

A Bayesian Method for Model Selection in Environmental Noise Prediction

L. Martín-Fernández^{1,*}, D. P. Ruiz¹, A. J. Torija², and J. Míguez³

¹*Department of Applied Physics, University of Granada, Avda. Fuentenueva s/n, Granada 18071, Spain*

²*Institute of Sound and Vibration Research, University of Southampton, Highfield, Southampton SO17 1BJ, UK*

³*Department of Signal Theory & Communications, Carlos III University of Madrid, Avda. de la Universidad 30, Leganés, Madrid 28911, Spain*

Received 6 August 2013; revised 28 May 2014; accepted 3 August 2014; published online 11 March 2016

ABSTRACT. Environmental noise prediction and modeling are key factors for addressing a proper planning and management of urban sound environments. In this paper we propose a maximum a posteriori (MAP) method to compare nonlinear state-space models that describe the problem of predicting environmental sound levels. The numerical implementation of this method is based on particle filtering and we use a Markov chain Monte Carlo technique to improve the resampling step. In order to demonstrate the validity of the proposed approach for this particular problem, we have conducted a set of experiments where two prediction models are quantitatively compared using real noise measurement data collected in different urban areas.

Keywords: environmental noise level prediction, MAP model selection, Monte Carlo sampling, nonlinear state-space model, particle filtering

1. Introduction

Environmental noise is one of the most important environmental problems in urban areas, since it has a great impact on the health and welfare of the exposed population (Laszlo et al., 2012). In urban agglomerations, road traffic is the main noise source and its effects on inhabitants are well known (Ouis, 2001; Dai et al., 2005). Several research works (e.g., Bjorkman, 1991; Lercher, 1996) have shown that noise affects daily activities and causes sleep disturbances. Therefore, in order to care for human health and provide dwellers with a high quality of life, urban spaces should be planned with an appropriate sound environment, keeping noise levels under control.

The accurate characterization of the sound environment is essential for urban planning. In urban areas there is a large variety of sound sources and conditions that, in turn, generate a wide variety of relevant acoustic situations. Thus, we can find locations with different composition of traffic, different urban settings, with presence of noise sources that are difficult to characterize (e.g., business or leisure areas, works), the existence of green areas, etc., that create sound spaces with different sound pressure levels and with large differences in their temporal evolution and spectral composition.

Accordingly, the complexity of urban agglomerations ma-

kes the environmental noise modeling and prediction a complex and nonlinear problem. Previous research has shown the influence of different variables on the characterization of the environmental noise level and in the description of the temporal evolution of the sound pressure level (Torija et al., 2007a; Torija et al., 2007b; Torija et al., 2007c; Torija et al., 2010; Sachakamol et al., 2011). In the literature, there are different methods and techniques that solve the problem of predicting environmental noise levels, from physical models to statistical models. For example, models based on physical principles of sound generation and propagation, as Harmonoise/Imagine model (Watts, 2005), Calculation of Road Traffic Noise (CORTN) (Anon., 1975), Nord2000 (Kragh et al., 2002), etc., have an outstanding performance in estimating sound pressure levels. These models are mainly used for predicting long-term sound pressure levels from road traffic but urban environments are characterized by the presence of other sound sources different than road traffic, such as leisure noise, commercial activities, etc., which have a great influence in the generation of environmental noise in urban agglomerations. In Josse (1972), Burgess (1977) and Bertoni et al. (1987) it is proposed to use statistical models. Although these models can describe nonlinear correlations, they do not provide an accurate enough approximation of the trend followed by the sound pressure level when this is affected by a large number of physical parameters. Later on, in Cammarata et al. (1993), it is proposed to apply an artificial neural network (ANN) for noise prediction. The method involves the training of a back-propagation network (BPN) (McClelland and Rumelhart, 1988) using an appropriate set of acoustic measurements and, in the subsequent phase, the network predicts the sound pressure level for various inputs. This method achieves good re-

* Corresponding author. Tel.: +349 58244161; fax: + 349 58243214.

E-mail address: lauramartin@ugr.es (L. Martín-Fernández).

sults because neural networks have a great capacity for approximating functions which are essentially nonlinear (Lapedes and Farber, 1987; Suykens et al., 1996), as it is the environmental noise prediction problem. From this point of view, other authors have developed complex neural networks with the objective of providing a tool for the design, planning and evaluation of urban sound environments and the ultimate goal of incorporating the needs of the population into the planning of urban agglomerations (Cammarata et al., 1995; Genaro et al., 2010; Torija et al., 2012).

Since several models may be available to describe the sound environment in a given urban area, the question of how to choose the fittest model given a record of data arises naturally. In this paper, we study dynamic models that can be put in a state-space form. In particular, we identify the sources of noise as the state variables of the model and allow them to evolve randomly over time. The observations, or measurements, for the model are indicators of the overall sound pressure level and 1/3-octave band sound levels (spectral composition). The relationship between the observations (sound pressure levels) and the state variables (noise sources) is represented by nonlinearities (such as, e.g., different neural network configurations (McClelland and Rumelhart, 1988)) and a random perturbation.

Given two state-space models, each one with a different nonlinear structure describing the relationship between the indicators of sound pressure level and the noise sources, our goal is to quantify the fitness of each model to predict environmental noise levels using a collection of real data sets and select the most suitable candidate. Following the general approach in Djuric (1998), we propose to score the competing models by way of their posterior probabilities conditional on the same data sets, i.e., we carry out maximum *a posteriori* (MAP) model selection. Our approach involves the computation of the evidence (as defined in MacKay (2003), Chapter 3) in favor of each one of the competing models. However, since the models of interest are dynamic and nonlinear, these evidences cannot be found in closed form. To circumvent this difficulty, we introduce a numerical approximation method based on the use of particle filters (Gordon et al., 1993; Doucet et al., 2000; Doucet et al., 2001; Djuric et al., 2003), similar to the model monitoring algorithm of Djuric (1999). The proposed filtering algorithm includes Markov Chain Monte Carlo (MCMC) moves (Gilks and Berzuini, 2001) to mitigate the diversity loss that follows the resampling step in conventional particle filters. It should also be noted that the proposed method can be stated in terms of Bayes factors (as defined in Bernardo and Smith (2009); see Chapter 6). In particular, when only two models are compared, MAP selection as described in this paper corresponds to the Bayesian test in Proposition 6.1 of Bernardo and Smith (2009). Unlike Bayes factors, though, the MAP scheme can also be applied in a straightforward manner when more than two models are competing.

This paper is organized as follows. In Section 2 we define the state-space models to be compared. In Section 3 we elaborate on the MAP criterion for model selection. The particle filtering algorithm applied for the numerical implementa-

Table 1. State Variables*

State Variable	Value Range
1 Time Period	Day / Evening
2 Commercial or Leisure Environment	No / Yes
3 Construction Work	No / Yes
4 Stabilization Time	[2-55] (minutes)
5 Average Speed	[5.38-52.12] (km/h)
6 Traffic Slope	[0-9] (%)
7 Number of Ascendant Lanes	[0-4] (lanes)
8 Number of Descendant Lanes	[0-4] (lanes)
9 Pavement Type	Porous asphalt / Smooth asphalt / Paved
10 Pavement Surface Condition	Good / Fair / Bad / Very Bad
11 Street Geometry	Type "U" / Type "J" / Type "L" / Type "Free Field"
12 Street Width	[3.8-104.67] (m)
13 Street Height	[0-32.55] (m)
14 Roadway Width	[3.8-23.42] (m)
15 Source-Receptor Distance	[2.6-16.7] (m)
16 Type of Traffic Flow	Constant fluid flow / Constant pulsed flow / Flow decelerated in pulses / Flow accelerated in pulses / Intermittent flow / Banked flow
17 Number of Vehicles with Sirens	[0-1] (vehicles)
18 Impulsive Sound Event related to Traffic	No / Yes
19 Impulsive Sound Event unrelated to Traffic	No / Yes
20 Ascendant Flow of Light Vehicles	[0-20]
21 Descendant Flow of Light Vehicles	[0-38]
22 Ascendant Flow of Heavy Vehicles	[0-10]
23 Descendant Flow of Heavy Vehicles	[0-4]
24 Ascendant Flow of Motorcycles	[0-20]
25 Descendant Flow of Motorcycles	[0-14]

*The stabilization time (variable 4), at a certain location, is the time needed to stabilize the sound pressure level within a previously defined range (Torija et al., 2011). The street geometry classification (variable 11) is taken from (NMPB, 1996) and refers to the type of buildings on the street sides, e.g., a "J"-type street contains tall buildings on one side and low buildings on the other side. The traffic flow magnitudes (variables 20-25) have the form $10\log z$, where z is the number of vehicles over one lane (either ascendant or descendant) every two minutes.

tion is described in Section 4. In Section 5 we compare the proposed MAP model selection scheme with other Bayesian techniques for model selection. In Section 6 we test the proposed methodology by using a series of measurements of sound pressure levels obtained experimentally in the city of Granada (Spain). The obtained results are shown and discussed here. Finally, the article ends with a summary and some conclusions in Section 7.

