

# Chapter 7

## Weight of Evidence in Geospatial Analysis

Rajesh Bahadur Thapa

### 7.1 Introduction

Weight of evidence (WofE) is a quantitative method for combining evidence in support of a hypothesis. An evidence-based approach involves an assessment of the relative values of different pieces of information that have been collected in previous steps. ECHA (2010) defines WofE as “the process of considering the strengths and weaknesses of various pieces of information in reaching and supporting a conclusion.” A representative value needs to be assigned to each piece of information using a formalized weighting procedure. The evidence can be called as a factor, and can often influence the weight given owing to the quality of the data, the consistency of results, the nature and severity of effects, and the relevance of the information.

WofE is entirely based on the Bayesian approach of conditional probability. This method combines spatial data from a variety of sources to describe and analyze interactions, provides evidence for decision making, and makes predictive models. Basically, the method concerns the probability of detecting a certain event, which could be a given category of land-use change, for example, possibly an event of land-use change from agricultural area to built-up surface in relation to potential evidence (proximity to urban centers, roads, water, etc.), often called the driving factors of change (Thapa and Murayama 2011).

Historical evidence shows that Peirce (1878) was very close to the best definition of WofE, namely the logarithm of a Bayes factor, which is the ratio of the posterior to the prior odds. A theoretical expression and discussion of WofE modeling at an earlier stage can be found in Good (1950, 1979). This method was originally developed

---

R.B. Thapa (✉)

Earth Observation Research Center, Space Applications Mission Directorate,  
Japan Aerospace Exploration Agency (JAXA), Tsukuba, Japan

Formerly in Division of Spatial Information Science, Graduate School of Life  
and Environmental Sciences, University of Tsukuba, Tsukuba, Japan

e-mail: rajesh.thapa@jaxa.jp

for a non-spatial application, and therefore its applications dominated the literature in statistics and medical related fields until the mid-1980s. Its application to medical fields was promising, for instance, when the evidence consisted of a set of symptoms, and the hypothesis was “this patient has disease x.” For each symptom, a pair of weights was calculated, one for the presence of the symptom and one for the absence of the symptom. The magnitude of the weights depended on a measured association between the symptom and the occurrence of the disease in a large group of patients. The weights could then be used to estimate the probability that a new patient would get the disease, based on the presence or absence of symptoms (Spiegelhalter 1986; Raines et al. 2000). In this chapter, I briefly review the progress in the methodology and applications of WofE in the field of geospatial analysis, discuss how the method works, and conclude with an outlook for the future.

## 7.2 Methodology and Applications in Geospatial Analysis

Since the late 1980s, the geoscience field has adopted WofE modeling for geospatial applications, the earliest of which were maps of mineral potential (Bonham-Carter et al. 1988). A pattern of mineral deposits is related to several map layers representing geological data that may be indicative of the occurrence of mineral deposits (Bonham-Carter 1994). The method has gradually been integrated with spatial databases, and has been used for a variety of purposes in various geographic regions. Aspinall (1992) described an inductive modeling procedure integrated with geographical information science (GIS) and the Bayesian theorem for wildlife habitat mapping. The use of the modeling procedure is illustrated through an analysis of the winter habitat relationships of red deer in the Grampian Region, north-east Scotland. The habitat data sets used to construct the model were the accumulated frost and altitude records obtained from maps, and land cover derived from satellite imagery.

Bonham-Carter (1994) illustrated the modeling process in a probabilistic framework, so that the weighting of individual map layers was based on a Bayesian probability model. In particular, the WofE model was presented in a map context, with examples showing applications to mineral-potential mapping in Meguma terrane, Nova Scotia, Canada. The relationships of WofE to the methods used in the expert system of the prospectors are explained with a very simple example of the system’s inference network.

Similarly, Cheng and Agterberg (1999) proposed a new approach to the WofE method based on fuzzy sets and fuzzy probabilities for mineral-potential mapping. This approach can be considered as a generalization of the ordinary weights of evidence method, which is based on binary patterns of evidence and has been used in conjunction with GIS for mineral-potential mapping. In the newly proposed method, instead of separating the evidence into binary form, fuzzy sets containing more subjective genetic elements are created; fuzzy probabilities are defined to construct a model for calculating the posterior probability of a unit area containing mineral deposits on the basis of the fuzzy evidence for the unit area. This method can be

treated as a hybrid method, which allows objective or subjective definitions of a fuzzy membership function of evidence augmented by an objective definition of fuzzy or conditional probabilities. The posterior probabilities calculated by this method would depend on existing data in a totally data-driven approach, but would also depend partly on expert knowledge when the hybrid method is used.

