



## Probabilistic disaggregation model with application to natural hazard risk assessment of portfolios

Custer, Rocco; Nishijima, Kazuyoshi

*Publication date:*  
2012

[Link back to DTU Orbit](#)

*Citation (APA):*

Custer, R., & Nishijima, K. (2012). Probabilistic disaggregation model with application to natural hazard risk assessment of portfolios. Paper presented at Fifth Asian-Pacific Symposium on Structural Reliability and its Applications , Singapore, Singapore.

---

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

# PROBABILISTIC DISAGGREGATION MODEL WITH APPLICATION TO NATURAL HAZARD RISK ASSESSMENT OF PORTFOLIOS

ROCCO CUSTER<sup>1</sup> and KAZUYOSHI NISHIJIMA<sup>2</sup>

*Department of Civil Engineering, Technical University of Denmark,  
Brovej Building 118, 2800 Lyngby, Denmark,*

*<sup>1</sup>E-mail: rocu@byg.dtu.dk <sup>2</sup>E-mail: kazn@byg.dtu.dk*

In natural hazard risk assessment, a resolution mismatch between hazard data and aggregated exposure data is often observed. A possible solution to this issue is the disaggregation of exposure data to match the spatial resolution of hazard data. Disaggregation models available in literature are usually deterministic and make use of auxiliary indicator, such as land cover, to spatially distribute exposures. As the dependence between auxiliary indicator and disaggregated number of exposures is generally imperfect, uncertainty arises in disaggregation. This paper therefore proposes a probabilistic disaggregation model that considers the uncertainty in the disaggregation, taking basis in the scaled Dirichlet distribution. The proposed probabilistic disaggregation model is applied to a portfolio of residential buildings in the Canton Bern, Switzerland, subject to flood risk. Thereby, the model is verified and the relevance to natural hazard risk assessment is illustrated.

*Keywords:* Disaggregation, probabilistic model, Dirichlet distribution, compositional data, indicator, insurance portfolio, natural hazard, flood risk

## 1. Introduction

Natural hazard risk models are widely utilized to assess the occurrence of natural hazard events, such as floods, earthquakes and tropical cyclones, and their impact on the built environment and more generally on societal activities.

Natural hazards often affect entire portfolios of buildings and structures. Thus for many applications it is not sufficient to consider buildings and structures individually. For instance, insurance and reinsurance companies need risk assessments for their entire portfolios to assure the company stays solvent in case of large events, see Kron (2005); governments faced with decisions regarding risk-reducing measures need an evaluation of the measures' impact on every structure that is potentially affected.

The level of sophistication of natural hazard models has been increasing over the last few decades and it has become possible to simulate hazard at a high spatial resolution. As a consequence the hazard information is often available at a higher spatial resolution than the exposure information, i.e. information on the elements at risk, which often comes spatially

aggregated. A common remedy to this mismatch is the disaggregation of exposure to a spatial scale matching the hazard resolution in the natural hazard model.

Wünsch, Herrmann et al. (2009), describes disaggregation as the transfer of the value of a variable from a coarse spatial level to a lower spatial level by means of auxiliary indicator.

Auxiliary indicator can be any variable available at a lower spatial level and correlates with the disaggregated variable; common examples of auxiliary data are land cover data, or census data. Several disaggregation methods have been developed and are available in literature, e.g. Thieken (2006), Wünsch, Herrmann et al. (2009) and Gallego (2010), which are deterministic.

However, the dependency between disaggregated variable and auxiliary variable is generally not perfect. Therefore, it is advisable to consider the disaggregation uncertainty through a probabilistic approach to disaggregation. No disaggregation model with a probabilistic approach seems available in literature.

The objective of this paper is to develop a probabilistic disaggregation model and to show its relevance in the context of natural hazard risk assessment. Whereas deterministic disaggregation models only consider one possible geographical distribution of exposures, a probabilistic disaggregation model accommodates uncertainty on spatial distribution of the exposures. This allows for assessing tail-risks of aggregated portfolios facing natural hazard more precisely.