2. Models

2.1. Sound Pressure Levels

Let us consider the problem of predicting the value of 23 descriptors of sound pressure level in urban areas from the

Table 2. Characteristics of the Studied Locations (Streets)

	Street	Time Period	Description
1	Camino de Ronda	21:30-22:30	Broad street with a narrow central reservation. Geometry type “U”. High flow of heavy vehicles.
2	Gran Vía	14:20-15:30	Geometry type “U”. High flow of heavy vehicles, light vehicles, and motorcycles. Commercial area.
3	Avenida de Murcia	10:45-11:35	Geometry type “U”. Great ascendant traffic slope. High flow of heavy vehicles.
4	Méndez Núñez	21:30-22:40	Geometry type “Free Field”. High traffic flow. University area.
5	Camino de las Vacas	9:40-10:50	Geometry type “Free Field”. High flow of heavy vehicles.
6	Nueva del Santísimo	20:40-21:30	Narrow street with geometry type “U”. Low traffic flow. Commercial area.
7	Reyes Católicos	20:15-20:55	Geometry type “U”. High traffic flow. Commercial area.
8	Doctor Olóriz	20:30-21:00	Geometry type “J”. Opposite the bullring and near hospitals. Leisure zone. Descendant traffic slope.
9	Real de la Cartuja	12:30-12:55	Geometry type “U”. Great ascendant traffic slope. Pavement type “Paved”.
10	Gran Capitán	11:40-12:20	Geometry type “U”. Pavement type “Porous asphalt”.
11	Gonzalo Gallas	16:55-18:05	Geometry type “L”. Leisure zone.

knowledge of a set of noise sources. To be specific, the indicators of interest are the A-weighted equivalent continuous sound pressure level, L_{Aeq} (57-85 dBA), the non-weighted equivalent continuous sound pressure level, L_{eq} (65-90 dB), and the sound level in 1/3 octave bands from 40 Hz to 4 kHz, L_f (10-80 dB), where $f = 40, 50, 63, 80, 100, 125, 160, 200, 250, 315, 400, 500, 630, 800, 1000, 1250, 1600, 2000, 2500, 3150$ and 4000 Hz. These magnitudes have been obtained from field measurements with S integration time ($S = 2$ minutes).

2.2. State Variables

A precise characterization of the urban sound environment requires the consideration of a wide range of different magnitudes related to the sound emission generated by the present sources of noise as well as the sound propagation in diverse geometrical configurations. In this paper, we collectively refer to these magnitudes as state variables. Specifically, we study 25 variables displayed in Table 1: 20 sound emission variables (entries 1-10 and 16-25 in Table 1) and 5 sound propagation variables (entries 11-15 in Table 1). The pavement type (entry 9 in Table 1) is also related to sound propagation because sound is propagated differently in paved surfaces (sound reflection) and porous surfaces (sound absorption) (Lui and Li, 2004). The selection and classification of the state variables is taken from Torija et al. (2010). Note that most of state variables are static over time (entries 1-15 in Table 1) except the type and magnitude of the traffic flows, the number of vehicles with sirens and impulsive sound events (Torija et al., 2011) (entries 16-25 in Table 1), which can change significantly depending on both time and location.

2.3 Nonlinear Prediction of Sound Pressure Levels

The prediction of sound pressure levels from noise sources can be carried out by different methods, e.g., physical models (Anon, 1975; Kragh et al., 2002; Watts, 2005) or statistical models (Josse, 1972; Burgess, 1977; Bertoni et al., 1987). However, since the relationship between the sound pressure level and the noise sources is highly nonlinear, neural net-

works have been advocated as efficient tools in urban agglomerations in Cammarata et al. (1995), Genaro et al. (2010) and Torija et al. (2012) because of their ability to approximate nonlinear functions (studied in, e.g., Lapedes and Farber (1987) and Suykens et al. (1996)).

In this paper, as prediction functions of the candidate models, we use two backpropagation networks (McClelland and Rumelhart, 1988) with different configurations (named Configuration 2 and Configuration 4 in Torija et al. (2012)). Nevertheless, note that the proposed method could be applied to compare dynamical models based on other types of prediction functions. We have chosen these network configurations due to their similarity and the difficulty of making an election *a priori* based on their goodness to fit data in different locations.

In particular, the studied neural networks have a common structure, i.e., 2 layers with 25 inputs related to the state variables, 23 neurons in the hidden layers and 23 outputs related to the sound pressure levels. Moreover, both nets use a hyperbolic tangent sigmoid transfer function (first layer) and a linear transfer function (second layer). However, they were calibrated with different training functions. Both functions update the weight and bias values according to the Levenberg-Marquardt optimization method (Marquardt, 1963; Hagan and Menhaj, 1994) avoiding the nets overfit the experimental data, but the training of the first backpropagation network (network 1, in the sequel) uses a Bayesian regularization process (MacKay, 1992; Foresee and Hagan, 1997) that is not applied when training the second network (network 2, in the sequel). The aim of this regularization algorithm is to improve the generalization capability of the network and, then, obtain good results given new input data. In Torija et al. (2012), the network configuration 1 slightly outperformed the network configuration 2.

Both networks were trained using the same database, which covers the heterogeneity of the city of Granada (Spain), a typical Southern Europe medium-sized city. The experimental measurements, 274 records in total, were collected on working days, at different time periods, in different urban set-

tings and different traffic conditions by researchers of the laboratory of Physics and Environmental Acoustics (Department of Applied Physics, University of Granada). The main features of each location are outlined in Table 2.

The measurement of environmental noise was carried out in slots of 30-90 minutes, with each experimental datum corresponding to a record of 2 minutes. According to that, in each collected time series, the first datum of, e.g. the “ascendant flow of light vehicles” (row 20 in Table 1) corresponds to the number of vehicles observed during the first two minutes, the second datum corresponds to the following two minutes, etc. For binary data, e.g., the “impulsive sound event” of row 18 in Table 1, the first datum indicates whether at least one impulsive event was observed during minutes 1 and 2, the second datum corresponds to minutes 3 and 4, and so on. Note that the short two-minute period was chosen to ease the observation of the short-term variability of the sound pressure level for each location. The measurements were obtained following international procedures of reference; all microphones were mounted away from reflecting facades, at a height of 4 meters above local ground level (Directive 2002/49/EC, 2002).

2.4. State-Space Models

As explained in Section 2.2, several state variables of the problem (entries 16-25 in Table 1) are time-varying. Therefore, any model aimed to predict the values of the sound pressure levels from these variables (e.g., traffic flows) should take into account their dynamics in order to produce adequate results. A compact manner to jointly represent the state variable dynamics and the nonlinear relationship between the variables and the indicators of sound pressure level is by way of a state-space model.

To be specific, in each one of our candidate state-space model, let:

$$x_t = (x_{1,t}, x_{2,t}, \dots, x_{10,t}), \quad t = 0, 1, \dots, T \quad (1)$$

be the vector of time-varying state variables (i.e., the type and magnitude of the traffic flows, the number of vehicles with sirens and impulsive sound events), where $x_{i,t}$, $i = 1, \dots, 10$ corresponds to the $(15+i)$ -th entry of Table 1 (i.e., $x_{1,t}$ is the “type of traffic flow” and $x_{10,t}$ is the “descendant flow of motorcycles”) and T is the number of time steps. We collect the static variables (related to the period of day, the geometrical configuration, the number of lanes, etc.) in a vector:

$$\theta = (\theta_1, \theta_2, \dots, \theta_{15}) \quad (2)$$

where $\theta_1, \theta_2, \dots, \theta_{15}$ correspond to entries 1-15 in Table 1.

