

## AUTOMATIC SURFACE CLASSIFICATION FOR RETRIEVING AREAS WHICH ARE HIGHLY ENDANGERED BY EXTREME RAIN

Peter Fischer<sup>a</sup>\*, Thomas Krauß<sup>a</sup>, Dr. Tim Peters<sup>b</sup>

<sup>a</sup> Remote Sensing Technology Institute, German Aerospace Center (DLR), Münchener Str. 20, 82234 Wessling, Germany - (Peter.Fischer, Thomas.Krauss)@dlr.de

<sup>b</sup> Westfälische Provinzial Versicherung AG, Provinzial-Allee 1, 48181 Münster - Tim.Peters@provinzial.de

### Commission VII, WG VII/4

**KEY WORDS:** Automatic DSM classification, Terrain Positioning Index, extreme rain, damage probability

### ABSTRACT:

In this case study, an approach for finding regions endangered by extreme rain is presented. The approach is based on the assumption that sinks in the surface are more endangered than their surroundings. The surface data, which are the source for the classification, are generated using a Cartosat stereo scene. The classification is performed by using an algorithm for retrieving the terrain positioning index. Different classification schemes are possible, therefore a set of input parameters is iteratively computed. The classification results are then evaluated. For validating the classification stock data of an insurance are used. We compare the position of the reported damages caused by extreme rain with our classification. By doing so we got the confirmation of the assumption.

### 1. INTRODUCTION

Human made objects face in general several types of dangers, which are related to small scale weather events, like storms, volley and extreme rain. The orography of a location strongly determines the influence of an extreme weather event on the human entities placed at the location. Finding indexes for spatial regions based on the orography which indicate the probability of being affected by an extreme weather event is therefore essential. Such indexes would enable decision makers to develop strategies where structures like roads, bridges and settlements can be placed with an minimal risk of being damaged by extreme weather phenomena.

There has been a lot of work done in the field of automated terrain classification. A broad range of indexes and classification algorithms is already developed, which take the presence of discrete regions like watersheds (Tagil and Jenness, 2008), or regions which can be described as geomorphological enclosed areas (Reu et al., 2013). These indexes describe the surface as is. This serves as entry point to our application, where we investigate the influence of orography to the relationship between extreme weather events and the probability of damage to human entities located on earth surface. This is a quite complex task, due to the differing dependencies of spatial exposition and damage event for various weather events. For example a short distance to a river may be a good indicator for a high risk of floods, but for a storm event the influence is probably negligible. We want to show the impact of the spatial exposition of entities on the probability of being affected by extreme rain events.

Insurance companies already have zoning systems which classify the risk of locations being affected by extreme rain and flooding. These existing systems work quite well for floods. But analysis of damage and stock data have shown that for backwater and extreme rain existing zoning systems don't deliver reliable results. This can be explained by the methodology used for zoning the space. In general the systems take the euclidean distance to the

next streaming water body into account. By doing so, in case of an flooding event which affects this water body for every point on the surface an estimation can be made about the probability of being affected by this flooding.

Extreme rain events in contrast are not bounded to streaming water bodies. They occur at almost every place on earth surface, and the risk of a damage on the ground is mainly related to the exposition of a body on the ground. In this paper we prove the assumption that sinks on the surface are higher endangered than hilltops. This assumption is based on the nature of fluids to follow the gradient of surfaces to minimize their potential.

### 2. DATA BASE

#### 2.1 Digital Elevation Model

As source data for the classification algorithm a digital elevation model retrieved from an Cartosat stereo scene is used, which is shown in figure 1. The parameters describing the spatial extent, resolution and the projection of the DEM are listed in table 1.

<b>Datum</b>	WGS 84 / UTM zone 32 N	
<b>Spatial resolution [m]</b>	5	
<b>Min/Max height [m]</b>	31.89, 372.23	
<b>Corner</b>	<b>Easting [m]</b>	<b>Northing [m]</b>
<b>Upper Left</b>	348100.00	5725430.00
<b>Lower Left</b>	348100.00	5687255.00
<b>Upper Right</b>	386055.00	5725430.00
<b>Lower Right</b>	386055.00	5687255.00

Table 1: Description of DEM

For determining the influence of the spatial resolution on the classification process and the final geo-product, also downsampled versions of this DEM were used. For downsampling bilinear interpolation (algorithm from Scipy (Jones et al., 2001--)) with a downsampling factor of two and four were used.

\*Corresponding author. Tel.: +49 8153 281561; fax: +49 8153 281444.