The paper is organized as follows. First, a probabilistic disaggregation model is proposed, then, its application is illustrated at the example of a flood event in Switzerland. Finally, the results are discussed and an outlook on further developments is given.

## 2. Probabilistic Disaggregation Model

### 2.1. Definitions

This paper proposes a probabilistic disaggregation model where  $x$  denotes the geographically aggregated variable to be disaggregated onto a study area with  $n$  cells; the quantity of  $x$  attributed to the  $i$  th cell is a random variable and is denoted by  $Y_i, i = 1, 2, \dots, n$ . By definition of disaggregation,  $Y_i, i = 1, 2, \dots, n$  are subject to the following constraint:  $\sum_{i=1}^n Y_i = x$ , therefore  $Y_i, i = 1, 2, \dots, n$  have the characteristics of compositional data, see e.g. Aitchison (1986).

An indicator  $z_i$  is assumed to be given for each cell.  $z_i$  can be any auxiliary indicator which shows a certain dependency with  $Y_i$ ; the vector  $\mathbf{z} = (z_1, z_2, \dots, z_n)$  comprises the indicators for all cells in the study area.

### 2.2. Objective

The objective of the probabilistic disaggregation model is to describe the characteristics of uncertainty of  $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)$  for given  $\mathbf{z}$  and  $x$  through their joint probability density function, i.e.,  $f_{Y_1, Y_2, \dots, Y_n | x, \mathbf{z}}(y_1, y_2, \dots, y_n | x, \mathbf{z})$ .

### 2.3. Model

The proposed model is based on the Dirichlet distribution; a multivariate probability distribution commonly used to model compositional data. The  $n$ -dimensional random vector of the Dirichlet distribution is denoted  $\Theta = (\Theta_1, \Theta_2, \dots, \Theta_n)$  where  $\Theta_i$  is a random variable representing the proportion of  $x$  attributed to cell  $i$ , i.e.  $\Theta_i = Y_i/x$ . The Dirichlet distribution is parameterized through vector  $\mathbf{a} = (\alpha_1, \alpha_2, \dots, \alpha_n)$ .

The joint probability density function of  $\Theta$  is given in Eq. (1).

$$f_{\Theta|\mathbf{a}}(\boldsymbol{\theta}|\mathbf{a}) = (B(\mathbf{a}))^{-1} \prod_{i=1}^n \theta_i^{\alpha_i-1}, \quad B(\mathbf{a}) = \prod_{i=1}^n \Gamma(\alpha_i) \left( \Gamma\left(\sum_{i=1}^n \alpha_i\right) \right)^{-1}, \quad (1)$$

where  $\Gamma(\cdot)$  is the complete Gamma function.

Following the assumption that a set of indicators  $\mathbf{z} = (z_1, z_2, \dots, z_n)$  is useful to disaggregate the total sum, the parameters  $\mathbf{a}$  are represented as a function of  $\mathbf{z}$ ; i.e.  $\mathbf{a} = (\alpha_1(\mathbf{z}), \alpha_2(\mathbf{z}), \dots, \alpha_n(\mathbf{z}))$ . By transforming  $\Theta$  to  $\mathbf{Y}$ , the joint probability density function of  $\mathbf{Y}$  is obtained as:

$$f_{y_1, y_2, \dots, y_n} \left( y_1, y_2, \dots, y_n \mid \mathbf{a}(\mathbf{z}), x \right) = x^{-k} \left( B(\mathbf{a}(\mathbf{z})) \right)^{-1} \prod_{i=1}^n y_i^{\alpha_i-1}. \quad (2)$$

This is the basic formulation of the probabilistic disaggregation model. In the following, a special form of the model is presented.