The indicators of sound pressure level and its spectral composition are represented by a 23 dimensional vector:

$$y_t = (y_{1,t}, y_{2,t}, \dots, y_{23,t}) \quad (3)$$

where $y_{1,t}$ represents the value of the L_{Aeq} indicator at time t ,

$y_{2,t}$ represents the L_{eq} indicator at time t and $y_{3,t}, \dots, y_{23,t}$ stand for the remaining indicators in 1/3 octave bands from 40 Hz to 4 kHz (see Section 2.1).

The dynamics of x_t is independent of the static variables in θ , but the sound pressure levels in y_t are modeled as a function of x_t , θ and the prediction method (in this case, a neural network), plus a random perturbation. In particular, we have two state-space models of the form (Note: We use p to denote probability functions, including densities and masses. The notation is argument-wise. For example, if x and y are continuous random variables, then $p(x)$ and $p(y)$ denote their probability density functions, possibly different. If x is a discrete random variable, then $p(x)$ denotes its probability mass function. Conditional densities and masses are indicated in the obvious way, e.g., $p(x|y)$). This notation is common in Bayesian analysis and the particle filtering literature.):

$$x_t \sim p(x_t | x_{t-1}) \quad (4)$$

$$y_t \sim p(y_t | x_t, \theta, m) \quad (5)$$

where $m \in \{1, 2\}$ is the model index, $p(x_t | x_{t-1})$ describes the conditional distribution of the dynamic variables (in particular, we assume that every $x_{i,t}$ is first order Markovian) and $p(y_t | x_t, \theta, m)$ is the conditional density of the sound pressure levels given the state variables and the neural network used for prediction (either network 1 or network 2, as indicated by m).

The conditional density of the sound pressure levels, often referred to as the likelihood of the variables, is assumed to be Gaussian, namely:

$$p(y_t | x_t, \theta, m) = N(y_t; g_m(x_t, \theta), D_m) \quad (6)$$

where $N(z; \mu, C)$ denotes the (multivariate) Gaussian probability density function (pdf) of z with mean vector μ and covariance matrix C , $g_m(x_t, \theta)$ represents the sound pressure levels predicted by the m -th neural network and:

$$D_m = \begin{bmatrix} \sigma_{m,1}^2 & 0 & L & 0 \\ 0 & \sigma_{m,2}^2 & L & 0 \\ M & M & O & M \\ 0 & 0 & L & \sigma_{m,23}^2 \end{bmatrix}$$

is a diagonal covariance matrix with the marginal variances $\sigma_{m,j}^2$, $j = 1, 2, \dots, 23$ fitted using the experimental data and the predictions from the m -th neural network. The Gaussian model is not a limitation of the method because it is straightforward to incorporate other distributions for modeling the likelihood of state variables (e.g., Gaussian mixture models (Don and Rees, 1985) or impulsive models (Garcia and Faus, 1991)).

The dynamic states $x_{1,i}$, $x_{2,i}$, $x_{3,i}$ and $x_{4,i}$ (type of traffic flow, number of vehicles with sirens and impulsive sound events, respectively) are discrete first-order Markov chains with transition probabilities given by the matrices:

$$P_1 = \begin{bmatrix} 0.23 & 0.30 & 0.05 & 0.11 & 0.12 & 0.19 \\ 0.05 & 0.34 & 0.17 & 0.16 & 0.28 & 0 \\ 0 & 0.46 & 0.17 & 0.37 & 0 & 0 \\ 0.19 & 0.37 & 0.08 & 0.31 & 0.05 & 0 \\ 0.03 & 0.10 & 0 & 0.07 & 0.55 & 0.25 \\ 0.19 & 0 & 0 & 0 & 0.25 & 0.56 \end{bmatrix}$$

$$P_2 = \begin{bmatrix} 0.99 & 0.01 \\ 1 & 0 \end{bmatrix}$$

$$P_3 = \begin{bmatrix} 0.42 & 0.58 \\ 0.53 & 0.47 \end{bmatrix} \text{ and } P_4 = \begin{bmatrix} 0.77 & 0.23 \\ 0.72 & 0.28 \end{bmatrix}$$

respectively (Note: We assume that a discrete variable $x_{i,t}$ can take the values described in the value-range column of Table 1. Let x_i^j be the j -th value in the range of variable $x_{i,t}$. The entry in the k -th row, l -th column of P_i , denoted $P_i^{k,l}$, is the probability mass $p(x_{i,t} = x_i^k | x_{i,t-1} = x_i^l)$. For the continuous variables (traffic flows), $x_{5,t}, \dots, x_{10,t}$, we have chosen a simple random walk model of the form:

$$p(x_{i,t} | x_{i,t-1}) = N(x_{i,t}; x_{i,t-1}, \alpha_i^2), \quad i = 5, \dots, 10 \quad (7)$$

where α_i^2 is a variance fitted using the collected data. We assume that the state variables evolve independently over time, i.e., $p(x_i | x_{i-1}) = \prod_{i=1}^{10} p(x_{i,t} | x_{i,t-1})$.

These two models have been chosen because they show a good and similar behavior in predicting sound pressure levels. For urban planning, the selection of the model with the best performance is important because, even if both models have similar accuracy, the difference between predictions is not trivial in Environmental Acoustics and can imply a substantial increase in the annoyance of the inhabitants or in the quantity of affected population.

3. Criterion for Model Selection

The aim of this paper is to formally show how, following a general Bayesian approach presented in Djuric (1998), it is possible to score, and then select, different state-space models to predict the indicators of sound pressure level. We have explicitly defined two models in Section 2.4, but the criterion (and procedure) to be introduced can be easily extended to an arbitrary number of candidate models and/or different models (i.e., with other model variables, different types of nonlinear prediction functions, etc.).

Let the index $m \in \{1, 2\}$ refer to the state-space model. We assign a prior probability $p(m)$ to each model m and, given a fixed record of experimental data $y_{1:T} \in \{y_1, y_2, \dots, y_T\}$, we use $p(m | y_{1:T}, \theta)$ to denote the *a posteriori* probability mass of the model m given the data and the static variables in θ . For the rest of this paper, we assume θ is known and leave the dependences on this fixed vector implicit. Therefore, we write $p(m | y_{1:T})$ for the posterior probability of the model m and

$p(y_{1:T} | m) = p(y_{1:T} | m, \theta)$ for the associated likelihood.

A natural criterion to choose the model that best fits the data is to solve the problem:

$$\hat{m} = \arg \max_{m \in \{1, 2\}} p(m | y_{1:T}) \quad (8)$$

where \hat{m} is the maximum *a posteriori* (MAP) model choice. For calculating $p(m | y_{1:T})$, we apply the Bayes' theorem:

$$p(m | y_{1:T}) = \frac{p(y_{1:T} | m) p(m)}{p(y_{1:T})} \quad (9)$$

where $p(y_{1:T} | m)$ is the joint probability density of the observations $y_{1:T}$, conditional on the model m , $p(m)$ is the prior probability of m and $p(y_{1:T})$ is the unconditional probability density of $y_{1:T}$ (which is constant with respect to m). After applying the chain rule to Equation (9), we rewrite (8) as:

$$\hat{m} = \arg \max_{m \in \{1, 2\}} \prod_{k=1}^T p(y_k | y_{1:k-1}, m) p(m) \quad (10)$$

where $p(y_k | y_{1:k-1}, m)$ is the predictive probability density of the sound pressure level indicators at time k given the previous indicators and the model m , for $k = 1, 2, \dots, T$. In order to ease the calculations, let us take the logarithm of the product in (10), to obtain:

$$\begin{aligned} \hat{m} &= \arg \max_{m \in \{1, 2\}} \log \left[\prod_{k=1}^T p(y_k | y_{1:k-1}, m) p(m) \right] \\ &= \arg \max_{m \in \{1, 2\}} \left[\sum_{k=1}^T \log p(y_k | y_{1:k-1}, m) + \log p(m) \right] \end{aligned} \quad (11)$$

Notice that the optimization problems (8) and (11) are equivalent because $\log(z)$ is a monotonically increasing function. Finally, for notational convenience, we define the objective function:

$$J_m(T) = \sum_{k=1}^T \log p(y_k | y_{1:k-1}, m) + \log p(m) \quad (12)$$

and reduce the model selection problem to choosing the index \hat{m} that yields the highest value of $J_m(T)$, i.e.:

$$\hat{m} = \arg \max_{m \in \{1, 2\}} J_m(T) \quad (13)$$

In many cases, it may be of interest to compare the candidate models in terms of their ability to predict a scalar observation $y_{i,t}$, $i = 1, 2, \dots$, instead of the joint vector y_t (e.g., in order to determine the sensitivity of the model for a given sound pressure level). The derivation of the MAP criterion is exactly the same as before. Namely, for the i -th observation, we select:

$$\hat{m} = \arg \max_{m \in \{1, 2\}} J_m^i(T) \quad (14)$$

where:

$$J_m^i(T) = \sum_{k=1}^T \log p(y_{i,k} | y_{i,1:k-1}, m) + \log p(m) \quad (15)$$

The MAP selection method that we propose in this paper is summarized as:

Initialization.