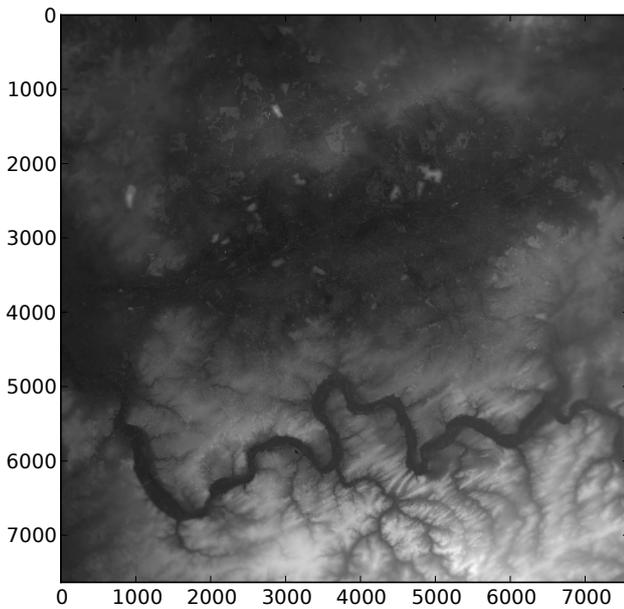


Figure 1: Digital Elevation Model

Furthermore we took an digital terrain model aquired by the Shuttle Radar Topography Mission (Farr et al., 2007), to compare the influence of spatial resolution but moreover the influence of different acquisition systems. By doing so, we can determine whether surface models retrieved by radar systems deliver at least comparable results to models retrieved from optical systems.

## 2.2 Geocoded insurance data

The data provided by the insurance are divided into stock data and damage event data, where only damage events caused by extreme rain were taken into account. The stock data contain information about all households who own an elementary insurance for extreme rain, the damage dataset lists households who were affected by an damage event in the time span from 2003 to 2013.

The stock data table has five attributes, which are described in table 2. Each dataset represents a household which is specified by an unique insurance number, 3D coordinates (UTM), the insured sum and auxiliary data. The stock data table contains 24997 households which own an elementary insurance.

attribute	type	meaning
VSNR	int	unique insurance id, primary key
x	double	easting [m]
y	double	northing [m]
z	double	height [m]
vers_summe	int	insurance sum [€]

Table 2: Description of Stock Data attributes

The damage data table has six attributes, three for the location of the damage, one for the contract id, one for the costs of the damage and one for the date when a damage occurred. Table 3 gives an overview of the attributes. The number of households which were affected by a damage is 735.

Figure 2 depict the distribution of the households in the test area. Cyan dots mark households which own an insurance, red dots mark the households which were affected by an damage event.

attribute	type	meaning
SCHDAT	date	date of damage
VSNR	int	unique insurance id, primary key
auf	String	costs ('aufwand') [€]
x	double	easting [m]
y	double	northing [m]
z	double	height [m]

Table 3: Description of Damage Data attributes

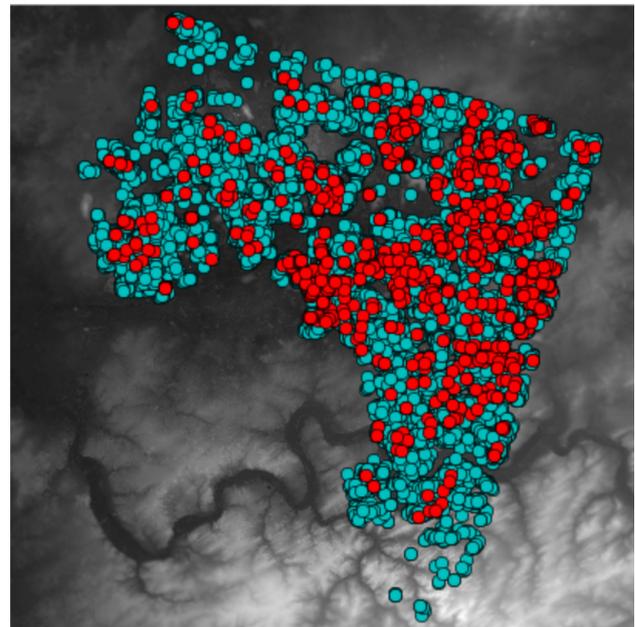


Figure 2: Distribution of households in test areal

## 3. TERRAIN CLASSIFICATION ALGORITHM

### 3.1 Theoretical Background

The TPI (1) is the basis of a classification system and is simply the difference between a cell elevation value  $z_0$  and the average elevation  $\bar{z}$  of the neighbourhood  $R$  around the cell (Weiss, 2001) (Wilson and Gallant, 2000).

$$TPI = z_0 - \bar{z} \quad (1)$$

$$\bar{z} = \frac{1}{n_R} \sum_{i \in R} z_i \quad (2)$$

There are two situations for a kernel map moving over a surface, which can be distinguished easily. On the one side the height value at a point  $pt$  is greater than the mean elevation  $\mu$  in the neighbourhood of this point. Therefore the TPI value is also positive. On the other side the mean elevation  $\mu$  is greater than the elevation at point  $pt$ , because of this the TPI value is lower than zero.

To distinguish between flat areas and points on slopes, the slope angles, described in (Zevenbergen and Thorne, 1987) are also taken into account. Such situations occur for example at hillside situations. For computing the slope angles, the following equations 3 and 4 are used.

$$s_{x,y} = \sqrt{\left(\frac{\Delta z}{\Delta x}\right)^2 + \left(\frac{\Delta z}{\Delta y}\right)^2} \quad (3)$$

$$s_{x,y}^{\circ} = \frac{\arctan(s_{x,y})}{\pi * 180} \quad (4)$$

In the project the DEV index is used which is based on the TPI. DEV (5) measures the topographic position as a fraction of local relief normalised to local surface roughness (Reu et al., 2013), the equations are given in 5 and 6.

$$DEV = \frac{z_0 - \bar{z}}{SD} \quad (5)$$

$$SD = \sqrt{\frac{1}{n_R - 1} \sum_{i=1} (z_i - \bar{z})^2} \quad (6)$$

The main influencing factor of the TPI and DEV value are the two parameters inner radius and outer radius, which define the resulting kernel map. In figure 3 two different kernel maps are shown. For both kernel maps, the outer radius is 10 pixels. For figure 3(a), the inner radius is 5 pixel, for figure 3(b) the inner radius is zero pixels. Thus the resulting number of height values taken into account is for the small kernel 252 in contrast to 348 height values for the big kernel map.