The vector  $\mathbf{a}$  is assumed to be parameterized as  $\mathbf{a} = k \cdot \mathbf{a}_0(\mathbf{z})$  where  $k$  is a scaling factor and  $\mathbf{a}_0(\mathbf{z}) = (\alpha_{0,1}(\mathbf{z}), \alpha_{0,2}(\mathbf{z}), \dots, \alpha_{0,n}(\mathbf{z}))$ . It is then assumed that parameters  $\alpha_{0,i}$  are modeled through disaggregation weights  $w_i(\mathbf{z}) (> 0)$  such that  $\alpha_{0,i} = w_i(\mathbf{z}) / \sum_{i=1}^n w_i(\mathbf{z})$ ; hence,  $E[\Theta_i] = \alpha_{0,i}$ .

The factor  $k$  plays a role in controlling the variability of disaggregated random variables; namely a higher value of  $k$  results in a smaller variability in  $\Theta_i$ , which implies that the ‘‘quality’’ of the disaggregation using indicator  $z_i$  is high.

The proposed parameterization of  $\mathbf{a}$  is made possible by a characteristic of the Dirichlet distribution, namely that  $Var[\Theta_i] \propto 1 / \sum_{i=1}^n \alpha_i$ . It is therefore possible to control the variance through scaling of  $\mathbf{a}$  without affecting  $E[\Theta_i]$ . Note that the model has no restriction in the definition of disaggregation weights  $w_i(\mathbf{z})$ . Further, note that for  $k \rightarrow \infty$  the probabilistic disaggregation model converges to a deterministic model.

### 3. Example

The performance of the model and its relevance to natural hazard risk assessment is illustrated at the example of a flood hazard scenario in the Canton Bern, Switzerland. The number of buildings in a portfolio is given aggregated at a communal level and is to be disaggregated to 100m grid using land cover data as indicator. The disaggregated portfolio is intersected with a recent flood scenario to determine the number of buildings affected by the flood. The result is the probability distribution of the number of affected buildings in each cell. For selected communes the result is compared to 1) the actual number of buildings affected by the scenario and 2) the result yielded by a deterministic implementation of the model ( $k \rightarrow \infty$ ). The number of buildings in each commune of the Canton  $x_c, c = 1, 2, \dots, l$  is provided by the Swiss Bureau of Statistics. Note that the disaggregation is performed in each commune separately.

The indicator  $z$  is derived from CORINE land cover data EEA (2006), which provides land cover data for European countries with a 100m resolution. For this application 52 CORINE land cover types are converted into six land cover types according to Table 1. Note that, in this example,  $z$  is a categorical variable.

Table 1. Conversion table from CORINE land cover type to the indicator  $z$ .

Indicator State	CORINE Land Cover Classes	Description
$z_i = \zeta_1$	111 and 112	Urban
$z_i = \zeta_2$	121-124 and 131-133	Industrial
$z_i = \zeta_3$	141 and 142	Urban Leisure
$z_i = \zeta_4$	211-213, 221-223, 231 and 241-244	Agricultural
$z_i = \zeta_5$	311-313, 321, 322-324 and 331-335	Forests and natural areas
$z_i = \zeta_6$	411, 412, 421-423, 511, 512 and 521-523	Wetlands and water bodies

The last major flood event affecting the Canton Bern, the August 2005 flooding, see e.g. Barredo (2007), is taken as a hazard scenario. The areas inundated during the event are available in a GIS environment.

### 3.1. Portfolio disaggregation

As indicator  $z$  is categorical, the function  $g(z_i)$  is given in tabular form, i.e. disaggregation weights  $w_i = g(z_i)$  are defined for every value of  $z$ .

The estimate of disaggregation weights is made from statistical observations. As no statistical data of direct use for this application is available, observations are made on satellite images of the Canton Bern using Google Earth. A 100m raster grid is overlaid the satellite image and the number of residential buildings per cell is manually counted. The parameters are estimated from the observations made in agricultural and urban cells in the communes of Burgdorf, Bremgarten bei Bern, Etzelkofen and Bern (partially). From 528 urban cells a sample mean  $\bar{n}_u = 7.24 \text{ buildings/cell}$  with a coefficient of variation  $CoV = 0.59$  are estimated, whereas 2158 agricultural cells result in a sample mean  $\bar{n}_a = 0.19 \text{ buildings/cell}$  with a coefficient of variation  $CoV = 4.17$ .