- a. Obtain the prior probability mass function $p(m)$ for the indices $m \in \{1, 2\}$.
- b. Let $J_m(0) = \log p(m)$, $m \in \{1, 2\}$.

Sequential update. For $t = 1, 2, \dots, T$,

- a. Compute the predictive pdf $p(y_t | y_{1:t-1}, m)$.
- b. Update the objective function, $J_m(t) = J_m(t-1) + \log p(y_t | y_{1:t-1}, m)$

Selection. Choose the model $\hat{m} = \arg \max_{m \in \{1, 2\}} J_m(T)$.

4. Numerical Implementation

The model selection method introduced in Section 3 demands the evaluation of the sequence of posterior pdf's $p(y_t | y_{1:t-1}, m)$, $t = 1, 2, \dots, T$, for each model. Unfortunately, due to the nonlinearity $g_m(x_t, \theta)$ in Equation (6), there is no closed-form expression for these densities, which, therefore, have to be approximated numerically.

In this paper, we propose to carry out these approximations using particle filtering (Gordon et al., 1993; Doucet et al., 2000). Specifically, we apply a variation of the standard particle filter (Gordon et al., 1993; Doucet et al., 2001) that includes a Markov Chain Monte Carlo (MCMC) move after the resampling step. The aim of this move is to enhance the diversity of the particle set (Gilks and Berzuini, 2001). The specific MCMC technique we have used in this paper is a Metropolis-Hastings algorithm, as described in Chib and Greenberg (1995). The algorithm is described as follows:

Initialization. Let M be the number of particles. Draw $x_0^{(i)}$ from $p(x_0)$, $i = 1, 2, \dots, M$.

Recursive steps. For $t = 1, \dots, T$:

- a. Draw $\tilde{x}_t^{(i)}$ from $p(x_t | x_{t-1}^{(i)})$, $i = 1, 2, \dots, M$.
- b. Update the importance weights, $\tilde{w}_t^{(i)} = p(y_t | \tilde{x}_t^{(i)}, m)$, and normalize them, $w_t^{(i)} = \tilde{w}_t^{(i)} / \sum_{j=1}^M \tilde{w}_t^{(j)}$, $i = 1, 2, \dots, M$.
- c. Multinomial resampling: for $i = 1, 2, \dots, M$, let $\bar{x}_t^{(i)} = \tilde{x}_t^{(k)}$ with probability $w_t^{(k)}$, $k \in \{1, \dots, M\}$.
- d. MCMC move: take one step of the Metropolis-Hastings method to generate $x_t^{(i)}$ from $\bar{x}_t^{(i)}$, $i = 1, 2, \dots, M$.

Let δ_x denote the unit delta measure centered at x and consider the two sets of particles (samples in the state space) $\{\tilde{x}_t^{(i)}\}_{i=1}^M$ and $\{x_t^{(i)}\}_{i=1}^M$ generated by the particle filter at each time step t . We can construct two random measures:

$$\pi_M(dx_t | y_{1:t-1}, m) = \frac{1}{M} \sum_{i=1}^M \delta_{\tilde{x}_t^{(i)}}(dx_t) \quad \text{and}$$

$$\pi_M(dx_t | y_{1:t}, m) = \frac{1}{M} \sum_{i=1}^M \delta_{x_t^{(i)}}(dx_t),$$

that enable the approximation of integrals with respect to

$p(x_t | y_{1:t-1})$ and $p(x_t | y_{1:t})$, respectively. Indeed, for any integrable real function f in the state-space:

$$\int f(x_t) p(x_t | y_{1:t-1}, m) dx_t \approx \int f(x_t) \pi_M(dx_t | y_{1:t-1}, m)$$

$$= \frac{1}{M} \sum_{i=1}^M f(\tilde{x}_t^{(i)})$$

$$\int f(x_t) p(x_t | y_{1:t}, m) dx_t \approx \int f(x_t) \pi_M(dx_t | y_{1:t}, m)$$

$$= \frac{1}{M} \sum_{i=1}^M f(x_t^{(i)})$$

The asymptotic convergence (as $M \rightarrow \infty$) of such approximations can be guaranteed under mild regularity assumptions (Crisan, 2001; Moral, 2004; Bain and Crisan, 2008). In particular, we note that the predictive pdf's $p(y_t | y_{1:t-1}, m)$ can be written as integrals of the form:

$$p(y_t | y_{1:t-1}, m)$$

$$\approx \int p(y_t | x_t, m) p(x_t | y_{1:t-1}, m) dx_t, \quad t = 1, 2, \dots, T,$$

and hence we compute their approximations as:

$$p_M(y_t | y_{1:t-1}, m) \approx p(y_t | y_{1:t-1}, m),$$

where:

$$p_M(y_t | y_{1:t-1}, m) = \int p(y_t | x_t, m) \pi_M(dx_t | y_{1:t-1}, m)$$

$$= \frac{1}{M} \sum_{i=1}^M p(y_t | \tilde{x}_t^{(i)}) \tag{16}$$

We use Equation (16) for the computation of the objective functions $J_m(t)$, $1 \leq t \leq T$, i.e., we let:

$$J_m(t) \approx \sum_{k=1}^T \log(p_M(y_k | y_{1:k-1}, m)) + p(m) \tag{17}$$

in all the experimental results to be shown in Section 6.

5. Comparison with Other Bayesian Methods for Model Selection

5.1. Bayes Factors

Given two competing models, $m \in \{1, 2\}$, and a sequence of observations $y_{1:T}$, the Bayes factor in favor of the model $m = 1$ (and against $m = 2$) is the ratio of model likelihoods (Bernardo and Smith, 2009; Definition 6.1):

$$B_{12}(y_{1:T}) = \frac{p(y_{1:T} | m = 1)}{p(y_{1:T} | m = 2)}$$

with $B_{21}(y_{1:T}) = B_{12}^{-1}(y_{1:T})$. Bayes factors are naturally compared to the ratio of prior model probabilities weighted by the loss or penalty associated to decision errors (see Proposition

6.1 in Bernardo and Smith (2009)). In particular, let l_{ij} be the penalty due to selecting model $m = i$ when the “true” model is $m = j$. Then, model $m = 2$ should be preferred over $m = 1$ if, and only if:

$$B_{12}(y_{1:T}) < \frac{l_{12}}{l_{21}} \times \frac{p(m=2)}{p(m=1)} \quad (18)$$

However, if $l_{12} = l_{21}$, then it is apparent that the criterion (18) reduces to the MAP criterion in Equation (14), since $B_{12}(y_{1:T}) < p(m=2) / p(m=1)$ if and only if $p(m=1 | y_{1:T}) > p(m=2 | y_{1:T})$. When the penalties are not symmetric it is straightforward to redefine the priors for the competing models as $\tilde{p}(m=2) = l_{12} p(m=2)$ and then the proposed MAP criterion (with priors $\tilde{p}(m)$, $m \in \{1, 2\}$) is, again, equivalent to the criterion given by Equation (18).

While the Bayes factors may be a convenient tool for comparing two models, the MAP criterion can be applied in a direct manner to problems involving many competing models.

5.2. Deviance Information Criterion

The deviance information criterion (DIC) (Spiegelhalter et al., 2002) has become a popular tool for Bayesian model selection in the last decade. It relies on the computation of a deviance or discrepancy between the model and the available data. For the class of state-space model in Section 3, the deviance takes the form:

$$D(x_{0:T}, m) = -2 \log p(y_{1:T} | x_{0:T}, m = 1)$$

and the DIC criterion consists in choosing the value of m for which the statistic:

$$DIC(m) = \int D(x_{0:T}, m) \pi(dx_{0:T} | y_{1:T}, m) + d_p \quad (19)$$

is minimum, where d_p is a penalty term (that penalizes higher dimensional models) of the form:

$$d_p = \int D(x_{0:T}, m) \pi(dx_{0:T} | y_{1:T}) - D\left(\int x_{0:T} \pi(dx_{0:T} | y_{1:T}), m\right) \quad (20)$$

Obviously, the integrals in Equations (19) and (20) cannot be computed exactly, but we can resort to the random measure:

$$\pi_M(dx_{0:T} | y_{1:T}, m) = \frac{1}{M} \sum_{i=1}^M \delta_{x_{0:T}^{(i)}}(dx_{0:T})$$

produced by the particle filter in order to approximate them.