An ArcView GIS extension of WofE, i.e., Arc-WofE, is publicly available and has applications to mineral potential (Kemp et al. 1999). This enhances further potential uses of WofE in the geospatial field. The system has four core steps: building a spatial digital database; extracting predictive evidence for a particular deposit type based on an exploration model; calculating weights for each predictive map or evidential theme; combining the evidential themes to make a prediction. This extension also provides an expert approach to weighting which can be used when no training points are available (Raines et al. 2000).

The popularity of the WofE method with more geospatial applications was further expanded in other environmental study fields in the first decade of the twenty-first century. A structure for simulating land-use change using the elementary probabilistic methods of the WofE approach was proposed by Almeida et al. (2003). The model framework has been applied to Bauru town in Brazil. This showed how various socio-economic and infrastructural factors can be combined using the WofE approach, which then enables them to predict the probability of changes between land-use types in different cells of the system. Another study over a larger area, which was also conducted in Brazil by Soares-Filho et al. (2004), shows an application of the WofE method to select the most important variables needed for land-cover change analysis and to quantify their influences on each type of land-use transition, e.g., deforestation, land abandonment, and re-growth clearing. They developed a land-cover change simulation model that is responsive to road paving and policy intervention scenarios in central Amazonia. The model assesses the impacts of road paving within the population, as well as policy intervention scenarios.

Romero-Calcerrada and Luque (2006) focused on boreal forest landscapes, and explored a multicriteria approach by using a predictive habitat suitability model for the three-toed woodpecker (*Picoides tridactylus*) based on WofE. Since the method depends on the indicator species which is used as a surrogate of biodiversity value, it can be applied to assess the biodiversity conditions of both managed and protected areas to help decision-making concerning the protection of valuable habitats. Thus, a map of habitat suitability representing a range of probabilities of occurrence offers an objective framework for evaluating the outcomes of different scenarios. Similarly, an objective assessment of habitat suitability provides a rational basis for management decisions incorporating the impact on species habitat. Romero-Calcerrada and Millington (2007) used both WofE and logistic regression to analyze the natural and human factors that contribute to wildfire on the Iberian Peninsula. Unlike expert knowledge approaches to modeling, the WofE approach derives the probabilities of fire occurrence based on the association between mapped occurrences and spatial evidence layers of biophysical data.

An application of the WofE method to perform a vulnerability assessment for the occurrence of elevated nitrate concentrations in the aquifer of Milan, Italy, is

given by Masetti et al. (2007). A comparison between the spatial distribution of vulnerability classes and the frequency of occurrences of nitrate in wells shows a high degree of correlation for both low and high nitrate concentrations. Groundwater-specific vulnerability was classified in terms of vulnerability classes and, according to the outcomes of the model, the population density can be considered to be the source of the greatest impact of nitrate. Mean annual irrigation and groundwater depth can be identified as influencing factors in the distribution of nitrate, while agricultural practice appears to be a negligible factor.

Dahal et al. (2008) and Pradhan et al. (2010) presented WofE modeling applications to landslide susceptibility mapping. The former applied the modeling to small catchments of Shikoku, Japan, while the latter applied it to a tropical hilly area in Malaysia. The Japanese case showed the usefulness and capability of the modeling in a small catchment area with a high-resolution data set. The Malaysian case showed the method of calculating the rating factor, and reported that the landslide susceptibility map and the verification results achieved a high predictive accuracy for the model.

Dilts et al. (2009) used WofE techniques to model spatial patterns of wildfire occurrence in relation to landscape-scale drivers of fire in Lincoln County, Nevada, USA. The spatial data sets which were used as potential predictors of fire occurrence included biophysical and socio-economic data. Models were developed and tested for lightning-caused fires over the entire county, and also in forested areas only. Higher fire density and higher lightning-strike density were observed in the eastern half of the county compared with the western half. Overall, the spatial distribution of wildfire occurrence was controlled more by ignition mechanisms than by processes influencing fuel moisture, accumulation, or both.

