C, Table 10), two times larger than the "Exact" model. The "93 Random" model has a slightly higher AUC (0.893) than the "Original" model, and an area of high susceptibility three times the size of the "Exact" model.

SUMMARY OF HIGHEST SUSCEPTIBILITY RANGE Susceptibility = 0.777479 - 0.97184 from 5 Natural Break (Jenks)				
Model	**# Events**	**#Pixels**	**%**	**AUC**
"Original"	99	8442	0.5	0.891
"93 Random"	93	4531	0.3	0.893
"LTE 5Km"	55	3550	0.2	0.896
"LTE 1Km"	24	1693	0.1	0.921
"Exact"	5	1700	0.1	0.930

295   **Table 10. Number of pixels and % area of susceptibility in the highest break of a five-break Natural Break (Jenks) categorization, for each of five models**

Due to the small number of points in the "Exact" model (five) and localized distribution, this model is not further considered
300   as a viable model.   As an enhanced view on susceptibility model choices, and with an attempt to better convey the 'uncertainties' to model users, mean and standard deviation maps were produced based on a combination of the two best models, "LTE 5Km" and "LTE 1Km", as input (Fig. 8). This result set provides the end user with a novel representation of uncertainty and 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.

305

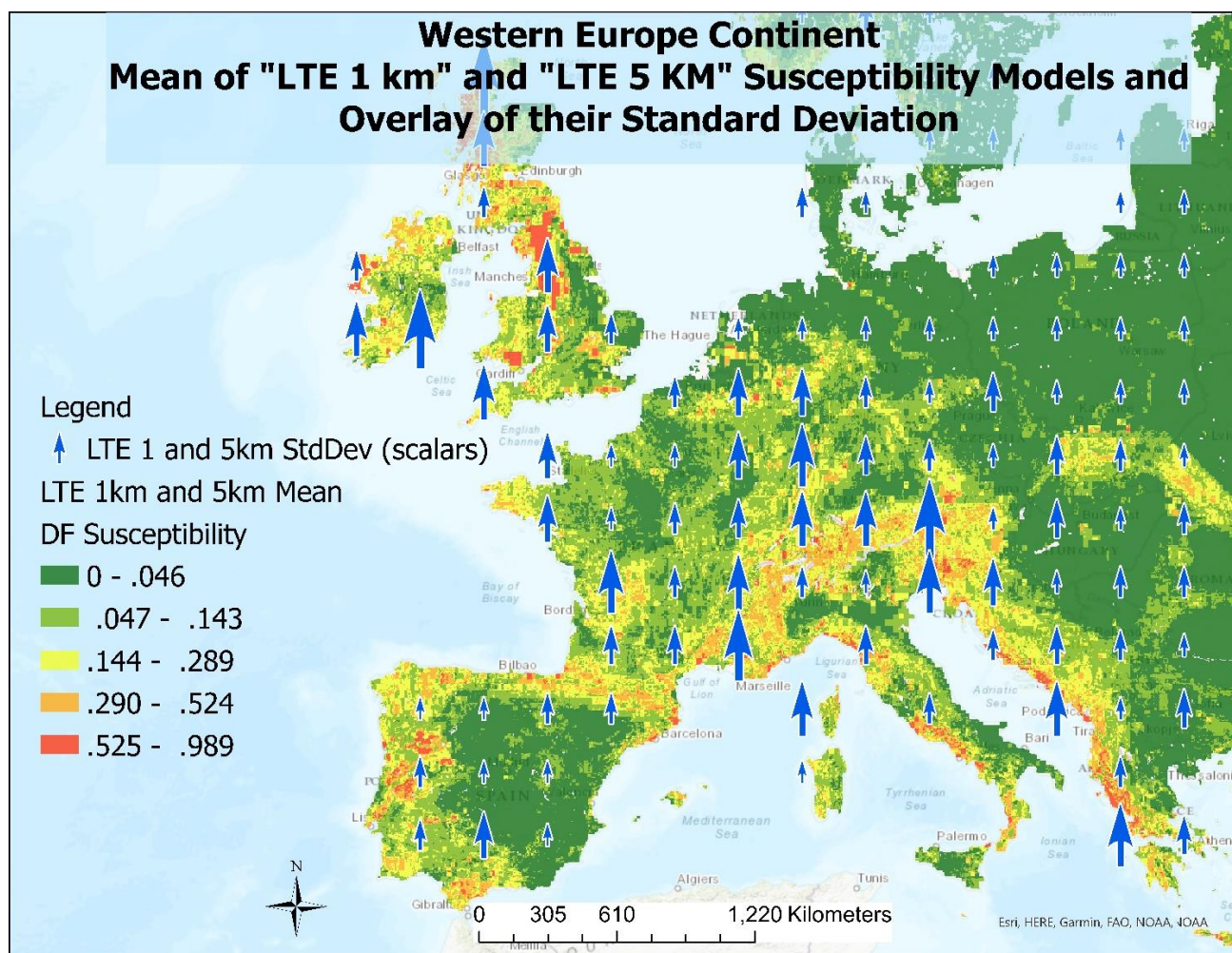



Figure 8. Overlay of susceptibility standard deviation and mean for two combined models ("LTE 5Km" and "LTE 1Km"). Standard deviation is represented by a scalar (blue arrows), small arrows = low values, etc. Warmers colors represent higher mean susceptibility Base map is from ArcGIS®, the intellectual property of Esri, used herein under license. Copyright © Esri.

6 Conclusion

88.8% of this study's Europe debris flow inventory has a locational uncertainty from 1 to 50 km, with 6% defined as "unknown". 82.5% of the complete inventory (all continents, all landslide types) have a locational uncertainty from 1 to 250 km, with 4.9% unknown (Table 11). It is believed these types of locational uncertainties may be representational of 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.

310

Locational Uncertainty Radius (km)	Frequency Distribution (Europe Debris Flows)	% of Total Debris Flow Events	Cum %	Frequency Distribution (World Debris Flows)	% of Total Debris Flow Events	Cum %	Frequency Distribution (World All Landslide Types)	% of Total Landslide Events	Cum %
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	544	4.9	100.0
<null>							2		
Total	99			2294			11033		

Table 11. Locational uncertainty frequency distribution of inventory: Europe debris flows (left), world debris flows (center), and world - all landslide types (right).