Therefore, a test based on the DIC criterion is, in principle, implementable. However, it is straightforward to realize that the numerical implementation of the MAP model selection technique is more efficient. Indeed, the implementation of the proposed MAP scheme only requires the approximation of the sequence of measures $\pi(dx_t | y_{1:t-1}, m)$ for model m (namely, to be used in Equation (16)), where x_t is a 10×1 vector for every $t = 0, 1, \dots, T$. To implement the DIC scheme,

on the contrary, we need to build an approximation of $\pi(dx_{0:T} | y_{1:T}, m)$ for each m , where the dimension of $x_{0:T}$ is $10(T+1) \times 1$. It is well known that PF's are efficient in the approximation of $\pi(dx_t | y_{1:t}, m)$ but not for the approximation of the higher dimensional $\pi(dx_{0:T} | y_{1:T}, m)$; see, e.g. Doucet et al. (2000), Cappé et al. (2007) or Künsch (2013) for a discussion. This can potentially impair the computation of the deviances for the competing models, specially (and paradoxically) as more data is accumulated.

6. Experimental Results

The proposed comparison method is validated using the two ANN-based state-space models defined in Section 2.4 and experimental data collected in the streets 2, 3 and 6 of Table 2. We have chosen these three streets because they present different traffic conditions and urban settings. The street number 2 is a large commercial area with high flow of heavy vehicles, light vehicles, and motorcycles, the street number 3 has a great ascendant traffic slope and high flow of heavy vehicles, and the street number 6 is a narrow street with low traffic flow.

A priori, model 1 should be expected to obtain the best results because its prediction function (network 1) has been calibrated using an additional Bayesian regularization process which has been shown to improve the generalization capabilities in neural networks (MacKay, 1992; Foresee and Hagan, 1997). Moreover, in Torija et al. (2012), the configuration of network 1 slightly outperformed the configuration of network 2, in terms of mean squared error (MSE) and linear regression.

In this paper, we assess the candidate models with the presented method in two ways. First, we compare their ability to predict the 23 indicators of sound pressure level jointly and, second, we evaluate their ability to predict 6 relevant scalar indicators individually. In each case, we apply the proposed MAP selection method to the two competing models using $M = 500$ samples in the particle filtering algorithm and data sequences of length $T = 25$ time steps (2 minutes each). We consider the same prior probability, $p(m) = 0.5$, for both models.

Moreover, we show the improvement of the best model (the model with the maximum value of $J(T)$) in all the experiments, i.e., we calculate the normalized difference between the goodness indicators $J(T)$ of the candidate models in percentage:

$$\text{Improvement (\%)} = 100 \times \frac{|J_1(T) - J_2(T)|}{\min(J_1(T), J_2(T))}$$

6.1. Model Comparison for the Complete Indicator Set

In this first set of experiments, the observation vectors $y_{1:L}, y_T$ are 23 dimensional, as shown in Equation (3), i.e., they contain all the indicators of sound pressure level described in Section 2.1. Using the particle approximation of Equation (17), we have evaluated the objective function $J_m(T)$ for the two competing models ($m = 1$ and $m = 2$) using separately the time series of data from streets 2, 3 and 6.

Table 3. Model Comparison for the Complete Indicator Set in Three Selected Streets*

Model (<i>m</i>)	$J_m(T)$		
	Street 2	Street 3	Street 6
1	-728.12	-764.99	-851.18
2	-939.20	-970.01	-1092.56

*The highest value of $J_m(T)$ which is obtained for each street is displayed in bold face.

Table 4. Model Comparison for Scalar Observations in Three Selected Streets*

Street	Observation (i)	$J_m^i(T)$		Improvement (%)
		Model 1 (<i>m</i> = 1)	Model 2 (<i>m</i> = 2)	
2	1	-44.34	-56.59	21.6
	2	-35.56	-40.90	13.1
	3	-37.53	-41.66	9.9
	4	-37.53	-44.95	16.5
	5	-43.78	-50.58	13.4
	6	-36.01	-40.70	11.5
3	1	-50.70	-57.16	11.3
	2	-50.01	-50.13	0.2
	3	-50.26	-52.42	4.1
	4	-48.08	-51.21	6.1
	5	-45.85	-50.12	8.5
	6	-38.68	-42.10	8.1
6	1	-68.81	-63.04	8.4
	2	-41.50	-41.37	0.3
	3	-54.56	-53.10	2.7
	4	-52.18	-51.40	1.5
	5	-52.41	-54.19	3.3
	6	-72.89	-64.61	11.4

*The highest value of $J_m^i(T)$ for each observation is shown in bold face.

The results are displayed in Table 3. We observe that $J_1(T) > J_2(T)$ for the three data sets. Specifically, model $m = 1$ attains an improvement of 22.5% for street number 2, 22.1% for street number 3 and 22.1% for street number 6. Therefore, model $m = 1$ is the MAP choice ($\hat{m} = 1$) for these three streets when the complete set of indicators (sound pressure level and its spectral composition) are considered.

Figure 1 shows the evolution of the objective functions $J_1(t)$ and $J_2(t)$ over time, for $1 \leq t \leq T = 25$ steps, for the three data sets corresponding to streets 2, 3 and 6, respectively. We observe that $J_1(t) > J_2(t)$ for all, even small, t . Moreover, the gap $J_1(t) - J_2(t)$ grows consistently over time, indicating that model $m = 1$ is a better fit than model $m = 2$ for these data sets.

This result confirm that the model which uses the neural network trained with an additional Bayesian regularization algorithm (network 1; studied in Torija et al. (2012)) obtains better predictions.

6.2. Model Comparison for Scalar Indicators

In this section, we show the results of using the proposed methodology to compare two competing models for a subset

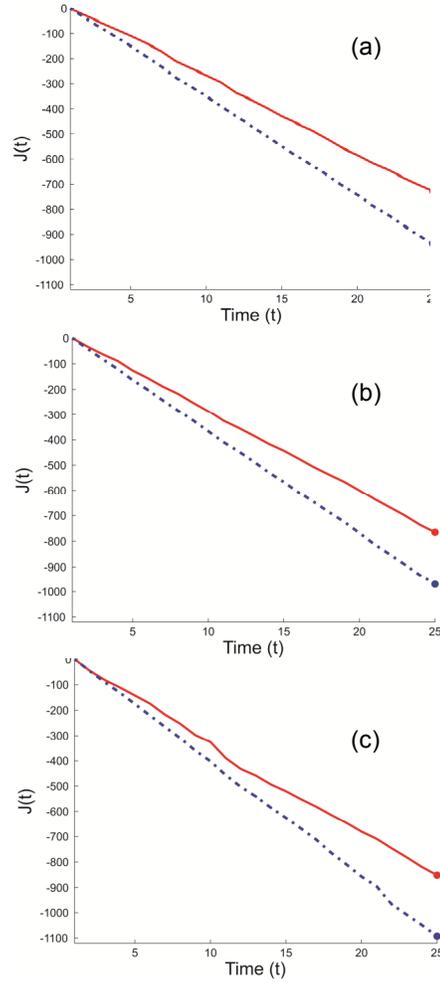


Figure 1. Model comparison in streets 2(a), 3(b) and 6(c) using all indicators (23 sound pressure levels). $J_1(T)$ and $J_2(T)$ are depicted in solid and dash-dotted lines, respectively, for $1 \leq t \leq T = 25$. The dots at $t = 25$ yield the values of $J_1(T)$ and $J_2(T)$.

of six (scalar) indicators: (1) L_{Aeq} , (2) L_{125Hz} , (3) L_{250Hz} , (4) L_{500Hz} , (5) L_{1KHz} , (6) $L_{3.15KHz}$. The studied descriptors characterize the overall sound pressure level and 1/3-octave band sound levels at low, medium and high frequencies. The aim is to show that the proposed technique can be successfully applied to compare the competing models in terms of their ability to estimate individual sound pressure level indicators and then achieve a deeper analysis of both candidates. Therefore, the observations y_t considered in this section are scalars and we successively compare the models for $y_t = y_{i,t}$, $i = 1, \dots, 6$, where $y_{i,t}$ represents the i -th indicator at time t . The comparisons are carried out using the objective functions:

$$J_m^i(t) \approx \sum_{k=1}^t \log(p_M(y_{i,k} | y_{i,1:k-1}, m)) + p(m), \quad i = 1, \dots, 6 \quad (21)$$

where the factors $p_M(y_{i,k} | y_{i,1:k-1}, m) = (1/M) \sum_{i=1}^M p(y_{i,k} | \tilde{x}_t^{\omega})$ are particle approximations of the true densities.