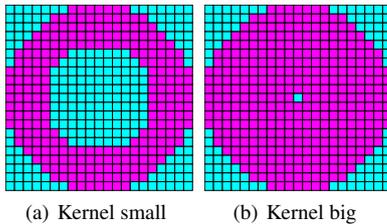


Figure 3: Comparison of different kernel sizes

Depending on the kernel size the classified result can diver quite much. If a small kernel map is used for classifying a DSM, small landforms can be distinguished in the resulting classification. If the same DSM is classified using a bigger kernel map, which is done by using a bigger outer radius, then the small landforms become generalized. The resulting output image therefore gives more small scale landforms. Finding the right kernel size is an iterative process, depending on the application (Tagil and Jenness, 2008).

The classification schema itself remains the same for different kernel maps, in most cases the one given in table 4 is used. This classification schema gives 6 classes, as input a DEV image and a slope image is needed. It is also possible to combine different kernel maps and therefor to gain more sophisticated classes.

Naming	class	DEV	Slope
valleys, base	1	$x \leq -1$	
lower slope	2	$-1 < x < 0.5$	
flat	3	$-0.5 < x < 0.5$	$< 5^{\circ}$
mid slope	4	$-0.5 < x < 0.5$	$> 5^{\circ}$
upper slope	5	$0.5 \leq x \leq 1$	
ridge, hilltop	6	$x > 1$	

Table 4: Classification rules - 6 classes, adapted from (Jenness, 2006)

Subsequent an overview of different classifications is shown. The goal is to show the influence of different inner and outer radii on the resulting classification. For outer and inner radius, the three tupels [300,100] m, [600, 200] m and [900, 300] m are used. Like already in section 2.1 mentioned, we also want to analyse the influence of different spatial resolutions of the input DSM on the

classification results. Grohmann (Grohmann et al., 2009) already showed that for surface roughness the influence of the spatial resolution can be neglected to some extent. To get a first clue, the source DSM with five metres resolution and a downsampled version with 20 metres resolution is used. Figure 4 shows the classification results based on the original DSM with five metres spatial resolution. In figure 4(a) as algorithm parameters the radii tuple [300, 100] m is used, this means the kernel is specified with an outer radius of 300 metres and an inner radius of 100 metres. With increasing radii in figure 4(b) and figure 4(c) the scale of the landforms classified decreases. Especially the river network gets very general in figure 4(c). The colour legend is given in table 4(g).

The same classifications were done on the downsampled DSM with an spatial resolution of 20 metres, the results are depicted in figure 4(d), 4(e) and 4(f) respectively. Compared to the classification results gained with the high resolution DSM, almost no difference can be detected visually. This supports the observations done by (Grohmann et al., 2009).

### 3.2 Proposed Algorithm

This classification algorithm is implemented as XDibias (Triendl et al., 1982) modul using Python (G. van Rossum and F.L. Drake, 2001) and Cython (Behnel et al., 2011) programming language. The idea of the algorithm consists of three steps.

1. The classification algorithm is applied to the three source DSMs with different combinations for the two parameters specifying the kernel map, inner and outer radius.
2. The classified raster dataset is merged with the point data from the insurance, for each single point in the two datasets the corresponding class is stored. Therefore a script based on the PointSampling Tool from Quantum GIS (QGIS Development Team, 2009) is implemented.
3. An grading matrix is computed to compare the classification results. The hypothesis is that households in valleys are more probable affected by damages than households on hilltops. This means the probability of being affected by an damage based on extreme rain should be higher for households in valleys. The structure of such a grading matrix is described in detail in the following.

The grading matrix, shown in equation 7, consists of the results from the mentioned three steps. There are 6 rows, which correspond to the total number of land form classes. The first column gives the name of the class  $a_x$ , for example  $a_1$  is valley. Second and third column give the total distribution of households  $b_x$  with contract and of damaged households  $c_x$ . Column four and five give the total costs for each class  $d_x$  and the total insured sum for each class  $e_x$  in euro. Column six gives the claims rate  $f_x$  for each class, which is computed with equation 8. Column seven gives the number of damaged households  $g_x$  in percent per class, which is computed with equation 9.