The disaggregation weights for the remaining values of indicator are estimated following Gallego (2010) where it is suggested that industrial areas, forests and urban leisure areas are 3 times less densely populated than agricultural areas, whereas wetlands and water bodies are assumed to be without population. As a result, following disaggregation weights were estimated:  $w_i(\zeta_1) = 110$ ,  $w_i(\zeta_2) = w_i(\zeta_3) = w_i(\zeta_5) = 1$ ,  $w_i(\zeta_4) = 3$ ,  $w_i(\zeta_6) = 0$ .

For each commune, a vector  $\mathbf{a}_0$  is determined from the estimated disaggregation weights  $w_i$ . Thereafter, parameter  $k$  is determined from  $\mathbf{a}_0$  and the observed coefficient of variation.  $k$  is estimated through an algebraic reformulation of the second moment of the Dirichlet distribution; as the great majority of buildings lay in urban areas,  $k$  is estimated based on observation made in urban cells (i.e.  $z_i = \zeta_1$ ). After estimating all parameters, a Monte Carlo simulation of the probabilistic disaggregation is performed.

The deterministic disaggregation is equivalent to the probabilistic model without disaggregation uncertainty. The number of disaggregated buildings is therefore  $y_i = \alpha_{0,i} \cdot x$ .

### 3.2. Number of affected buildings

The number of buildings affected by the scenario is determined by counting the buildings that come to stand in inundated cells. As disaggregation uncertainty is considered, the number of affected building is also uncertain.

## 4. Results

In Figure 1 a comparison is made between modeled and observed numbers of buildings per cell for the communes of Ittingen and Mühlethurnen. Note that the observations from these communes are excluded in estimating the model parameters.

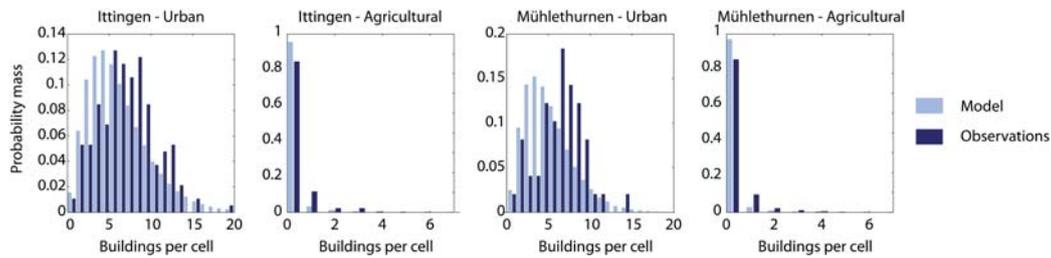


Figure 1. Observed vs. Modeled buildings per cell in communes of Ittingen and Mühlethurnen.

The number of buildings affected by the flood scenario is computed with the probabilistic disaggregation model as well as with the deterministic disaggregation model. Figure 2 illustrates the results of both disaggregation models for the communes of Thun, Wilderswil and Brienz and compares them to the number of buildings affected during the August 2005 floods.

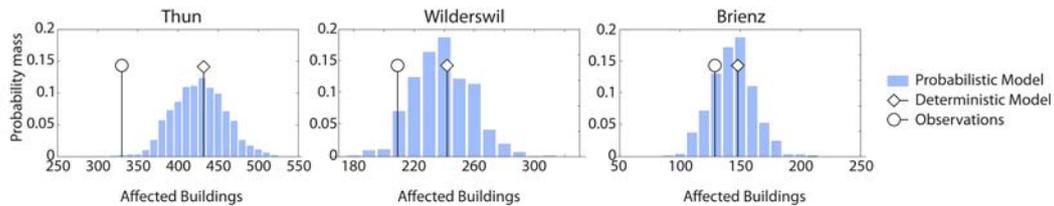


Figure 2. Numbers of affected buildings resulting from the probabilistic disaggregation model and from the deterministic disaggregation model and actual numbers from observations.

