International Environmental Modelling and Software Society (iEMSs)
2012 International Congress on Environmental Modelling and Software
Managing Resources of a Limited Planet, Sixth Biennial Meeting, Leipzig, Germany
R. Seppelt, A.A. Voinov, S. Lange, D. Bankamp (Eds.)
http://www.iemss.org/society/index.php/iemss-2012-proceedings

Approaching Uncertainties in Land-Use Change Modeling in the Amazon Rainforest with Bayesian Belief Networks

Carsten Krüger[a], Dennis Funke[a], Tobia Lakes[a]
[a] *Humboldt-Universität zu Berlin, Geography Department, Geomatics Lab, Unter den Linden 6, 10099 Berlin, carsten.krueger@geo.hu-berlin.de*

Abstract: In recent years, modeling techniques to study land-use change and to develop scenarios for highlighting possible future pathways of land-use change have been increasingly applied and studied. While modeling results always incorporate uncertainties in terms of input data, model parameters, model boundaries, and model structure, up to now, only few land-use change modeling approaches have explicitly focused on uncertainties using e.g. Monte-Carlo simulation or Fuzzy logic. Decision makers, however, are not only interested in understanding possible consequences, but also in evaluating the reliability of these. Since land-use change is spatially dependent on its neighborhood, we especially address uncertainties associated with different settings of a spatial land use in neighborhood variable. We apply our land-use change model to Southern Amazon, Brazil, where rapid deforestation of large areas occurred in the past with high impacts on soil erosions, droughts, floods, and migration. To explicitly account for uncertainty, we apply a Bayesian Belief Network (BBN) approach. Uncertainty is integrated in terms of conditional probabilities between spatial variables; i.e. one state of one variable is conditionally dependent on the states of other variables. Our results highlight the capability of the BBN approach to investigate observed spatial dynamics of land-use change. Additionally, we are able to evaluate the uncertainty associated with a specific input source. Data uncertainty in terms of the represented spatial neighborhood substantially influences the modeling results. A further possible enhancement is the integration of qualitative expert knowledge. The investigation of uncertainty propagation to land-use change scenarios is a future challenge.

***Keywords**: Spatial Modeling, Land-Use Change Modeling, Spatial Weighting Matrix, Spatial Neighborhood, Uncertainty Analysis*

1. INTRODUCTION

The Amazon forest is considered as the largest remaining tropical forest in the world, suffering high deforestation rates within the last decades. Estimates by Araujo et al. [2009] claim that by 2007 the total deforested area in the Brazilian Amazon accumulated up to 700,000 km², with 80 % of those being deforested after 1977. The impact on essential ecosystem services, such as the disturbance of the global carbon cycle, habitat loss and fragmentation for a variety of animal species or the disturbance of the regional water cycle, are related to deforestation [*Walker*, 2012]. Agricultural and infrastructural development for economic growth and participation in the global market on the one hand and sustainable development with the focus on environmental protection to reduce deforestation and its impacts on the other hand characterize the opposing land-use interests in Brazil's less developed states of Amazonia [*Coy and Klingler*, 2011].

By means of land-use change modeling we try to understand patterns and dynamics of deforestation to offer new insights for future development paths. Spatial patterns and neighborhoods are particularly important as explanatory variables for land-use change [*Anselin*, 2002; *Hoymann*, 2010].

Within the process of land-use change modeling uncertainty is inherently integrated at several stages, such as in the choice of integrated explanatory variables, the investigated relations between variables and the chosen parameterization of the model [*Walker et al.*, 2003; *Warmink et al.*, 2010]. Uncertainty can be understood as the deviation from an unknown real value within the mentioned categories [*Walker et al.*, 2003]. These stages cannot be considered independently from each other, since uncertainties propagate themselves through the whole modeling process [*Refsgaard et al.*, 2007].

Uncertainty investigations comprise manifold different methods. Refsgard et al. developed a conceptual guideline to deal with uncertainties in environmental modeling [2007]. Van der Sluijs et al. presented collections of tools to address uncertainty in environmental applications [2004]. Up to now, only few studies transfer concepts of uncertainty assessment to the domain of land-use change modeling [*Pontius and Neeti*, 2010]. Moreover, uncertainty is addressed by extrapolating measured accuracies of modeling results into the future [*Pontius Jr and Spencer*, 2005] rather than investigating uncertainty within the stages of modeling described above.