$$M = \begin{pmatrix} a_1 & b_1 & c_1 & d_1 & e_1 & f_1 & g_1 \\ a_2 & b_2 & c_2 & d_2 & e_2 & f_2 & g_2 \\ a_3 & b_3 & c_3 & d_3 & e_3 & f_3 & g_3 \\ a_4 & b_4 & c_4 & d_4 & e_4 & f_4 & g_4 \\ a_5 & b_5 & c_5 & d_5 & e_5 & f_5 & g_5 \\ a_6 & b_6 & c_6 & d_6 & e_6 & f_6 & g_6 \end{pmatrix} \quad (7)$$

$$f_x = \frac{d_x}{e_x} * 1000 \quad (8)$$

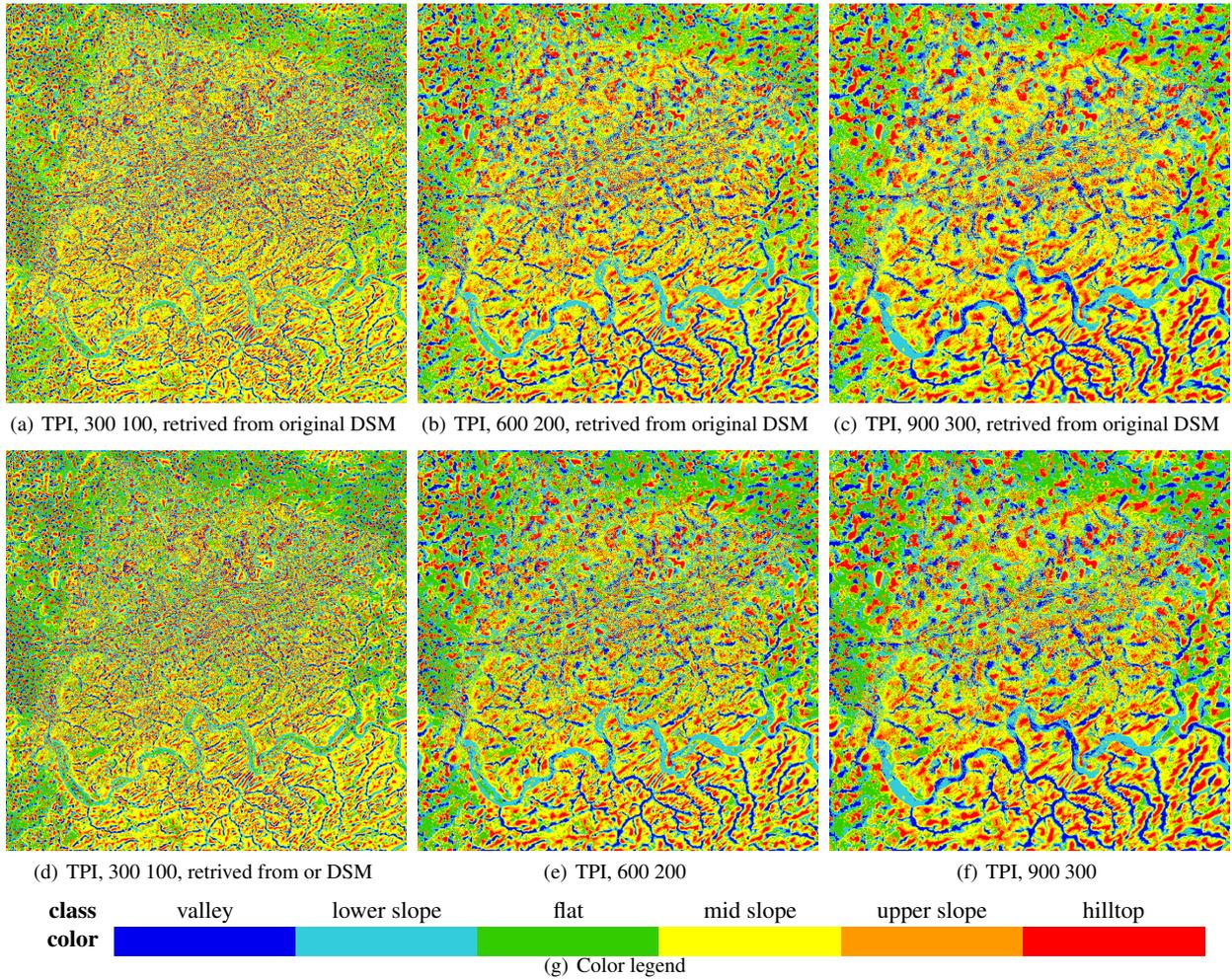


Figure 4: Comparison classification results for different kernel masks and spatial resolutions of DSM

$$g[\%] = \frac{c_x}{b_x} * 100 \quad (9)$$

4. The variance in the percentage of affected households over the six classes, see equation 13.

#### 4. RESULTS AND DISCUSSION

Four DSMs were processed, having a spacial resolution of 5,10 and 20 and 66 meters respectively. The parameter for the inner kernel radius ranges from 50 to 700 meters with a step length of 50 meters, the parameter for the outer radius ranges from 100 to 900 meters, step length is also 50 meters. This leads to 441 classified images for each of the four DSMs, all of them linked to a grading matrix.

The main objective is to answer the question, whether households in depressions are more affected by extreme rain, and if yes, which kernel radii fit best for the land form classification. To answer both questions, four descriptors are computed for the different surface models and the different radii combinations, and then stored in the four matrixes  $F, G, H, I$ .