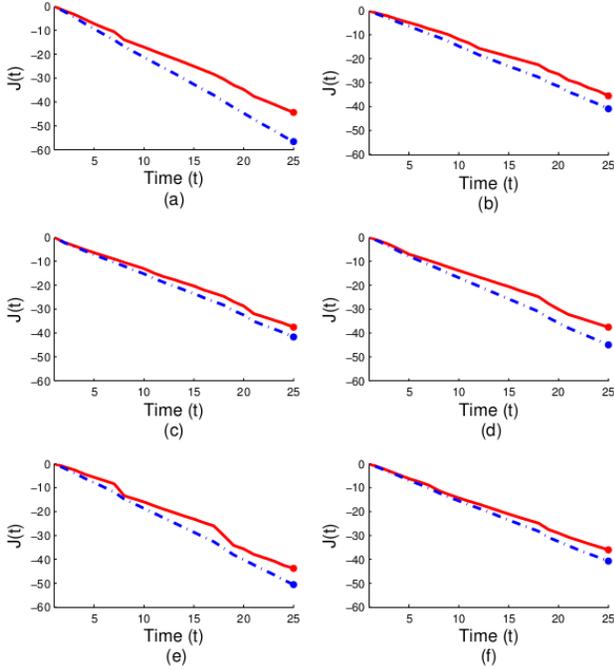


Figure 2. Street 2: Evolution of the functions $J_m^i(t), \dots, J_m^6(t)$, $m = 1, 2$, over time for $1 \leq t \leq T = 25$. (The curves for model $m = 1$ are depicted as solid lines, while the curves for model $m = 2$ are dash-dotted. Each plot corresponds to a different indicator of sound pressure level: (a) L_{Aeq} , (b) L_{125Hz} , (c) L_{250Hz} , (d) L_{500Hz} , (e) L_{1KHz} , and (f) $L_{3.15KHz}$. Same formats for other Figures)

6.2.1. Large Commercial Area with High Traffic Flow (Street 2)

Table 4 displays the values of $J_m^i(T)$ for $m = 1, 2$, and $i = 1, \dots, 6$, obtained for the data set of street number 2. The same as in Section 6.1, model $m = 1$ is clearly superior to model $m = 2$. Depending on the indicator of sound pressure level being observed, the improvement of model $m = 1$ over model $m = 2$ ranges from 9.9% up to 21.6%.

Figure 2 depicts the evolution of the objective functions $J_m^i(t)$, $m = 1, 2$, and $i = 1, \dots, 6$, over $1 \leq t \leq T = 25$. They show how model $m = 1$ improves over model $m = 2$ for all studied indicators when $t > 8$.

As in Section 6.1, the results are coherent with the fact that model 1 uses a neural network for noise prediction that works better than the network used in model 2 (see Torija et al., 2012).

6.2.2. Street with a Great Ascendant Traffic Slope and High Flow of Heavy Vehicles (Street 3)

We have conducted a similar study for street number 3. The final values of the objective functions, $J_m^i(T)$ for $m = 2$, $i = 1, \dots, 6$ and $T = 25$ time steps, are shown in Table 4. As in Section 6.2.1, model $m = 1$ is consistently better than model $m = 2$, although for the indicator of sound pressure level $y_{2,t} = L_{125Hz}$ the improvement is marginal and the two models can be considered equally good.

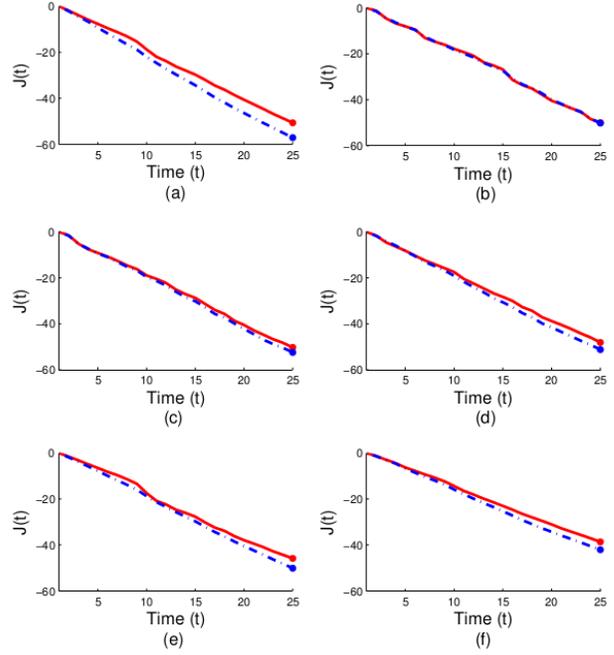


Figure 3. Street 3: Evolution of the functions $J_m^i(t), \dots, J_m^6(t)$, $m = 1, 2$, over time for $1 \leq t \leq T = 25$.

Figure 3 shows the evolution over time of the functions $J_m^i(t)$, $m = 1, 2$, $i = 1, \dots, 6$ and $1 \leq t \leq T = 25$. We observe that the two models perform equivalently for $y_{2,t} = L_{125Hz}$ but the difference $J_1(t) - J_2(t) > 0$ grows consistently over time for the other five indicators of sound pressure level.

As expected, model 1 (based on network 1) is also chosen as the best predictor of the studied scalar indicators by the proposed model selection method in this street.

6.2.3. Narrow Street with Low Traffic Flow (Street 6)

The same kind of experiment yields rather different results for the street number 6. Table 4 displays the values of the objective function $J_m^i(T)$ for $m = 1, 2$ and $i = 1, \dots, 6$. We observe that $J_2^i(t) > J_1^i(t)$ (i.e., model $m = 2$ has a larger posterior probability) for all studied indicators except $y_{5,t} = L_{1KHz}$. The improvement attained by model $m = 2$, however, is only relevant for the indicators L_{Aeq} and $L_{3.15KHz}$ ($i = 1$ and 6), respectively.

Figure 4, that shows the evolution over $1 \leq t \leq T = 25$ of the functions $J_m^i(t)$, $m = 1, 2$, $i = 1, \dots, 6$, allows us to observe that models $m = 1$ and $m = 2$ perform approximately equivalently for the indicators of sound pressure level $i = 2, 3, 4$ and 5. For $i = 1, 6$, we find that $J_2^i(t) > J_1^i(t)$ for $t \geq 10$.

The reasons why model $m = 2$ is a better predictor for the indicators of sound pressure level in this street number 6 while model $m = 1$ is superior in streets 2 and 3 can be assessed by considering the correlation between the dynamic state variables and the distinct indicators in each street.

Table 5 displays the Pearson correlation coefficients (Rodgers and Nicewander, 1988) between each one of the ten

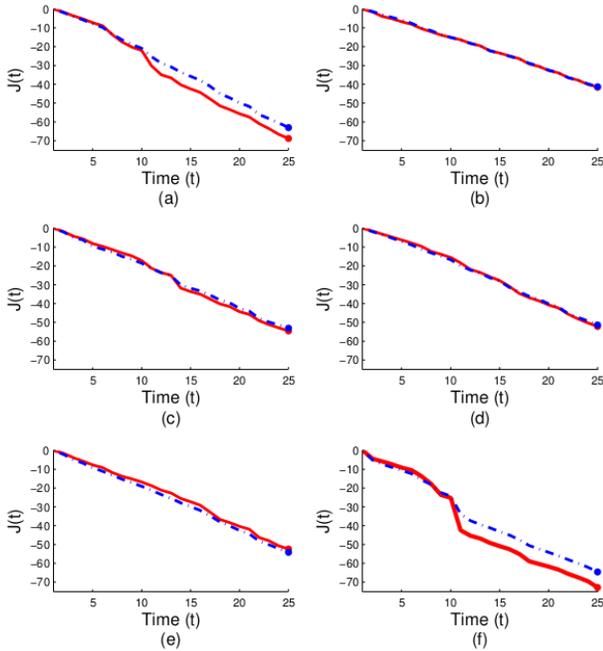


Figure 4. Street 6: Evolution of the functions $J^1_m(t), \dots, J^6_m(t)$, $m = 1, 2$, over time for $1 \leq t \leq T = 25$.