One appropriate method to incorporate uncertainty in a model is the application of Bayesian Belief Networks (BBNs) [*Stassopoulou et al.*, 1998]. BBNs are a graphical representation of the joint probability distribution of a set of variables in the form of directed acyclic graphs [*Pearl*, 1988]. Driving variables of land-use change are represented by nodes and their relationships by edges between the nodes. Relationships are mirrored by conditional probabilities, and thus, can reflect parameter uncertainty. The final outcomes of a BBN are probabilities for every state of the investigated target variable. Earlier studies have already applied BBNs for land-use modeling but did not explicitly focus on uncertainty involved [*Bashari et al.*, 2008; *McCloskey et al.*, 2011]. In our study, we understand uncertainty as an equal probability for every possible state of a certain node in a BBN.

Monte-Carlo simulations and Fuzzy logic are two alternatives to quantify uncertainties. Nash & Hannah used BBNs and Monte-Carlo simulations to examine the effect of the usage of fertilizer [2011]. While the results are comparable, the authors concluded that the BBNs are superior because of the possibilities to include evidence in the network (fix one node at a certain state), and to have the opportunity of diagnostic and predictive reasoning. Another approach to include uncertainty and in particular vagueness in a model is Fuzzy logic. In comparison, probabilities such as in a BBN, can represent uncertainty caused by various sources of errors [*Longley et al.*, 2010].

In this study, we address the topic of uncertainty in land-use change modeling with the following two research questions:

1. How can uncertainty be addressed in the land-use modeling process by means of BBNs?
2. How much uncertainty is related to the input data, exemplarily addressed by using different spatial neighborhoods?

2. DATA AND METHODS

The study area of approximately 32,000 km² includes parts of 11 municipalities located within the border region of northern Mato Grosso and southern Pará in the Southern Amazon. High deforestation rates between 1992 -2010 of 12.9 % (Mato Grosso) and 8.6 % of the forested area (Pará) are characterizing the forest development during the last two decades *[INPE*, 2012]. The western part of the study area is traversed by the federal highway BR-163 considered as a 'corridor of modernization and destruction' [*Coy and Klingler*, 2011].

We derive land-use change maps from Landsat classifications (see table 1). Socioeconomic data are resampled to a resolution of 100x100 m. Additionally, we derive proximity data by calculating Euclidean distances. The neighborhood variables are calculated on the Landsat classifications.

Table 1: Data description

Variable	*Year*	*Resolution*	*Source*
Forest cover change	1992-2001, 2001-2010	100x100 m	Landsat classification
Distance to indigenous areas	1992, 2001	100x100 m	MMA
Distance to protected areas	1992, 2001	100x100 m	MMA
Distance to roads	2006	100x100 m	IBAMA
Distance to rivers	1992	100x100 m	IBAMA
Distance to urban areas	≤ 2003	100x100 m	SEPLAN
Elevation	1992	100x100 m	USGS SRTM data
Land use in neighborhood	**1992 & 2001**	**100x100 m**	**Based on classification maps**
Slope	1992	100x100 m	USGS SRTM data
Agricultural GDP per capita	1992, 2001	Municipality	IPEA
Cattle density	1992, 2001	Municipality	IBGE
Employement share in primary sector	1992, 2001	Municipality	IBGE
Employement share in industrial sector	1992, 2001	Municipality	IBGE
GDP per capita	1992, 2001	Municipality	IPEA
Population density	1991, 2000	Municipality	IPEA
Population growth	1991-2000	Municipality	IPEA
Urban population density	1991, 2000	Municipality	IPEA

IBAMA - Instituto Brasileiro do Meio Ambiente e dos Recursos Naturais
IBGE - Instituto Brasileiro do Geografia e Estatistica
IPEA - Instituto de Pesquisa Economica Aplicada
MMA - Ministerio do Meio Ambiente Brazil
SEPLAN - Secretaria de Estado de Planejamento e Coordenação Geral

We apply Bayesian Belief Networks (BBNs) to study deforestation in Southern Amazon to explicitly address uncertainties in our land-use change model. We use forest cover change as the target variable. The structure of a BBN can be defined either by expert knowledge or by learning algorithms [*Neapolitan*, 2004]. The first approach is helpful in case of rare data sets but increases also subjectivity. The difficulty of the second possibility is its complexity. A relatively small set of six variables has more than 3.7 million different possible combinations to construct BBNs. To address this problem, approximate learning algorithms have been developed [*Neapolitan*, 2004]. Some of these are incorporated in the GNU-R packages DEAL [*Bøttcher and Dethlefsen, 2003*] and MASTINO [*Mascherini et al.*, 2008], however, they are still restricted in the number of variables. Due to the high complexity incorporated by 19 model variables, we are not able to select the best structure by learning. Therefore, we develop a strategy to select a subset of variables by means of logistic regressions. We choose those variables that were significant at the 5 % significance level, after applying a stepwise logistic regression following the forward selections procedure bases on the Wald statistic (see table 2).