Identifying specific factor classes (e.g., Cambisol soil class) which have a dominant association with past debris flow events is essential in preparing debris flow susceptibility maps across an area of study (Meten et al., 2015). 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 resolution of the factor itself, the extent of location uncertainty of the debris flow events being used to train a susceptibility model, and the risk associated with an incorrect model (Carrara et al., 1992; Ardizzone et al., 2002).

In this study, five MaxEnt models were developed, each based on a level of location uncertainty in the training data. The predisposing factors (precipitation, fault density, soil class), with highest contribution to the susceptibility results of three of the five models, were investigated to identify and demonstrate the scope and impact of uncertainty on factor class determinations. First the variances in factor classes within an event's locational uncertainty buffer were compared to the values associated with the event point location to expose the problem. Then models were developed based on event location uncertainty and evaluated on AUC, and model comparisons of susceptibility mean and standard deviation across two of the best performing models. The purpose of a study, 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. In the context of this study, a debris flow susceptibility of the continent of Europe, where susceptible areas can be further investigated through a drill-down to a larger scale study, the error of commission may be preferable.

The result of this research primarily demonstrates the importance of event location accuracy and shows that there is not a single definitive solution to the uncertainty problem. Including all data uncertainties (in addition to locational) may help refine the results but is a much more comprehensive and complex model.

340

Utilizing the MaxEnt statistical results, a model can be chosen which may best meet the context, scope, and objectives of a European continent debris flow susceptibility study, given the inventory of historic datapoints. The "LTE 5Km" model provides a good intermediate result, utilizing fifty-five points, about half of the original data, fairly well distributed across the study area, includes data with a relatively small geographic locational uncertainty from 1 to 5 km, and includes the "exact"

345 known points. This model is an intermediate compromise slightly favoring error of commission over error of omission. Furthermore, by developing susceptibility maps of the mean and standard deviation derived from the two best models ("LTE 5km", "LTE 1km") provides additional tools for the end user to choose sites (areas of high susceptibility mean and low standard deviation) for further investigation.