A recent application of WofE is found in urban growth modeling. Thapa and Murayama (2011) adopted the WofE method integrated in a cellular automata framework for predicting the future spatial patterns of urban growth in the Kathmandu metropolitan region. The model was validated by achieving a highly accurate prediction of urban development patterns for the future under the current scenario across the metropolitan region. Depending on local characteristics and land-use transition rates, the model produced a noticeable spatial pattern of changes in the region. The application of WofE to urban growth modeling can be found in Chap. 13.

### 7.3 WofE Model: How Does It Work?

Let's now consider a landscape (Fig. 7.1) which has three spatial patterns: forest, road buffer, and an area which is changing from forest to non-forest land. Landscape change is observed along the road network, so the road is considered to be the major driver of the change. If we considered this in binary terms, area change is represented as 1 and no change as 0. Similar assumptions can be made for the road layer, i.e., inside the road buffer as 1 and outside as 0. In this particular case, the WofE concerns the probability of detecting land change (deforestation) influenced by the driver (road).

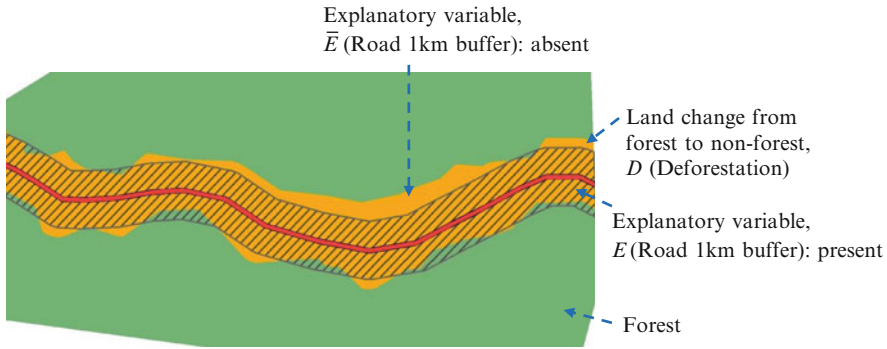


Fig. 7.1 Schematic drawing to illustrate the WofE method

To understand this deforestation process and detect the probability using the WofE technique, the WofE model is synthesized from Bonham-Carter (1994). The areas of deforestation,  $D$  (landscape change from forest to non-forest), and the explanatory variable,  $E$  (road buffer), are known, and then the probability of locating the occurrence of deforestation given the presence of the explanatory variable can be expressed by the conditional probabilities given in (7.1).

$$P(D | E) = \frac{P(D \cap E)}{P(E)} \tag{7.1}$$

The symbol  $\cap$  is a logical intersection or Boolean AND operation. The conditional probability of  $D$  occurring given the presence of  $E$  is written as  $P(D | E)$ . Thus, the probability of a deforestation pattern ( $D$ ) occurring given the presence of explanatory variable ( $E$ ) can be expressed as a probability ratio, which follows from the basic definition of conditional probability, followed by the substitution of area proportions as estimates of probabilities, and finally as a ratio of areas. The probability of land change in the WofE modeling framework is expressed as *odds*. *Odds* ( $O$ ) are defined as  $O = P/(1-P)$ , a ratio of the probability that an event will or will not occur (Bonham-Carter 1994). Now, (7.1) can be converted into odds (7.2).

$$O(D | E) = \frac{P(D | E)}{1 - P(D | E)} = \frac{P(D | E)}{P(\bar{D} | E)} \tag{7.2}$$

where  $\bar{D}$  represents the absence of deforestation, i.e., no change occurred in the landscape.  $1 - P(D | E)$  becomes  $P(\bar{D} | E)$  when we consider the probability of  $D$  (deforestation) being absent given the presence of  $E$  (explanatory variable, road). *Odds* values of less than 1 correspond to probabilities less than 0.5, and very small probabilities are nearly the same as *odds*. Accordingly, a similar argument is used to derive an expression for the conditional odds of  $D$  given the absence of  $E$ :