Table 5. Pearson’s Correlation Coefficients between Dynamical State Variables and L_{Aeq}

State Variables (x_i)	Street 2	Street 3	Street 6
1 Type of Traffic Flow	0.246	0.127	-0.717**
2 Ascendant Flow of Light Vehicles	0.058	0.312	-
3 Descendant Flow of Light Vehicles	-0.113	-	0.679**
4 Ascendant Flow of Heavy Vehicles	0.672**	0.716**	-
5 Descendant Flow of Heavy Vehicles	0.147	-	-
6 Ascendant Flow of Motorcycles	0.249	0.264	-
7 Descendant Flow of Motorcycles	-0.048	-	0.465*
8 Number of Vehicles with Sirens	-	-	-
9 Impulsive Sound Event related to Traffic	0.705**	0.756**	0.622**
10 Impulsive Sound Event unrelated to Traffic	0.525**	-0.180	0.398*

Note: Bilateral Significance: * $P \leq 0.05$, ** $P \leq 0.01$.

dynamic state variables, $x_{1,t} \dots x_{10,t}$, and the indicator of sound pressure L_{Aeq} (corresponding to $i = 1$). It is observed that the output L_{Aeq} is highly correlated with the dynamic variables: (a) $x_{4,t}$ (ascendant flow of heavy vehicles), $x_{9,t}$ (impulsive sound event related to traffic) and $x_{10,t}$ (impulsive sound event unrelated to traffic) in the street number 2, (b) $x_{4,t}$ and $x_{10,t}$ in the street number 3 and (c) $x_{1,t}$ (type of traffic flow), $x_{3,t}$ (descendant flow of light vehicles), $x_{7,t}$ (descendant flow of motorcycles) and $x_{9,t}$ (impulsive sound event related to traffic) for the street number 6.

It is apparent that the key state variables that determine the value of the indicator L_{Aeq} are different for the streets 2 and 3, on one hand, and the street number 6, on the other hand. Thus, according to the results of the comparison method, we deduce that the model $m = 1$ is a better predictor of sound

pressure levels in areas with a high traffic flow of heavy vehicles (streets 2 and 3), and the model $m = 2$ is a better predictor of the indicators in narrow streets with a low non-heavy traffic flow (street number 6).

Note that the proposed MAP model selection method is more general than the study of MSE values and regression coefficients (because it takes into account the complete pdf $p(y_{1:T} | m)$ for each competing model) and yields a single quantity, $J^i_m(T)$, to be compared between models, rather than calculating several performance indicators.

7. Conclusions

In this paper we have introduced a method for comparing dynamic models that predict indicators of the sound pressure level and its spectral composition. The proposed technique is based on the calculation of the posterior probability of each candidate model from a time series of measurements. The prediction models are formally described as dynamic systems in state-space form. Since the observations are nonlinear transformations of the state variables, the posterior probabilities cannot be calculated analytically and it is necessary to approximate them numerically. For this task, we use a particle filtering algorithm with Markov chain Monte Carlo moves in the resampling step to increase the diversity of the particles.

The selection of the model with the best performance is important because, even if both models have similar accuracy, the difference can imply wrong decisions in urban planning; hence a substantial increase in the annoyance of the inhabitants or in the quantity of affected population. The proposed comparison method enables us to select the best model (specifically, the one with the largest posterior probability given the available data) between similar competing models. For testing the procedure, we have chosen two ANN-based models as candidates. The reason for choosing these models is that they show a good behavior in predicting environmental noise levels from real data. For the comparison, we have used a series of experimental observations of sound pressure levels which were measured in the city of Granada (Spain). According to the results, on one hand, model number 1 is a better predictor than model number 2 for the complete indicator vector in three studied streets, i.e, the model number 1 makes the most reliable predictions of 23 sound pressure level indicators according to the MAP criterion that we propose in this paper. On the other hand, the model number 1 is a better predictor for all sound indicators which we studied individually in streets with a high traffic flow of heavy vehicles (streets 1 and 2), and the model number 2 has a higher ability to predict most of the studied scalar descriptors in narrow streets with a low non-heavy traffic flow (street 3).

These results confirm the experimental experience of our research group, since we had empirically found model number 1 to be the fittest for predicting the complete set of sound pressure level descriptors (the A-weighted sound pressure level L_{Aeq} , the no-weighted sound pressure level L_{eq} , and the sound level in 1/3 octave bands from 40 Hz to 4 kHz) in the given examples.

Acknowledgments. This work has been supported by the “Ministerio de Economía y Competitividad” of Spain under project TEC 2012-38883-C02-02. J.M. acknowledges the financial support of the Ministry of Science and Innovation of Spain (program Consolider-Ingenio 2010 CSD2008-00010 COMONSENS and project COMPREHENSION TEC2012-38883-C02-01). A.J.T. acknowledges the support of the University of Malaga and the European Commission under the Agreement Grant no. 246550 of the seventh Framework Programme for R & D of the EU, granted within the People Programme, Co-funding of Regional, National and International Programmes (COFUND)