Table 2: Results of the logistic regressions

1992-2001		2001-2010	
Variable	Effect coefficient	Variable	Effect coefficient
Urban population density	0.848	Elevation	0.996
Distance to rivers	0.99996	Agricultural GDP per capita	0.999
Distance to protected areas	0.999992	Distance to protected areas	1.000008
Distance to roads	1.00002	Population density	1.00002
Elevation	1.011	Slope	1.062
Employment share, primary sector	1.036	Population growth	1.141
Land use in neighborhood	2.414	Land use in neighborhood	5.853

In addition we incorporate expert knowledge in model structure development to some degree by defining impossible links and directions of links between variables. We assume that land-use change can only be dependent on other variables, but not the other way around. Furthermore, elevation, distance to rivers and slope (except of elevation) cannot be influenced by other variables.
We integrate different weighting matrices in the BBNs to investigate the influence of spatial neighborhood effects on deforestation. Each matrix reflects the amount of land-use change in a defined environment. The first weighting matrix is defined as a circle with a radius of 200 m (see figure 1). All pixels which are situated entirely within the scope of this matrix are equally weighted. Partially included pixels are weighted with their proportion falling into the area of the weighting matrix. The next three matrices are represented by different annuluses with increasing inner and outer radius: 200 m to 400 m, 400 m to 600 m, and 600 m to 800 m. The weighting coefficients are calculated analogously to the first matrix. A further matrix should incorporate the effects of all previously mentioned matrices. This matrix has decreasing weights if the distance to the observed raster cell increases.
We use two time periods to take the temporal dynamics in the Amazon rainforest into account [*Fearnside*, 2008]. We randomly select 90 % out of 7000 points for training and 10 % for testing the model. To apply the BBN, we discretize every variable in three classes, except of the neighborhood variable. Therefore, we chose two classes, since most elements are allocated in a class with only forest pixels in the defined neighborhood. After constructing and training the model, a sensitivity analysis is undertaken, to estimate the individual influence of any variable, and of any specification of the neighborhood variable.

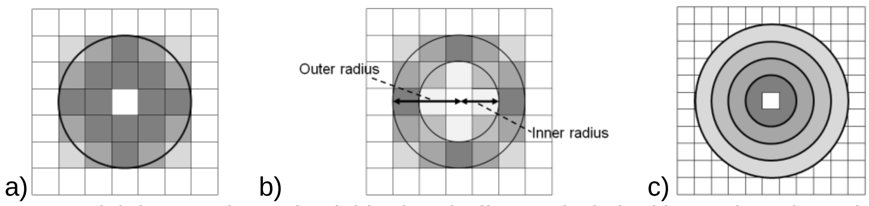
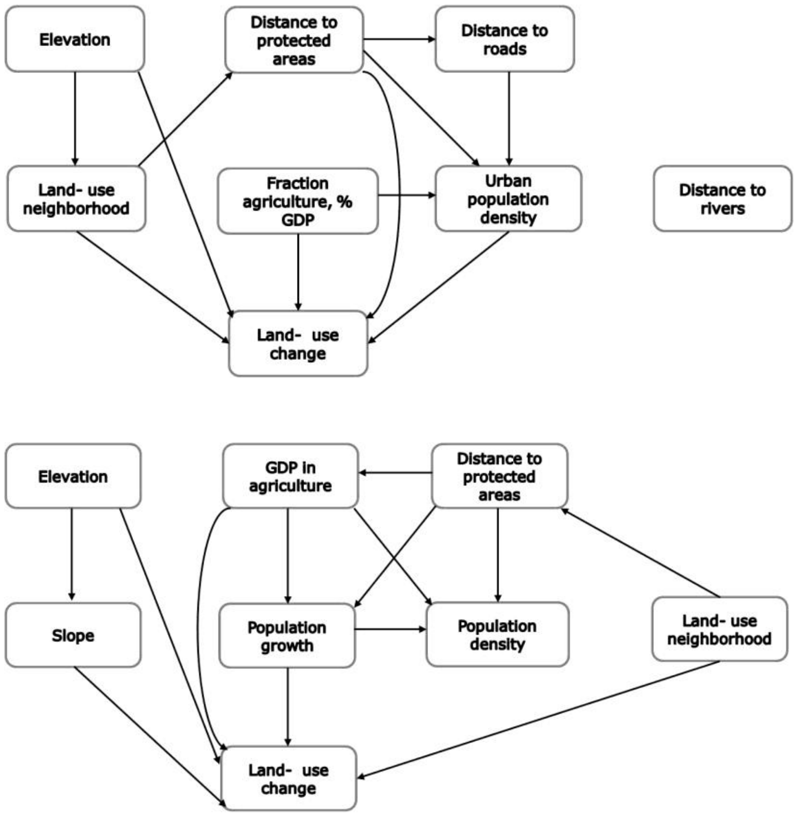