$$O(D | \bar{E}) = \frac{P(D | \bar{E})}{P(\bar{D} | \bar{E})} \tag{7.3}$$

The WofE method can combine several explanatory variables to predict similar patterns of land change. A pair of weights  $W^+$  (presence) and  $W^-$  (absence) can be determined for each predictor pattern (road, and other predictor if any), depending on the measured spatial association with the pattern of land change. The weights may be combined from each pattern to make a predictive map for the change. Taking a single predictor pattern,  $D$ , the positive weight  $W^+$  and the negative weight  $W^-$  can be expressed as the difference between the prior and posterior logit of  $D$ , as follows:

$$W^+ = \ln O(D | E) - \ln O(D) = \ln \left[ \frac{O(D | E)}{O(D)} \right] \quad (7.4)$$

$$W^- = \ln O(D | \bar{E}) - \ln O(D) = \ln \left[ \frac{O(D | \bar{E})}{O(D)} \right] \quad (7.5)$$

The WofE method uses the natural logarithm of *odds*, known as *log odds* or *logit*. The logit scale is centered at 0, corresponding to a probability of 0.5, with negative values for *odds* less than 1 and positive values for *odds* greater than 1. After computing the weights, the posterior logit can be generated using the following equations:

$$\ln O(D | E) = \ln O(D) + W^+ \quad (7.6)$$

$$\ln O(D | \bar{E}) = \ln O(D) + W^- \quad (7.7)$$

More explanatory variables can be incorporated with an assumption that the variables are conditionally independent with respect to land change (i.e., deforestation). The following expression can be written for more explanatory variables:

$$\ln O(D | E_1^k \cap E_2^k \cap \dots \cap E_n^k) = \ln O(D) + \sum_{j=1}^n W_{E_j}^k \quad (7.8)$$

where the superscript  $k$  is positive (+) or negative (-) depending on whether the explanatory variable is present or absent, respectively.

The explanatory variables are usually either discrete (e.g., a land-use planning map or other socio-economic data) or continuous (e.g., a slope, proximity to road, river, etc.). Continuous variables need to be transformed into discrete variables. Thence, each explanatory theme has  $k=1,2,\dots,m$  discrete class values or states which can be associated with weights in respect to the occurrence of events.

The WofE application on landslide susceptibility mapping synthesized from Dahal et al. (2008) is presented as an example. The application test area, with less than 400 ha of the Moriyuki catchment, is located in the northeast of the Shikoku region, Japan. The catchment had as many as 76 landslides (Fig. 7.2) due to heavy rainfall which occurred in October 2004. Thematic maps, i.e., slope, aspect, relief,