350 The cause, nature, and handling of mapping errors (uncertainties) has been the subject of extensive research (Maffini et al., 1989; Openshaw, 1989; Hunter and Goodchild, 1996; Hunter, 1999; Wechsler, 1999; Ardizzone et al., 2002; Shi, 2010; Zufle et al., 2017). 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 (Veregin, 1989). 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

355 desirable (Soma, 2018). 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 (Kirschbaum et al., 2015), and the 'epistemic quality' and 'activity confidence' attributes of the

360 GEM Active Faults data (Styron and Pagani, 2020). 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.

Although this project is singularly focused on the impact of landslide inventory locational uncertainty, additional sources of data uncertainty may be inherent in other attributes of the inventory, as well as uncertainties associated both with the spatial

365 assignment and attributes of the environmental factors. The impact of the locational uncertainty will vary according to the scale of the study area and to the size of the locational uncertainty buffer vis-à-vis the scale of lateral change for a factor class. It would be ideal to create susceptibility maps with an attribute and graphical representation of certainty, but not practicable due to the array, complexity, and relationship of all uncertainties, without modeling across all these variables. According to Goodchild et al 1993 as cited in (Hunter et al., 1994) there are three options for handling uncertainty: (1) omit all reference to

370 it; (2) attach a descriptor to the output; and (3) show samples from the range of possible maps. The latter approach is 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.

Code and data availability

375 ArcGIS Pro, a proprietary software of Esri, and public domain MaxEnt software were utilized. All datasets acquired for use

in this study, are publicly accessible through various sources. Additional information about those datasets can be obtained

from L. J. Kurilla, laurie.kurilla@edu.unito.it.

Author contributions

LJK designed the project, conducted the data collection and preparation, analyses of results, model validations, and prepared

380 the draft manuscript. GF provided advice and feedback in the process.

Competing interests

The authors declare that they have no conflicts of interest.

Acknowledgements

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

385 and suggestions.

References

Ardizzone, F., Cardinali, M., Carrara, A., Guzzetti, F., and Reichenbach, P.: Impact of mapping errors on the reliability of landslide hazard maps, Natural Hazards and Earth System Sciences, 2, 3-14, DOI 10.5194/nhess-2-3-2002, 2002.

390 Brabb, E. E., Colgan, J. P., and Best, T. C.: Map showing inventory and regional susceptibility for Holocene debris flows and related fast-moving landslides in the conterminous United States, 10.3133/mf2329, 1999.

Bridges, E. M.: World Geomorphology, Cambridge University Press, Cambridge, 10.1017/cbo9781139170154, 2012.

Brighenti, R., Segalini, A., and Ferrero, A. M.: Debris flow hazard mitigation: A simplified analytical model for the

395 design of flexible barriers, Computers and Geotechnics, 54, 1-15, 10.1016/j.compgeo.2013.05.010, 2013.

Campbell, R. H.: Debris flows originating from soil slips during rainstorms in Southern California, Journal of Engineering Geology, 7, 339-349, 1974.

Carrara, A., Cardinali, M., and Guzzetti, F.: Uncertainty in assessing landslide hazard and risk, ITC Journal, 2, 172-183, 1992.

400 Chrisman, N.: Modeling error in overlaid categorical maps, in: The Accuracy of Spatial Databases, edited by: Goodchild, M., and Gopal, S., Taylor & Francis, Pennsylvania, 21-43, 1989.

Convertino, M., Troccoli, A., and Catani, F.: Detecting fingerprints of landslide drivers: A MaxEnt model, Journnal of Geophysical Research: Earth Surface, 118, 1367-1386, 10.1002/jgrf.20099,, 2013.

Costa, J. E.: Physical Geomorphology of Debris Flows, in: Developments and Applications of Geomorphology,
405 edited by: Costa, J. E., and Fleisher, P. J., Springer Berlin Heidelberg, Berlin, Heidelberg, 268-317, 10.1007/978-3-642-69759-3_9, 1984.

