Environmental Research 135 (2014) 148–155

Contents lists available at ScienceDirect

Environmental Research

journal homepage: www.elsevier.com/locate/envres

Impact of input data uncertainty on environmental exposure assessment models: A case study for electromagnetic field modelling from mobile phone base stations

Johan Beekhuizen [a], Gerard B.M. Heuvelink [b], Anke Huss [a], Alfred Bürgi [c], Hans Kromhout [a], Roel Vermeulen [a,*]

[a] Institute for Risk Assessment Sciences (IRAS), Division Environmental Epidemiology, Utrecht University, Yalelaan 2, 3584 CM, Utrecht, The Netherlands
[b] Soil Geography and Landscape, Environmental Sciences Group, Wageningen University, PO Box 47, 6700 AA Wageningen, The Netherlands
[c] ARIAS umwelt.forschung.beratung, CH-3011 Bern, Switzerland

ARTICLE INFO

Article history:
Received 2 April 2014
Received in revised form
22 May 2014
Accepted 26 May 2014

Keywords:
Uncertainty analysis
Monte Carlo simulation
Electromagnetic fields
Exposure assessment
Mobile phone base stations

ABSTRACT

Background: With the increased availability of spatial data and computing power, spatial prediction approaches have become a standard tool for exposure assessment in environmental epidemiology. However, such models are largely dependent on accurate input data. Uncertainties in the input data can therefore have a large effect on model predictions, but are rarely quantified.
Methods: With Monte Carlo simulation we assessed the effect of input uncertainty on the prediction of radio-frequency electromagnetic fields (RF-EMF) from mobile phone base stations at 252 receptor sites in Amsterdam, The Netherlands. The impact on ranking and classification was determined by computing the Spearman correlations and weighted Cohen's Kappas (based on tertiles of the RF-EMF exposure distribution) between modelled values and RF-EMF measurements performed at the receptor sites.
Results: The uncertainty in modelled RF-EMF levels was large with a median coefficient of variation of 1.5. Uncertainty in receptor site height, building damping and building height contributed most to model output uncertainty. For exposure ranking and classification, the heights of buildings and receptor sites were the most important sources of uncertainty, followed by building damping, antenna- and site location. Uncertainty in antenna power, tilt, height and direction had a smaller impact on model performance.
Conclusions: We quantified the effect of input data uncertainty on the prediction accuracy of an RF-EMF environmental exposure model, thereby identifying the most important sources of uncertainty and estimating the total uncertainty stemming from potential errors in the input data. This approach can be used to optimize the model and better interpret model output.

1. Introduction

An important part of epidemiological studies is the exposure assessment of the population to environmental pollutant(s). Traditionally, routine and purpose-designed monitoring have been used to determine exposure levels (Briggs et al., 2009), but with the increase in geographical data and modelling techniques, it has become feasible to assess exposure for large populations using spatial prediction models. Examples can be found in the modelling of air pollution (Beelen et al., 2013; Jerrett et al., 2005), noise (Murphy et al., 2009; Gan et al., 2012) and radio frequency electromagnetic fields (RF-EMF) (Bürgi et al., 2008; Beekhuizen et al., 2013).

However, even the most advanced models are by definition simplified representations of reality and cannot predict true exposure levels with 100% accuracy and precision. Spatial exposure models consider the source of the pollutant and surroundings to determine the strength and propagation of the exposure through the environment. These models need accurate geographical as well as source-related input data, as uncertainty in the inputs can have a large effect on the predictions (McIntyre et al., 2002; Jacquemin et al., 2013; Ramsey, 2009). For an overview of uncertainty analysis methods in exposure and risk assessment, we refer to review papers by Ramsey (2009) and Briggs et al. (2009) and more recently by Mesa-Frias et al. (2013).