Figure 1: Weighting matrices of neighborhood effects. a) circle, b) annulus, c) combination of a) and b). The brighter the shade the lower the weights.

3. RESULTS

Our resulting BBNs reflect the capabilities of the selected variables to explain deforestation in our study area (see absolute error in table 3). The model structure and the learned dependencies are represented in figure 2. Except of distance to rivers, all variables of the first time period are either directly or indirectly connected with the target variable land-use change between 1992 and 2001. In contrast, the learned network of the second time period has every input variable included in the network structure.



a)

b)

Figure 2: Learned structures of the BBNs for a) 1992-2001 and b) 2001-2010

To investigate the effect of different input data - and in particular different spatial neighborhoods - on uncertainty in a land-use change model, we tested and compared different modeling results. For both time periods land-use in neighborhood is a substantially influencing variable in our land-use change models (see table 3). A sensitivity analysis highlights its strong impact on deforestation compared to the remaining variables and reveals the second largest (1992-2001) and the largest (2001-2010) influence. The weighting matrix with only the direct spatial neighborhood included, leads to a probability of deforestation of 30.8 % if sparse forested area is inside the neighborhood. In comparison, the probability of deforestation is only about one third of that probability, if the neighborhood is densely forested. Both time periods reveal the same tendency: with increasing inner and outer radius of the weighting matrix, the differences in probability between both mentioned states of the neighborhood variable decrease.

If we understand a probability distribution of 50 % for both states as a situation with complete uncertainty, then the uncertainty decreases simultaneously with increasing inner and outer radius. At the same time, the second time period shows a higher average uncertainty in all models.

Table 3: Sensitivity analysis

Neighborhood	1992-2001			2001-20010		
	Error, %	State 1	State 2	Error, %	State 1	State 2
0-200 m	10.92	30.8	12.4	20.13	45.8	16.9
200-400 m	12.14	23.8	11.8	18.97	38.7	14.5
400-600 m	12.62	21.1	11.4	19.23	35.9	12.7
600-800 m	12.62	20.3	10.6	19.74	33.5	11.9
0-800 m	12.86	19.5	10.4	19.49	32.4	11.3

State 1/2: Probability in % for deforestation, if the neighborhood node is fixed at state 1 (sparsely forested) or state 2 (densely forested)
Error: absolute error

4. DISCUSSION

In this study, we applied BBNs for modeling deforestation in Southern Amazon for two different time periods. The models were able to adequately reflect the process of deforestation in that area.
Referring to our first question, we want to stress the abilities of BBNs to address uncertainty in the land-use modeling process. Uncertainty was reflected by the learned conditional probabilities between variables in the different models. By characterizing complete uncertainty with an equal probability for every state, we were able to differentiate degrees of uncertainties for the different spatial neighborhoods included in the models.
One challenge in the application of BBNs is the definition of the structure. Since structure learning reaches its limitations in a large set of variables, we defined a subset, based on results of a logistic regression. Additionally, expert knowledge could be integrated to more adequately reflect the complexity of the land-use system in terms of variables and links between them [*Marcot et al.*, 2006].
Despite the problem of network construction, we got valuable results for land-use change and could address uncertainties related to the spatial neighborhood. One difficulty is the separation of real uncertainty as the deviation from a uniform distribution and a simply higher probability of deforestation if the definition of the neighborhood changes. A solution for this matter can be the calculation of possible ranges of the conditional probabilities by varying input parameters.
Referring to our second question, input data showed a significant influence on uncertainty for the different spatial neighborhoods. The further away the defined spatial neighborhood is, the lower the probabilities for deforestation are. This effect can be due to two reasons. On the one hand, the deforestation frontier encroaches at first in the direct neighborhood and then on areas further away [*Aguiar et al.*, 2007]. On the other hand, we have less knowledge about the direct neighborhood, reflected in the deviation of the resulting probability distributions from a uniform distribution. Moreover, the influence of the direct neighborhood cannot be only due to spread over reasons of deforestation, it can also be a consequence of missing variables [*Müller et al.*, in press].
Both time periods are characterized by declining differences of both states of the neighborhood variable with increasing inner and outer radius. This tendency highlights a higher sensitivity of land-use change towards the direct neighborhood. These results are comparable with other land-use modeling studies [e.g. *Hoymann*, 2010].
Another aspect which is reflected in the conditional probabilities is a higher probability for deforestation in the second time period. This may be due to deforestation is much more affected by the direct neighborhood in the second time period, or the uncertainty related to the link between land use in neighborhood and land-use change is higher between 2001 and 2010.
In comparison with other methods to analyze uncertainties, such as Monte-Carlo simulations and Fuzzy logic, BBNs are an appropriate tool to explicitly address uncertainty in land-use change modeling, because of several possibilities to look for consequences of different error sources: diagnostic (e.g. driver analysis) as well as