Devkota, K. C., Regmi, A. D., Pourghasemi, H. R., Yoshida, K., Pradhan, B., Ryu, I. C., Dhital, M. R., and 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, Natural Hazards, 65, 135-165,
410 10.1007/s11069-012-0347-6, 2012.

Dou, J., Tien Bui, D., Yunus, A. P., Jia, K., Song, X., Revhaug, I., Xia, H., and Zhu, Z.: Optimization of Causative Factors for Landslide Susceptibility Evaluation Using Remote Sensing and GIS Data in Parts of Niigata, Japan, PLoS One, 10, e0133262, 10.1371/journal.pone.0133262, 2015.

Dowling, C. A. and Santi, P. M.: Debris flows and their toll on human life: a global analysis of debris-flow fatalities
415 from 1950 to 2011, Natural Hazards, 71, 203-227, 10.1007/s11069-013-0907-4, 2013.

Esri, I.: ArcGIS Pro (Version 2.7), Esri, Inc. [code], 2020.

Froude, M. J. and Petley, D. N.: Global fatal landslide occurrence from 2004 to 2016, Natural Hazards and Earth System Sciences, 18, 2161-2181, 10.5194/nhess-18-2161-2018, 2018.

Froude, M. J. and Petley, D. N.: Guide to the Global Fatal Landslide Database on ArcGIS Online [dataset], 2019.
420 GÁL, A., Poszet, S. L., and Kerekes, A. H.: Landslide susceptibility assessment using the maximum entropy model in a sector of the Cluj–Napoca Municipality, Romania, Revista de Geomorfologie, 20, 130-146, 10.21094/rg.2018.039, 2018.

Grozavu, A. and Patriche, C. V.: Landslide Susceptibility Assessment: GIS Application to a Complex Mountainus Environment, 10.1007/978-3-642-12725-0_4, 2013.

425 Gunther, A., Reichenbach, P., Guzzetti, F., and Richter, A.: Criteria for the identification of landslide risk areas in Europe: the Tier 1 approach, Joint Research Center, Institute for Environment and Sustainability, European Commission, Luxembourg, 37-40, 10.2788/63147, 2007.

Highland, L. M. and Bobrowsky, P.: The Landslide Handbook - A Guide to Understanding Landslides, 2008.

Hunter, G. J.: Managing uncertainty in GIS, Geographical Information Systems, 2, 633-641, 1999.
430 Hunter, G. J. and Goodchild, M.: Communicating uncertainty in spatial databases, Transactions in GIS, 1, 13-24, doi:10.1111/j.1467-9671.1996.tb00030.x, 1996.

Hunter, G. J., Goodchild, M., and Robey, M.: A Toolbox for Assessing Uncertainty in Spatial Databases, 22nd Annual Conference of the Australasian Urban and Regional Information Systems Association, Inc., Sydney1994.

ISPRA: World Landslides Forum: Each year, landslide cause damage to 6 billion Euros, 2020.
435 Iverson, R. M.: The Physics of Debris Flows, American Geophysical Union, 35, 245-296, 1997.

Kirschbaum, D. and Stanley, T.: Satellite-Based Assessment of Rainfall-Triggered Landslide Hazard for Situational Awareness, Earth's Future, 6, 505-523, 10.1002/2017ef000715, 2018.

Kirschbaum, D., Stanley, T., and Zhou, Y. P.: Spatial and temporal analysis of a global landslide catalog, Geomorphology, 249, 4-15, 10.1016/j.geomorph.2015.03.016, 2015.

440 Kornejady, A., Ownegh, M., and Bahremand, A.: Landslide susceptibility assessment using maximum entropy model with two different data sampling methods, Catena, 152, 144-162, 10.1016/j.catena.2017.01.010, 2017.

Lombardo, L., Fubelli, G., Amato, G., and Bonasera, M.: Presence-only approach to assess landslide triggering-thickness susceptibility: a test for the Mili catchment (north-eastern Sicily, Italy), Natural Hazards: Journal of the International Society for the Prevention and Mitigation of Natural Hazards, 84, 565-588, 2016.