Even though the effect of uncertainty in exposure modelling is recognized by the research community as an important issue (Briggs et al., 2009), uncertainty analyses are rarely applied in environmental epidemiological studies. Uncertainty in environmental exposure models can be assessed with Monte Carlo simulation

* Corresponding author. Fax: +31 30 2539499.
E-mail address: r.c.h.vermeulen@uu.nl (R. Vermeulen).

http://dx.doi.org/10.1016/j.envres.2014.05.038
0013-9351/

(Hammersley and Handscomb 1964). In Monte Carlo simulation, the uncertainty in model outputs (i.e. the exposure estimates) is derived by repeatedly sampling from probability distributions of the model inputs. Monte Carlo simulation is easy to implement, applicable to a wide range of models and results in an entire probability distribution of the model output (Heuvelink, 1998).

There is a need for accurate exposure assessment of RF-EMF from mobile phone base stations for epidemiological studies (IARC, 2013; HPA, 2012), as crude proxies of exposure are not sufficient to prove or dismiss associations between RF-EMF and health effects. The input data requirements for RF-EMF prediction models are demanding, as detailed data on all mobile phone base stations and buildings in the study area are essential inputs. Furthermore, as RF-EMF levels can change rapidly at distances less than a few metres, accurate positional information of the receptor site (e.g. the bedroom of participants in an epidemiological study) is needed. Therefore, errors in input data of RF-EMF models and receptor locations can greatly affect model accuracy. As we do not know the actual error, we assessed the uncertainty in RF-EMF modelling. It is important to make this distinction, as error and uncertainty are often used interchangeably. Error is the difference between the true value and the modelled value, while uncertainty is the acknowledgement of error by characterising the range of values within which the true value is asserted to lie (Heuvelink et al., 2007).

We computed the total uncertainty in RF-EMF modelling caused by uncertainty in model inputs using Monte Carlo simulation, and assessed which inputs have the largest influence on the model's ability to classify exposure by comparing model predictions with measurements performed in Amsterdam, The Netherlands. The uncertainty analysis identifies input data sources that should be prioritized for improving model performance and the total prediction uncertainty caused by model inputs, indicating the reliability and potential range of RF-EMF levels at all receptor sites.

2. Methods

2.1. Model description

We applied the NISMap model (Bürgi et al., 2008) to predict total downlink RF-EMF exposure from mobile phone base stations. NISMap has been used previously in studies in Switzerland and The Netherlands (Bürgi et al., 2010; Beekhuizen et al., 2014a; Beekhuizen et al., 2013) and we refer to these studies for a detailed model description. In short, NISMap is a spatial radio wave propagation model that calculates the spread of radio waves from antennas in 3D space. NISMap considers the properties of the source of radio waves (antenna location, height, direction, tilt, output power and pattern), terrain height, the (simplified) physics of the radio wave propagation (using a radio wave propagation algorithm, such as the Double Powerlaw (ITU, 2009) and damping and diffraction of the wave by buildings.

2.2. Input data and validation measurements

We studied uncertainty in RF-EMF modelling for the urban area of Amsterdam, The Netherlands, using measurement and input data from an indoor model validation study (Beekhuizen et al., 2014a). Antenna data were obtained from the three Dutch mobile phone network operators, 3D building data from Eurosense (Wemmel, Belgium) and the Dutch national Digital Terrain Model (DTM) (Actueel Hoogtebestand Nederland, AHN2). The total downlink RF-EMF from mobile phone base stations was measured with an EME-SPY 140 device (SATIMO, France). We collected 15-minute spot measurements in 263 rooms in 101 primary schools and 30 private homes in Amsterdam. We left 11 sites out of the analysis as they were located at the outer edge of the city, which was not completely surrounded by the original Amsterdam building data for which we built an uncertainty model.

2.3. Monte Carlo uncertainty propagation analysis

The first step in a Monte Carlo analysis is determining the probability distribution function (PDF) of the model inputs. Next, the model is run for N samples randomly drawn from these PDFs. Finally, the model uncertainty can be inferred from the statistical evaluation and visualisation of the simulated model outputs. The minimum number of necessary model runs was determined by plotting the standard deviations of a batch of N simulations at each site against the standard deviations of a second batch with as many model runs. If the resulting uncertainties at each site for one batch of model runs are similar to those of the second, the graph will show a high correlation, indicating that a sufficiently large number of model runs has been used (Nol et al., 2010). For $N=200$ simulations, we found an acceptable R^2 of 0.96 between the two sets. The Appendix shows a scatterplot of the standard deviations of both sets.