prognostic (e.g. scenario-building) reasoning for land use change modeling. Moreover, it is possible to include evidence in the network and to examine the resulting changes of uncertainty. Therefore, in a future step BBNs could be used to investigate the propagation of uncertainties to land-use change scenarios.

5. CONCLUSIONS

This study serves to analyze uncertainties in land-use change modeling, especially connected to the spatial neighborhood. We applied different BBNs and analyzed the learned conditional probabilities reflecting the relationships between our model variables. Our results show a considerable potential of BBNs to analyze uncertainties in land-use change although the interpretation of the conditional probabilities is ambivalent. One explanation for decreasing probabilities with increasing distance of the involved neighborhood is less influence of areas further away. At the same time, the direct land use in neighborhood could lead due a higher uncertainty due to possible missing explaining variables. Further analysis could incorporate possible ranges of the probabilities to support the interpretation of the respective variables. We tested that input data has considerable influence on uncertainty – in a next step further stages of land-use change modeling could be addressed, up to the propagation of uncertainties to land-use change scenarios to make use of the beneficial properties of BBNs, e.g. reasoning in two directions and incorporation of evidence.

ACKNOWLEDGMENTS

The presented work is supported by Deutsche Forschungsgemeinschaft GRK 1324/1 (METRIK). The authors would like to thank Florian Gollnow for helpful comments and fruitful discussions.

REFERENCES

Aguiar, A. P. D., G. Câmara, and M. I. S. Escada (2007), Spatial statistical analysis of land-use determinants in the Brazilian Amazonia: Exploring intra-regional heterogeneity, *Ecological Modelling, 209*(2–4), 169–188, http://www.sciencedirect.com/science/article/pii/S0304380007003377.

Anselin, L., Under the hood Issues in the specification and interpretation of spatial regression models, *Agricultural Economics, 27*(3), 247–267, 2002.

Araujo, C., C. A. Bonjean, J.-L. Combes, P. Combes Motel, and E. J. Reis, Property rights and deforestation in the Brazilian Amazon, *Ecological Economics, 68*(8–9), 2461–2468, http://www.sciencedirect.com/science/article/pii/S0921800908005417, 2009.

Bashari, H., C. Smith, and O. J. H. Bosch, Developing decision support tools for rangeland management by combining state and transition models and Bayesian belief networks, *Agricultural Systems, 99*(1), 23–34, http://www.sciencedirect.com/science/article/pii/S0308521X08000966, 2008.

Bøttcher, S. G., and C. Dethlefsen, Learning Bayesian Networks with R, in *DSC Working Papers*, 2003.

Coy, M., and M. Klingler, Pionierfronten im brasilianischen Amazonien zwischen alten Problemen und neuen Dynamiken. Das Beispiel des "Entwicklungskorridors" Cuiabá (Mato Grosso) -Santarém (Pará), *Innsbrucker Geographische Gesellschaft: Innsbrucker Jahresbericht 2008-2010*, 109–129, 2011.

Fearnside, P. M., Amazonia and deforestation, in *The Oxford Companion to Global Change.*, edited by A. S. Goudie, and D. J. Cuff, pp. 21–27, Oxford University Press. New York, USA, 2008.

Hoymann, J., Spatial allocation of future residential land use in the Elbe River Basin, *Environment and Planning B: Planning and Design, 37*(5), 911–928, 2010.