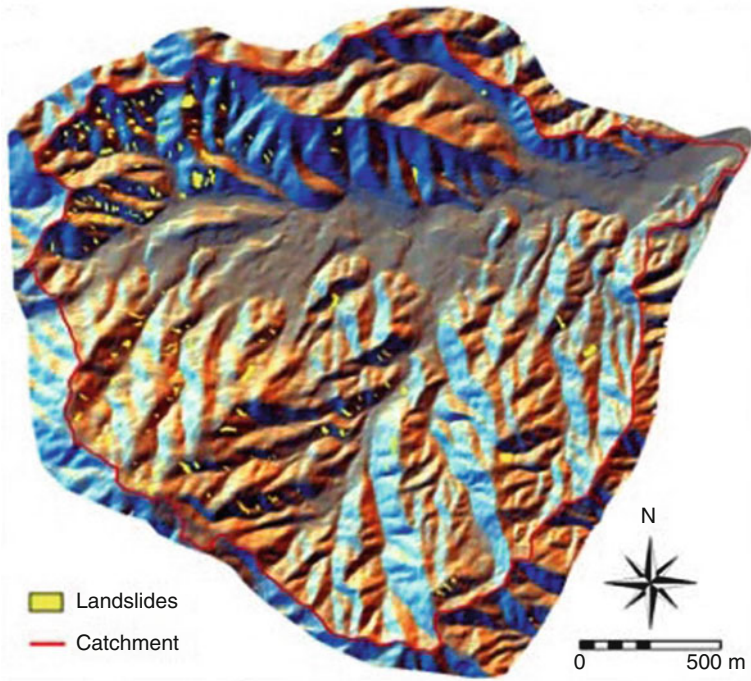
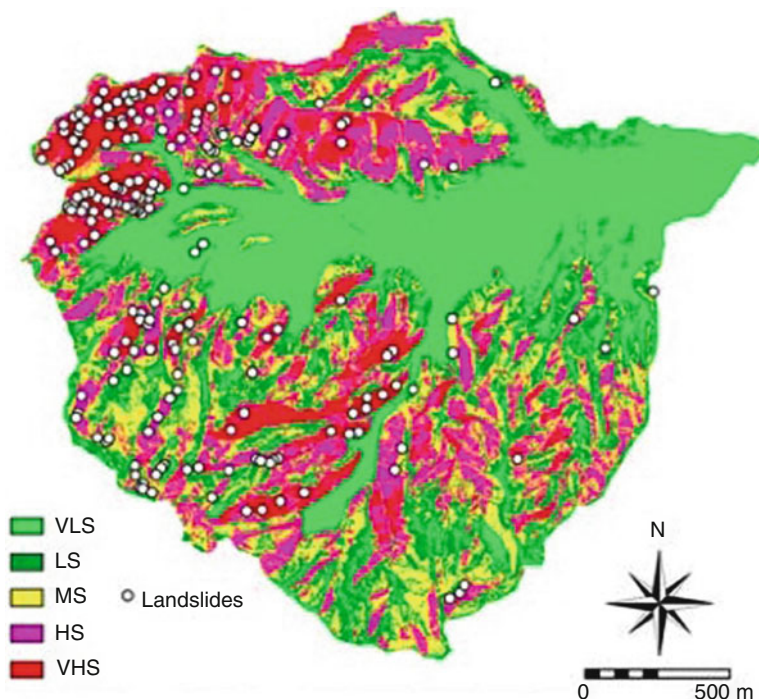


Fig. 7.2 Landslide events in the Moriyuki catchment (reproduced from Dahal et al. 2008)

flow accumulation, soil depth, soil type, land use, and distance to road, were used as landslide predictive factors. The size of the landslide and the number of potential events varies in different landscapes depending on local geo-environmental characteristics. Therefore, each thematic map was logically classified into various category ranges in order to estimate geographically varying weights. The maps of categorical factors were overlaid with the landslide event map, and this produced weights for each map using WofE techniques. The weighted factor maps were linearly combined to create a landslide susceptibility index map. The index map was cross-validated with a landslide event map and showed a considerable success rate, i.e., 80.7% of the WofE-based predictive model. Based on the success rate ratio, a map with five landslide susceptibility zones, i.e., very low, low, moderate, high, and very high, was established and is shown in Fig. 7.3.

## 7.4 Future Outlook

From the discussion and examples above, it is clear that the WofE method can combine spatial data from diverse sources to describe and analyze interactions, provide support for decision makers, and make predictive models. The statistical association between an event and the associated factors determines the weights. The WofE



**Fig. 7.3** Landslide susceptibility map of the Moriyuki catchment. *VHS* very high susceptibility, *HS* high susceptibility, *MS* moderate susceptibility, *LS* low susceptibility, *VLS* very low susceptibility (reproduced from Dahal et al. 2008)

method itself is combined with Bayes' rule of probability, with an assumption of conditional independence. The model is given in log-linear form so that the weights from the evidential themes can be added.

In this method, weight values are easy to interpret. A positive weight for a particular evidential-theme indicates that a larger proportion matched the conditions of that theme than would occur due to chance, whereas the inverse is true for negative weights. A weight of zero indicates that the training points are spatially uncorrelated to the theme. The range in weight values for a particular evidential theme gives an overall measure of how important the theme is for modeling. Uncertainties due to variances of weights and missing data allow the relative uncertainty in posterior probability to be estimated and mapped. Because conditional independence is never completely satisfied, the posterior probabilities are usually overestimated in absolute terms. However, the relative variations in posterior probability (as observed in spatial patterns on the response map) are usually not much affected by violations of this assumption, as stated by Raines et al. (2000).