2.4. Uncertainty in RF-EMF model inputs

To quantify uncertainty in input parameters we used (i) reference data to assess errors in the data, (ii) information on uncertainty in the antenna input parameters obtained through consultation with an expert on network planning and an expert on network optimization from one of the Dutch network providers (July 2013), and (iii) information on all uncertainty PDFs obtained through consultation at an expert elicitation meeting with Alfred Bürgi (August 2013), the developer of NISMap and an expert on its properties and inputs. The following uncertainty sources and PDFs were selected:

2.4.1. Building height

The uncertainty in building height was estimated by computing differences between heights in the Eurosense building data set with heights from an unfiltered version of the high-resolution (0.5 by 0.5 m) DTM (AHN2) for three 1×1 km^2 areas in Amsterdam, containing over 900 buildings. The errors were truncated at -5 and 5 m (2.3% of the errors), as errors > 5 m are likely caused by temporal changes in building data between AHN2 and Eurosense, and thus do not indicate the height error in our building dataset. Fig. 1 shows the obtained empirical error distribution, from which we sampled building height errors. If the simulated building height was smaller than 2 m, we set the building height to 2 m (the minimum realistic building height). We assumed that errors in building height are spatially correlated; i.e. we decided that it is unrealistic to raise one building and lower a neighbouring building with the same height and building age. We therefore created building clusters based on building height, age and proximity to other buildings using the ArcGIS 10.2 Grouping Analysis tool. Each building cluster is a group of buildings that have been built in a similar time period, have a similar height and are located nearby each other. There were on average 20 buildings per cluster, but (depending on the cluster-criteria) the cluster size could range from a few to hundreds of buildings.

2.4.2. Building damping

The degree of building damping can be set to different values for wall, roof and inside. As the damping factor is dependent on many factors, such as wall material, thickness, and incidence angle of the radio wave, the damping differs for each building. However, NISMap has the limitation that damping factors can only be set to one value for all buildings, restricting the simulation of damping values. Typical Dutch home and school buildings are mostly made from bricks, thus we based our damping parameters on values for bricks (BUWAL, 2002).

With the exception of high-performance heat-insulating windows, normal windows are virtually transparent for RF-EMF. Therefore, the actual damping of walls can be distinctly smaller; if the antenna is in direct line of sight of a room with a large window, then there is little building damping. There is less variation for roof damping as there are typically no windows on the roof. The inside damping factor is used to predict the damping of the radio waves due to obstacles indoors, such as inside walls and furniture, and can therefore greatly vary between buildings. We chose a normal

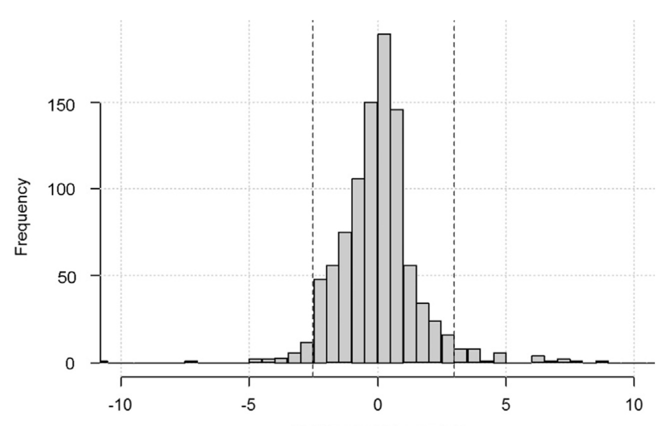

Fig. 1. Differences between building heights from a reference height data set (AHN2) and building height dataset used in our modelling (Eurosense), for a total of ~900 buildings in three 1×1 km^2 areas in Amsterdam, The Netherlands. The black striped lines represent the 2.5 and 97.5 percentiles [-2.51 m; 2.98 m].