Insituto Nacional de Pesquisas Espaciais (INPE), *Taxas anuais do desmatamento – 1988-2011*, accessed February 24, 2012.

Longley, P. A., M. F. Goodchild, D. J. Maguire, and D. W. Rhind (2010), *Geographic Information System and Science*, 3rd ed., John Wiley & Sons, Chichester, England.

Marcot, B. G., P. A. Hohenlohe, S. Morey, R. Holmes, R. Molina, M. C. Turley, M. H. Huff, J. A. Laurence, and F. S. USDA, Characterizing Species at Risk II: Using Bayesian Belief Networks as Decision Support Tools to Determine Species Conservation Categories Under the Northwest Forest Plan, *Ecology and society a journal of integrative science for resilience and sustainability*(2), http://hdl.handle.net/10113/29211, 2006.

Mascherini, M., F. Frascati, and F. M. Stefanini, MASTINO: Learning Bayesian Networks Using R, *Proceeding of COMPSTAT 2008 – International Conference on Computational Statistics*, 2008.

McCloskey, J. T., R. J. Lilieholm, and C. Cronan, Using Bayesian belief networks to identify potential compatibilities and conflicts between development and landscape conservation, *Landscape and Urban Planning, 101*(2), 190–203, http://www.sciencedirect.com/science/article/pii/S0169204611000697, 2011.

Müller, R., D. Müller, F. Schierhorn, G. Gerold, and P. Pacheco, Proximate causes of deforestation in the Bolivian lowlands: an analysis of spatial dynamics, *Regional Environmental Change*, 1–15, in press.

Nash, D., and M. Hannah (2011), Using Monte-Carlo simulations and Bayesian Networks to quantify and demonstrate the impact of fertiliser best management practices, *Environmental Modelling & Software, 26*(9), 1079–1088, http://www.sciencedirect.com/science/article/pii/S1364815211000843.

Neapolitan, R. E. , *Learning Bayesian Networks*, Prentice Hall, Chicago, Illinois, 2004.

Pearl, J. , *Probabilistic Reasoning in Intelligent Systems. Networks of Plausible Inference*, Morgan Kaufman Publishers, San Francisco, 1988.

Pontius Jr, R. G., and J. Spencer, Uncertainty in extrapolations of predictive land-change models, *Environment and Planning B: Planning and Design*(32), 211–230, DOI:10.1068/b31152, 2005.

Pontius, R., and N. Neeti, Uncertainty in the difference between maps of future land change scenarios, *Sustainability Science, 5*(1), 39–50, 2010.

Refsgaard, J. C., J. P. van der Sluijs, A. L. Højberg, and P. A. Vanrolleghem, Uncertainty in the environmental modelling process – A framework and guidance, *Environmental Modelling & Software, 22*(11), 1543–1556, http://www.sciencedirect.com/science/article/pii/S1364815207000266, 2007.

Stassopoulou, A., M. Petrou, and J. Kittler, Application of a Bayesian network in a GIS based decision making system, *International Journal of Geographical Information Science, 12*(1), 23–46, 1998.

van der Sluijs, J. P., P. H. M. Janssen, A. C. Petersen, P. Kloprogge, J. S. Risbey, W. Tuinstra, and J. R. Ravetz, *RIVM/MNP Guidance for Uncertainty Assessment and Communication. Tool Catalogue for Uncertainty Assessment*, Copernicus Institute for Sustainable Development and Innovation, Utrecht/Bilthoven, 2004.

Walker, R. , The scale of forest transition: Amazonia and the Atlantic forests of Brazil. Environmental Kuznets Curves and Environment-Development Research, *Applied Geography, 32*(1), 12–20, http://www.sciencedirect.com/science/article/pii/S0143622810001402, 2012.

Walker, W. E., P. Harremoës, J. Rotmans, J. P. van der Sluijs, M. B. A. van Asselt, P. Janssen, and M. P. von Krayer Krauss, Defining Uncertainty: A Conceptual Basis for Uncertainty Management in Model-Based Decision Support. Integrated Assessment, *Integrated Assessment, 4*(1), 5–17, 2003.

Warmink, J. J., J. A. E. B. Janssen, M. J. Booij, and M. S. Krol, Identification and classification of uncertainties in the application of environmental models, *Environmental Modelling & Software, 25*(12), 1518–1527, http://www.sciencedirect.com/science/article/pii/S1364815210001179, 2010.