1. The difference between class one and six for the claim rate, computed with equation 10.
2. The difference between the percentage of affected households for class one and six, computed with equation 11.
3. The variance in the claim rates over the six classes, see equation 12.

$$F_{i,j} = \sum_{i=1}^{15} \sum_{j=2}^{18} f_{i*50,j*50,6} - f_{i*50,j*50,1} \quad (10)$$

$$G_{i,j} = \sum_{i=1}^{15} \sum_{j=2}^{18} g_{i*50,j*50,6} - g_{i*50,j*50,1} \quad (11)$$

$$H_{i,j} = \sum_{i=1}^{15} \sum_{j=2}^{18} \frac{\sum_{k=1}^6 (f_{i*50,j*50,k} - \bar{f})^2}{6} \quad (12)$$

$$I_{i,j} = \sum_{i=1}^{15} \sum_{j=2}^{18} \frac{\sum_{k=1}^6 (g_{i*50,j*50,k} - \bar{g})^2}{6} \quad (13)$$

Figure 5 gives an overview for the range of the difference between the claim rate in class one and class 6. The maximum difference in the claim rate is achieved using the downsampled surface model with 10 meters spatial resolution, which is shown in figure 5(b). Both the original surface model and the downsampled with 20 meter spatial resolution don't reach such a high

difference. The claim rate differences resulting from the processing with SRTM data are smallest, not reaching values higher than 7. In Figure 6 a detailed view is given for the variance in the claim rates. It is shown that the variance reaches also its maximum values when using the Cartosat data downsampled with factor two.

The same evaluation was made for the damage probabilities, figure 7 depicts the difference between the number of affected households per class [%] in class one and class six. Figure 8 shows the variance in the single number of affected households per class [%].

We see that the observation from the claim rates is somehow mirrored into this plots. Also for the probabilities a big outer radius and a small inner radius seems to be a good advice. And also here, the influence of the spatial resolution of the input DSM seems almost negligible for the Cartosat DSMs. Probably the point that the downsampled DSM with 10 m spatial resolution delivers the best results can be explained when taking into account, that vegetation and other error sources for land form classification are filtered by the downsampling. On the other hand taking the radar data from SRTM does not deliver that good results, probably because of the lower spatial resolution. Furthermore when it comes to find the maximum difference for the number of affected households between hilltops and valley, a small outer radius of 450 m and an inner radius of 50 m seems to deliver also reliable results.

Exemplary we compare the classification result retrieved with using 900 meters as outer radius and 150 meters as inner radius for optical data and 450 meters as outer radius and 50 meters as inner radius for SRTM data. These combinations give us the most promising results in the previous investigations. Figure 7(a) shows the decrease of the percentage of affected households from the valley to the hilltop. This trend is shown, no matter whether the classification is based on optical or radar data. In the claim rates this trend is depicted even stronger. This means that not just more households in valleys are affected by extreme rain than on hilltops, but also that the costs of the single damages are in general higher in valleys.

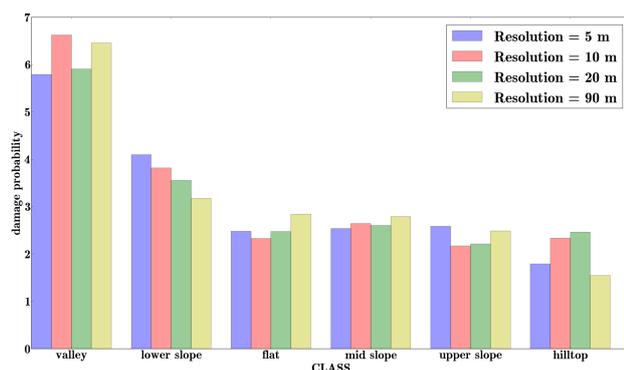


Figure 9: Percentage of affected households per class

## 5. CONCLUSIONS AND FUTURE WORK

By combining the outcomes from claim rate differences and claim rate variance, we come to the conclusion that big outer radii in general suite best for reaching good classification results. The inner radius in contrast should be quite small, not bigger than 300 meters. An outer radius of 900 meters and an inner radius of 150 meters seem to deliver the most satisfying results for high resolution optical data. The influence of the spatial resolution is at least not that big - this can be explained with keeping in mind that the

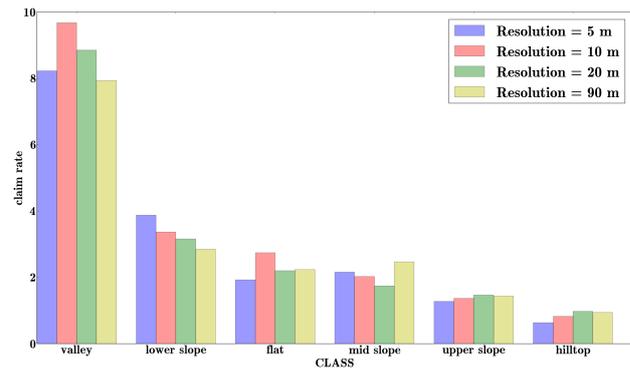


Figure 10: Claims Rate for different resolutions

lower resolution DSMs are retrieved by downsampling the high resolution DSM using linear interpolation.

For low resolution radar data the inner radius was taken as low as possible, the outer radius was not bigger than 450 m. More investigations should be done on the influence of the different data record methods, here radar and optical systems. For example the degree of soil moisture and vegetation cover at record time have different influences on the resulting surface model - which influences the land form classification.