Being a Bayesian method, it has a number of advantages over other more traditional spatial modeling methods. Much appreciated are the ability to accommodate



both categorical and ordered data, the robustness to small sample sizes, and the ability to use data that are not normally distributed (Bonham-Carter et al. 1988). As we have observed several applications in the geospatial field, i.e., mineral-potential mapping, landslide susceptibility, habitat suitability, wildfire, land-use and land-cover change analysis, and urban growth modeling, the future of this method is emerging. It may be possible that the wider applications of WofE will soon be used to solve global environmental problems, such as in REDD+.

## References

- Almeida CM, Batty M, Monteiro AVM, Camara G, Soares-Filho BS, Ceroueira GC, Pennachin CL (2003) Stochastic cellular automata modelling of urban land use dynamics: empirical development and estimation. *Comput Environ Urban Syst* 27:481–509
- Aspinall RJ (1992) An inductive modeling procedure based on Bayes' theorem for analysis of pattern in spatial data. *Int J Geogr Inform Syst* 6:105–121
- Bonham-Carter G (1994) *Geographic information systems for geoscientists: modeling with GIS*. Pergamon, New York
- Bonham-Carter GF, Agterberg FP, Wright DF (1988) Integration of geological datasets for gold exploration in Nova Scotia. *Am Soc Photogram Rem Sens* 54:1585–1592
- Cheng Q, Agterberg FP (1999) Fuzzy weights of evidence method and its application in mineral potential mapping. *Nat Resour Res* 8:27–35
- Dahal RK, Hasegawa S, Nonomura A, Yamanaka M, Masuda T, Nishino K (2008) GIS-based weights-of-evidence modelling of rainfall-induced landslides in small catchments for landslide susceptibility mapping. *Environ Geol* 54:311–324
- Dilts TE, Sibold JS, Biondi F (2009) A weights-of-evidence model for mapping the probability of fire occurrence in Lincoln county, Nevada. *Ann Assoc Am Geogr* 99:712–727
- ECHA (European chemical Agency) (2010) *Practical guide 2: how to report weight of evidence*. <http://echa.europa.eu/>. Accessed 4 Mar 2011
- Good IJ (1950) *Probability and the weighing of evidence*. C. Griffin, London
- Good IJ (1979) Studies in the history of probability and statistics: A. M. Turing's statistical work in World War II. *Biometrika* 66:393–396
- Kemp LD, Bonham-Carter GF, Raines GL (1999) Arc-WofE: arcview extension for weights of evidence mapping. <http://www.ige.unicamp.br/wofe>. Accessed 7 Mar 2011
- Masetti M, Poli S, Sterlacchini S (2007) The use of the weights-of-evidence modeling technique to estimate the vulnerability of groundwater to nitrate contamination. *Nat Resour Res* 16:109–119
- Peirce CS (1878) The probability of induction. *Popular Science Monthly* 12:705–718. Reprinted (1956) in Newman JR (ed) *The world of mathematics*, vol 2. Simon and Schuster, New York, pp 1341–1354
- Pradhan B, Oh H, Buchroithner M (2010) Weights-of-evidence model applied to landslide susceptibility mapping in a tropical hilly area. *Geomat Nat Hazards Risk* 1:199–233
- Raines GL, Bonham-Carter G, Kemp L (2000) Predictive probabilistic modeling: using ArcView GIS. *ArcUser*, April–June, pp 45–48. <http://www.esri.com>
- Romero-Calcerrada R, Luque S (2006) Habitat quality assessment using weights-of-evidence based GIS modelling: the case of *Picoides tridactylus* as species indicator of the biodiversity value of the Finnish forest. *Ecol Model* 196:62–76
- Romero-Calcerrada R, Millington JDA (2007) Spatial analysis of patterns and causes of fire ignition probabilities using logistic regression and weights-of-evidence based GIS modeling. *Geophys Res Abstr* 9:01337

- Soares-Filho BS, Alencar A, Nespad D, Cerqueira GC, Dial M, Del C, Solozarno L, Voll E (2004) Simulating the response of land-cover changes to road paving and governance along a major Amazon Highway: the Santarem–Cuiaba corridor. *Glob Chang Biol* 10:745–764
- Spiegelhalter DJ (1986) Probabilistic prediction in patient management and clinical trials. *Stat Med* 5:421–33
- Thapa RB, Murayama Y (2011) Urban growth modeling of Kathmandu metropolitan region, Nepal. *Comput Environ Urban Syst* 35:25–34