References

- Anon. (1975). *Calculation of road traffic noise*, London: United Kingdom Department of the Environment and Welsh Office Joint Publication/HMSO.
- Bain, A., and Crisan, D. (2008). *Fundamentals of Stochastic Filtering*, Springer.
- Bernardo, J.M., and Smith, A.F.M. (2009). *Bayesian Theory*, vol. 405, Wiley.
- Bertoni, D., Franchini, A., and Magnoni, M. (1987). *Il Rumore Urbano e l'Organizzazione de Territorio*, Pitagora, Bologna, Italy.
- Björkman, M. (1991). Community noise annoyance: Importance of noise levels and the number of noise events. *J. Sound Vibrat.*, 151(3), 497-503. [http://dx.doi.org/10.1016/0022-460X\(91\)90549-Y](http://dx.doi.org/10.1016/0022-460X(91)90549-Y)
- Burgess, M.A. (1977). Noise prediction for urban traffic conditions-related to measurements in the Sydney Metropolitan Area. *Appl. Acoust.*, 10(1), 1-7. [http://dx.doi.org/10.1016/0003-682X\(77\)90002-0](http://dx.doi.org/10.1016/0003-682X(77)90002-0)
- Cammarata, G., Cavalieri, S., and Fichera, A. (1995). A neural network architecture for noise prediction. *Neural Networks*, 8(6), 963-973. [http://dx.doi.org/10.1016/0893-6080\(95\)00016-S](http://dx.doi.org/10.1016/0893-6080(95)00016-S)
- Cammarata, G., Cavalieri, S., Fichera, A., and Marletta, L. (1993). Noise prediction in urban traffic by a neural approach. In: *International Workshop on Artificial Neural Networks*, IWANN93, vol. 1, Barcelona, Spain, pp. 611-619.
- Cappé, O., Godsill, S.J., and Moulines, E. (2007). An overview of existing methods and recent advances in sequential Monte Carlo. *Proc. IEEE*, 95(5), 899-924. <http://dx.doi.org/10.1109/JPROC.2007.893250>
- Chib, S., and Greenberg, E. (1995). Understanding the Metropolis-Hastings algorithm. *Amer. Statist.*, 49(4), 327-335. <http://www.jstor.org/stable/2684568>
- Crisan, D. (2001). Particle filters-A theoretical perspective. In: Doucet, A., de Freitas, N., Gordon, N. (Eds.), *Sequential Monte Carlo Methods in Practice*, Springer, Ch. 2, pp. 17-42. http://dx.doi.org/10.1007/978-1-4757-3437-9_2
- Dai, L., Cao, J., Fan, L., and Mobed, N. (2005). Traffic noise evaluation and analysis in residential areas of Regina. *J. Environ. Inf.*, 5(1), 17-25. <http://dx.doi.org/10.3808/jei.200500042>
- Directive 2002/49/EC (2002). *Directive of the European Parliament and of the Council of 25 June 2002*, relating to the assessment and management of environmental noise.
- Djuric, P.M. (1998). Asymptotic MAP criteria for model selection. *IEEE Trans. Signal Process.*, 46(10), 2726-2735. <http://dx.doi.org/10.1109/78.720374>
- Djuric, P.M. (1999). Monitoring and selection of dynamic models by Monte Carlo sampling. *Proc. IEEE Signal Proc. Workshop on Higher-Order Statistics*, 191-194. <http://dx.doi.org/10.1109/HOST.1999.778723>
- Djuric, P.M., Kotecha, J.H., Zhang, J., Huang, Y., Ghirmai, T., Bugallo, M. F., and Míguez, J. (2003). Particle filtering. *IEEE Signal Proc. Maga.*, 20(5), 19-38. <http://dx.doi.org/10.1109/MSP.2003.1236770>
- Don, C. G., and Rees, I. G. (1985). Road traffic sound level distributions. *J. Sound Vibrat.*, 100(1), 41-53. [http://dx.doi.org/10.1016/0022-460X\(85\)90341-4](http://dx.doi.org/10.1016/0022-460X(85)90341-4)
- Doucet, A., de Freitas, N., and Gordon, N. (2001). An introduction to sequential Monte Carlo methods. In: Doucet, A., de Freitas, N., Gordon, N. (Eds.), *Sequential Monte Carlo Methods in Practice*, Springer, Ch. 1, pp. 4-14. <http://dx.doi.org/10.1007/978-1-4757-3437-9>
- Doucet, A., Godsill, S., and Andrieu, C. (2000). On sequential Monte Carlo sampling methods for Bayesian filtering. *Stat. Comput.*, 10(3), 197-208. <http://dx.doi.org/10.1023/A:1008935410038>
- Foresee, F.D., and Hagan, M.T. (1997). Gauss-Newton approximation to Bayesian regularization. *Proceedings of the 1997 International Joint Conference on Neural Networks*, 3, pp. 1930-1935.
- Garcia, A., and Faus, L.J. (1991). Statistical analysis of noise levels in urban areas. *Appl. Acoust.*, 34(4), 227-247. [http://dx.doi.org/10.1016/0003-682X\(91\)90007-2](http://dx.doi.org/10.1016/0003-682X(91)90007-2)
- Genaro, N., Torija, A., Ramos-Ridao, A., Requena, I., Ruiz, D. P., and Zamorano, M. (2010). A neural network based model for urban noise prediction. *J. Acoust. Soc. Am.*, 128(4), 1738-1746. <http://dx.doi.org/10.1121/1.3473692>
- Gilks, W.R., and Berzuini, C. (2001). Following a moving target-Monte Carlo inference for dynamic Bayesian models. *J. Roy. Stat. Soc. Ser. B. (Stat. Method.)*, 63(1), 127-146. <http://dx.doi.org/10.1111/1467-9868.00280>
- Gordon, N.J., Salmond, D.J., and Smith, A.F.M. (1993). Novel approach to nonlinear and non-Gaussian Bayesian state estimation. *IEE Proceedings-F*, 140(2), 107-113. <http://dx.doi.org/10.1049/ip-f-2.1993.0015>
- Hagan, M.T., and Menhaj, M. (1994). Training feed-forward networks with the Marquardt algorithm. *IEEE Trans. on Neural Network*, 5(6), 989-993. <http://dx.doi.org/10.1109/72.329697>
- Josse, R. (1972). *Notions d'acoustique*, Eyrolles, Paris, France.
- Kragh, J., Svein, A., and Jonasson, H.G. (2002). *Nordic environmental noise prediction methods*, Nord2000, Summary report, Denmark: DELTA.
- Künsch, H.R. (2013). Particle filters. *Bernoulli*, 19(4), 1391-1403. <http://dx.doi.org/10.3150/12-BEJSP07>
- Lapedes, A., and Farber, R. (1987). *Nonlinear signal processing using neural networks: Prediction and system modeling*, Technical Report LA-UR-87-2662, Los Alamos National Laboratory.
- Laszlo, H., McRobie, E., Stansfeld, S., and Hansell, A. (2012). Annoyance and other reaction measures to changes in noise exposure-a review. *Sci. Total Environ.*, 435-436(0), 551-562. <http://dx.doi.org/10.1016/j.scitotenv.2012.06.112>
- Lercher, P. (1996). Environmental noise and health: an integrated research perspective. *Environ. Int.*, 22(1), 117-128. [http://dx.doi.org/10.1016/0160-4120\(95\)00109-3](http://dx.doi.org/10.1016/0160-4120(95)00109-3)
- Lui, W.K., and Li, K.M. (2004). A theoretical study for the propagation of rolling noise over a porous road pavement. *J. Acoust. Soc. Am.*, 116(1), 313-322. <http://dx.doi.org/10.1121/1.1751153>
- MacKay, D.J.C. (1992). Bayesian interpolation. *Neural Comput.*, 4(3), 415-447. <http://dx.doi.org/10.1162/neco.1992.4.3.415>
- MacKay, D.J.C. (2003). *Information Theory, Inference and Learning Algorithms*, Cambridge University Press.
- Marquardt, D. (1963). An algorithm for least-squares estimation of nonlinear parameters. *SIAM J. Appl. Math.*, 11(2), 431-441. <http://dx.doi.org/10.1137/0111030>
- McClelland, J.L., and Rumelhart, D.E. (1988). *Explorations in parallel distributed processing*, MA: MIT Press, Cambridge.
- Moral, P.D. (2004). *Feynman-Kac Formulae: Genealogical and Interacting Particle Systems with Applications*, Springer. <http://dx.doi.org/10.1007/978-1-4684-9393-1>

- NMPB (1996). *Road traffic noise: new French calculation method including meteorological effects (NMPB-1996 method)*, Experimental version, CERTU: France.
- Ouis, D. (2001). Annoyance from road traffic noise: A review. *J. Environ. Psychol.*, 21(1), 101-120. <http://dx.doi.org/10.1006/jevp.2000.0187>
- Rodgers, J.L., and Nicewander, W.A. (1988). Thirteen ways to look at the correlation coefficient. *Amer. Statist.*, 42(1), 59-66. <http://dx.doi.org/10.2307/2685263>
- Sachakamol, P., Dai, L., Quinn, M., Alexander, S., Heck, N., Chernoff, G., Huang, G. W., Li, Y.P., Huang, G.H., Sun, W., et al. (2011). Parametric influence on prediction of sound absorption coefficients for asphalt pavements. *J. Environ. Inf.*, 18(1), 1-11. <http://dx.doi.org/10.3808/jei.201100193>
- Spiegelhalter, D.J., Best, N.G., Carlin, B.P., and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *J. Roy. Stat. Soc. Ser. B. (Stat. Method.)*, 64(4), 583-639. <http://dx.doi.org/10.1111/1467-9868.00353>
- Suykens, J.A.K., Vandewalle, J., and de Moor, B.L.R. (1996). *Artificial neural networks for modelling and control of nonlinear systems*, Kluwer Academic Publishers. <http://dx.doi.org/10.1007/978-1-4757-2493-6>
- Torija, A.J., Genaro, N., Ruiz, D.P., Ramos-Ridao, A., Zamorano, M., and Requena, I. (2010). Priorization of acoustic variables: Environmental decision support for the physical characterization of urban sound environments. *Build. Environ.*, 45(6), 1477-1489. <http://dx.doi.org/10.1016/j.buildenv.2009.12.011>
- Torija, A.J., Ruiz, D.P., and Ramos-Ridao, A. (2007a). Characterization of the different types of vehicles flow in traffic, *Proceedings of the 19th International Congress on Acoustics*, Madrid, 5(9), p. 53.
- Torija, A.J., Ruiz, D.P., and Ramos-Ridao, A. (2007b). A method for prediction of the stabilization time in traffic noise measurements, *Proceedings of the 19th International Congress on Acoustics*, Madrid.
- Torija, A.J., Ruiz, D.P., and Ramos-Ridao, A. (2007c). Obtaining of a factor to describe the anomalous sound events in traffic noise measurements, *Proceedings of the 19th International Congress on Acoustics*, Madrid, 5(9), p. 53.
- Torija, A.J., Ruiz, D.P., and Ramos-Ridao, A. (2011). Required stabilization time, short-term variability and impulsiveness of the sound pressure level to characterize the temporal composition of urban soundscapes. *Appl. Acoust.*, 72(2), 89-99. <http://dx.doi.org/10.1016/j.apacoust.2010.09.011>
- Torija, A.J., Ruiz, D.P., and Ramos-Ridao, A.F. (2012). Use of back-propagation neural networks to predict both level and temporal-spectral composition of sound pressure in urban sound environments. *Build. Environ.*, 52, 45-56. <http://dx.doi.org/10.1016/j.buildenv.2011.12.024>
- Watts, G. (2005). *Harmonoise prediction model for road traffic noise*, PPR 034.