The results showed that the spatial exposition of houses has great on the claim rate and probability of the occurrence of a damage caused by extreme rain. By using more sophisticated algorithms for land form classification and taking more spatial descriptors like curvature and land cover into account, probably the model could be even more accurate.

## ACKNOWLEDGEMENTS

This project is funded by the department of Technology Marketing of German Aerospace Center, and carried out in a strategic innovation partnership with Westfälische Provinzial Versicherung Aktiengesellschaft, Germany. The authors gratefully acknowledge the support and assistance of Robert Klarner, DLR TM RBO OP. Furthermore the authors would like to thank Frank Bistrick and Roman Kolbe from Westfälische Provinzial Versicherung for the initial stock and damage data and discussions.

## REFERENCES

- Behnel, S., Bradshaw, R., Citro, C., Dalcin, L., Seljebotn, D. and Smith, K., 2011. Cython: The best of both worlds. *Computing in Science Engineering* 13(2), pp. 31–39.
- Farr, T., Rosen, P., Caro, E., Crippen, R., Riley, D., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L., Seal, D., Shaffer, S., Shimada, J., Umland, J., Werenr, M., Oskin, M., Burbank, D. and Alsdorf, D., 2007. The shuttle radar topography mission. *Reviews of Geophysics*.
- G. van Rossum and F.L. Drake, 2001. *Python Reference Manual*. PythonLabs, Virginia, USA.
- Grohmann, C., Smith, M. and Riccomini, C., 2009. Surface roughness of topography: A multi-scale analysis of landform elements in midland valley, scotland. *Proceedings of Geomorphometry 2009* pp. 140–148.
- Jenness, J., 2006. Topographic position index (tpi\_jen.avx) extension for arcview 3.x, v.1.3.a. Jenness Enterprise.

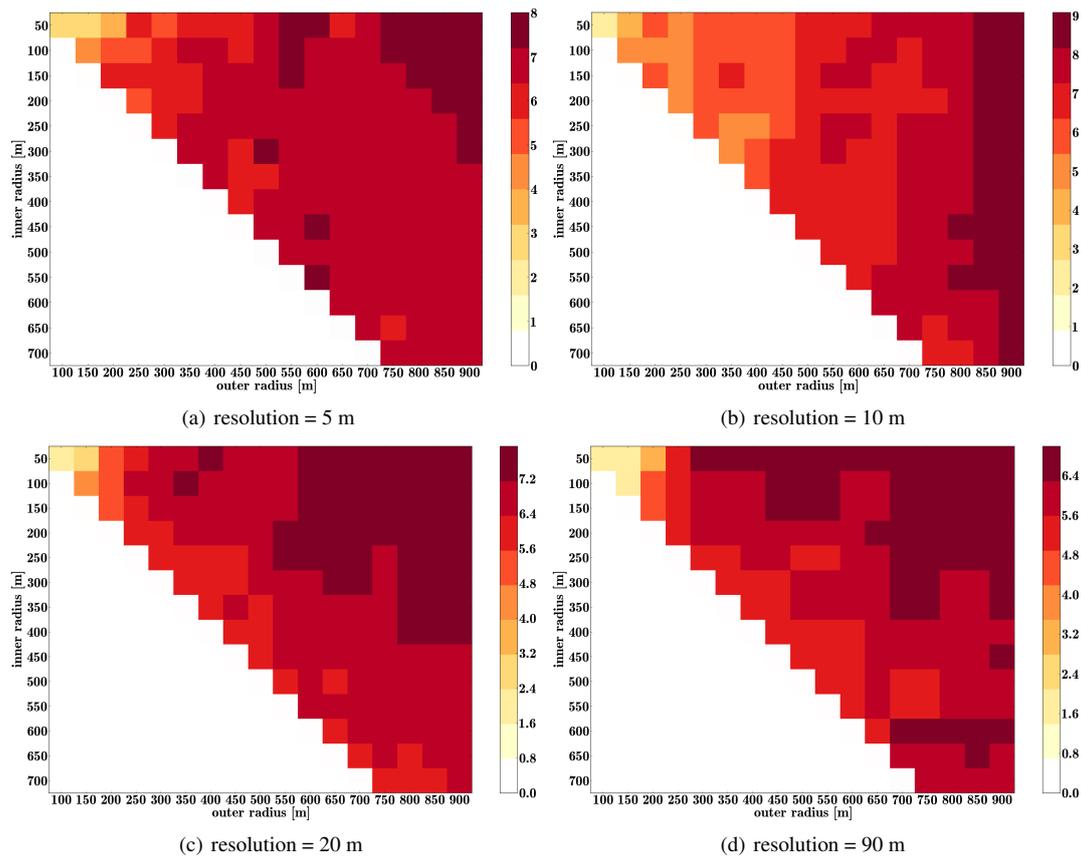


Figure 5: Claim Rate Difference Class 1 - Class 6 for different resolutions

Jones, E., Oliphant, T., Peterson, P. et al., 2001–. SciPy: Open source scientific tools for Python. [Online; accessed 2014-07-24].

QGIS Development Team, 2009. QGIS Geographic Information System. Open Source Geospatial Foundation.