## 5. Discussion

The probabilistic disaggregation model performs as expected and disaggregation uncertainty is reasonably reproduced, see Figure 1. Although the observed and the modeled distributions shown in Figure 1 do not fully match, the probabilistic disaggregation represents a significant improvement over deterministic disaggregation. According to the results in Figure 2, the number of buildings affected by the flood scenario is overestimated by both disaggregation models. It suggests that the bias results from the chosen indicator which may convey insufficient information to characterize building densities in proximity of water bodies. For instance it can plausibly be assumed that, for a given indicator, flood plains show a lower building density than other regions. In an improved model multiple indicators may be utilized, e.g. in addition to land cover data, flood risk zone information may be included as auxiliary indicators in the disaggregation process.

The relevance of the probabilistic disaggregation for natural hazard risk assessment can be evident from Figure 2. When the disaggregation considers uncertainty, the number of buildings affected by a natural hazard event is uncertain. As a consequence the tail risk in a natural hazard risk assessment may increase.

The proposed model allows modeling the uncertainty of disaggregated variable  $Y_i$ . Nevertheless, it should be noted that the dependency between  $Y_i$  and  $Y_j$  for any  $i$  and  $j$  which is not captured by the indicators cannot be represented, as the Dirichlet distribution implies a strictly negative correlation structure, see e.g. Monti, Mateu-Figueras et al. (2011). A possible extension of the proposed model in this regard is to treat the parameters  $\alpha(\mathbf{z})$  as random variables; i.e. to model through the conditional probability  $p_A(\alpha|\beta, \mathbf{z})$  where  $\beta$  is the set of hyper parameters. Such a hierarchical modeling adds flexibility on the proposed model, whereby circumventing the aforementioned limitations.

Moreover, Dirichlet distribution is valid for continuous support variables, whereas  $Y_i$  may be discrete, requiring it to be rounded. Particular attention has to be given to cases where  $E[Y_i] \ll 1$  as a simple rounding of  $Y_i$  to the closest integer generally leads to  $\sum_{i=1}^n Y_i$  being significantly smaller than  $x$ .

## 6. Conclusions

A probabilistic disaggregation model has been proposed. It allows for disaggregating a geographically aggregated variable according to a spatial indicator with due consideration of disaggregation uncertainty. The model performance is investigated at the example of portfolios of residential buildings subject to flood risk. It is found that the model adequately disaggregates the portfolios and generally outperforms its deterministic counterpart using the same indicator. It is demonstrated how the risk tails of portfolios subject to natural hazard increases by considering disaggregation uncertainty. Based on the numerical investigation of the performance of the proposed model, several issues are identified that require further improvement of the model, and a possible extension of the model is addressed.

## References

- Aitchison, J., *The statistical analysis of compositional data*. Chapman & Hall, Ltd., 1986
- Barredo, J., *Major flood disasters in Europe: 1950–2005*, Nat. Hazards, 42(1), 125-148, 2007
- EEA, *CORINE Land Cover Data*, EEA European Environmental Agency. 2006
- Gallego, F., *A population density grid of the European Union*, Population & Environment, 31(6), 460-473, 2010
- Kron, W., *Flood Risk = Hazard • Values • Vulnerability*, Water Int'l, 30(1), 58-68, 2005
- Monti, G., G. Mateu-Figueras, et al., *The shifted-scaled Dirichlet distribution in the simplex. Proceedings of the 4th International Workshop on Compositional Data Analysis (2011)*, J. J. Egozcue, Tolosana-Delgado, R. and Ortego, M.I., Girona, Spain. 2011
- Thieken, A. H., Müller, M., Kleist, L., Seifert, I., Borst, D., and Werner, U., *Regionalisation of asset values for risk analyses*, Nat Hazards Earth Syst, 4, 167:178, 2006
- Wünsch, A., U. Herrmann, et al., *The Role of Disaggregation of Asset Values in Flood Loss Estimation: A Comparison of Different Modeling Approaches at the Mulde River, Germany*, Environmental Management, 44(3), 524-541, 2009