445 Lorente, A., García-Ruiz, J. M., Beguería, S., and Arnáez, J.: Factors Explaining the Spatial Distribution of Hillslope Debris Flows, Mountain Research and Development, 22, 32-39, 10.1659/0276-4741(2002)022[0032:Fetsdo]2.0.Co;2, 2002.

Maffini, G., Arno, M., and Bitterlich, W.: Observations and comments on the generation and treatment of error in digital GIS data, in: Accuracy of Spatial Databases, edited by: Goodchild, M., and Gopal, S., Taylor and Francis,
450 Philadelphia, 55-68, 1989.

Malamud, B. D., Turcotte, D. L., Guzzetti, F., and Reichenbach, P.: Landslide inventories and their statistical properties, Earth Surface Processes and Landforms, 29, 687-711, 10.1002/esp.1064, 2004.

Meten, M., PrakashBhandary, N., and 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,
455 2, 10.1186/s40677-015-0016-7, 2015.

Nettleton, I. M., Martin, S., Hencher, S., and Moore, R.: Debris Flow Types and Mechanisms, in: Scottish Road Network Landslide Study Scotland, 2005.

Nsengiyumva, J. B., Luo, G., Nahayo, L., Huang, X., and Cai, P.: Landslide Susceptibility Assessment Using Spatial Multi-Criteria Evaluation Model in Rwanda, Int J Environ Res Public Health, 15, 10.3390/ijerph15020243, 2018.

460 Openshaw, S.: Learning to live with errors in spatial databases, in: Accuracy of Spatial Databases, edited by: Goodchild, M., and Gopal, S., Taylor and Francis, Philadelphia, 263-276, 1989.

Park, N.-W.: Using maximum entropy modeling for landslide susceptibility mapping with multiple geoenvironmental data sets, Environmental Earth Sciences, 73, 937-949, 10.1007/s12665-014-3442-z, 2014.

Phillips, S. and Dudik, M.: Modeling of species distributions with Maxent: new extensions and a comprehensive
465 evaluation, Ecography, 31, 161-175, 10.1111/j.2007.0906-7590.05203.x, 2008.

Phillips, S., Dudik, M., and Schapire, R. E.: Maxent software for modeling species niches and distributions (Version 3.4.1) [code], 2021.

Ritchie, H. and Roser, M.: Urbanization, 2018.

Shi, W.: Principles of Modeling Uncertainties in Spatial Data and Spatial Analyses, CRC Press, Florida2010.

470 Soma, A.: Landslide susceptibility Map using Certainty Factor for Mitigation in Mountainous Area of Ujung-Loe Watershed South Sulawesi Indonesia, 2018.

Styron, R. and Pagani, M.: The GEM Global Active Faults Database, Earthquake Spectra, 36, 160-180, 10.1177/8755293020944182, 2020.

Toth, G., Montanarell, L., Stolbovoy, F., Bodis, K., Jones, A., Panagos, P., and Liedekerke, M.: Soils of the European
475 Union, European Commissiono Joint Research Centre Institute for Environment and Sustainability, Italy1018-5593, 10.2788/87029, 2008.

Veregin, H.: Error modeling for the map overlay operation, in: The Accuracy of Spatial Databases, edited by: Goodchild, M., and Gopal, S., Taylor & Francis, Pennsylvania, 3-18, 1989.

Wechsler, S. P.: Digital Elevation Model (DEM) Uncertainty: Evaluation and Effect on Topographic Parameters,
480 ESRI User Conference, San Diego, CA USA1999.

Yuan, S., Huang, G., Xiong, H., Gong, Q., Wang, J., and Chen, J.: Maximum Entropy-Based Model of High-Threat Landslide Disaster Distribution in Zhaoqing, China, Journal of Risk Analysis and Crisis Response, 7, 108-126, 2017.

Zufle, A., Trajcevski, G., Pfoser, D., Renz, M., Rice, M. T., Leslie, T., Delamater, P., and Emrich, T.: Handling Uncertainty in Geo-Spatial Data, 2017 IEEE 33rd International Conference on Data Engineering (ICDE),
485 10.1109/icde.2017.212, 2017.