Reu, J. D., Bourgeois, J., Bats, M., Zwertvaegher, A., Gelorini, V., Smedt, P. D., Chu, W., Antrop, M., Maeyer, P. D., Finke, P., Meirvenne, M. V., Verniers, J. and Cromb, P., 2013. Application of the topographic position index to heterogeneous landscapes. *Geomorphology* 186(0), pp. 39 – 49.

Tagil, S. and Jenness, J., 2008. Gis-based automated landform classification and topographic, landcover and geologic attributes of landforms around the yazoren polje, turkey. *Journal of Applied Sciences* 8(6), pp. 910 – 921.

Triendl, E., Lehner, M., Fiedler, R., Helbig, H., Kritikos, G. and Kbler, D., 1982. *Dibias handbook*, technical report. DFVLR Oberpfaffenhofen.

Weiss, A.-D., 2001. Topographic position and landforms analysis. Poster Presentation, ESRI Users Conference, San Diego, CA.

Wilson, J. and Gallant, J., 2000. *Terrain Analysis: Principles and Applications*. Earth sciences: Geography, Wiley.

Zevenbergen, L. and Thorne, C., 1987. Quantitative analysis of land surface topography. *Earth Surface Processes and Landforms* 12(2), pp. 47–56.

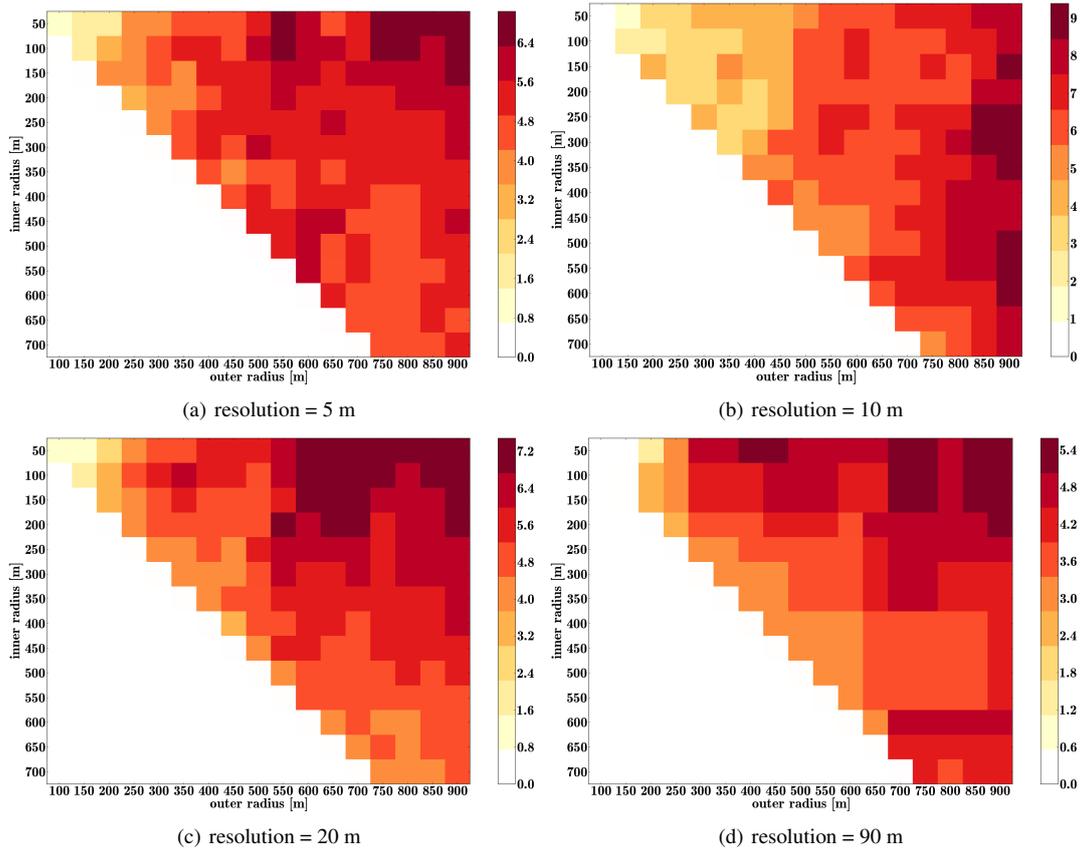


Figure 6: Variance in Claim Rates for different resolutions

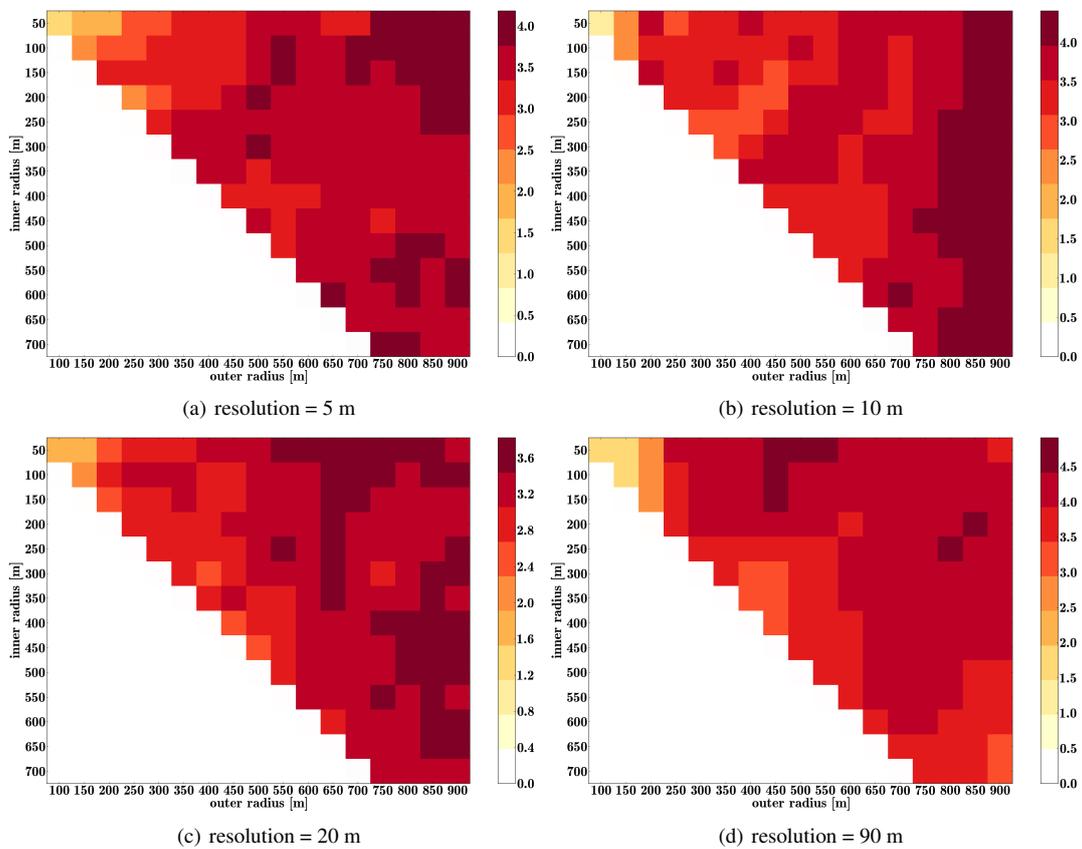


Figure 7: affected households [%] Class 1 - Class 6 for different resolutions

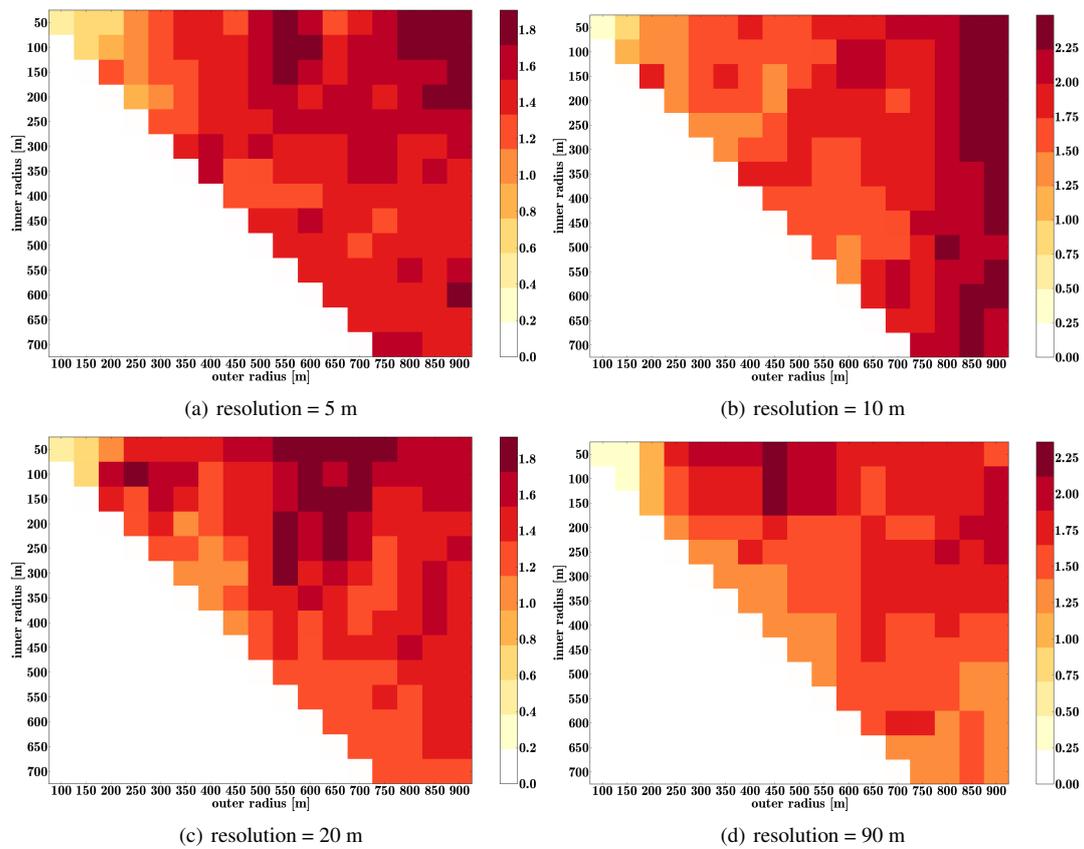


Figure 8: Variance in number of affected households per class [%] for different resolutions