

Application of the Bayesian Inference for Estimation of the Long-Term Noise Indicators and Their Uncertainty

W. BATKO* AND B. STĘPIEŃ

Department of Mechanics and Vibroacoustics, AGH University of Science and Technology
al. A. Mickiewicza 30, 30-059 Kraków, Poland

The problem of estimation of the environmental noise hazard indicators and their uncertainty is presented in the hereby paper. The main attention is focused on the estimation process of the long-term noise indicators and their type *A* standard uncertainty defined by the standard deviation of the mean of the measurement results. The rules given in the ISO/IEC Guide 98 are used in the calculations. It is usually determined by means of the classic variance estimators, at the assumption of the normality of measurements results. However, such assumption in relation to the acoustic measurements is rather questionable. This is the reason that the authors indicated the necessity of implementation of non-classic statistic solutions. There was formulated the estimation idea of seeking density function of long-term noise indicators distribution by the Bayesian inference, which does not generate limitations for form and properties of analyzed statistics. There was presented theoretical basis of the proposed method, and the example of calculation process which make possible determining searched estimators of expected value and variance of long-term noise indicators L_{DEN} and L_N . The illustration for indicated solutions and usefulness analysis was constant monitoring results of traffic noise recorded on one of the main arteries of Kraków, Poland.

PACS: 43.50.Rq, 43.50.Yw

1. Introduction

Directive 2002/49/EC of the European Parliament [1] requires carrying out a long-term policy of the environment protection against noise in the European Union countries. Its realisation is based on the estimation of long-term noise indicators L_{DEN} and L_N in sites being under the protection.

The average *A*-weighted long-term sound levels L_{DEN} and L_N in dB are determined on the basis of noise annoyance indicators $L_{DEN,i}$ for $i = 1, 2, \dots, M$ (where $M = 365$ or $M = 366$ for the leap year) of all days in the calendar year at the day-evening-night periods [2]:

$$L_{DEN,i} = 10 \log \left(\frac{1}{24} \left(12 \times 10^{0.1 L_{D,i}} + 4 \times 10^{0.1(L_{E,i}+5)} + 8 \times 10^{0.1(L_{N,i}+10)} \right) \right), \quad (1)$$

where $L_{D,i}$ — *A*-weighted sound level, determined from the day-time noise exposure i.e. from 6:00 a.m. to 6:00 p.m., dB, $L_{E,i}$ — *A*-weighted sound level, determined from the noise exposures from 6:00 p.m. to 10:00 p.m., dB, $L_{N,i}$ — *A*-weighted sound level, determined for the night periods i.e. from 10:00 p.m. to 6:00 a.m., dB, and night periods $L_{N,i}$ for $i = 1, 2, \dots, M$ determined by relation [2]:

$$L_{N,i} = 10 \log \left(\frac{1}{K} \sum_{i=1}^K 10^{0.1(L_{Aeq,T})_i} \right), \quad (2)$$

where K — sample size, $(L_{Aeq,T})_i$ — equivalent sound level for the i -th sample, dB.

Estimation of the long-term indicators of the acoustic hazard for the environment L_{DEN} and L_N being the average value calculated from all calendar days

$$L_{DEN} = 10 \log \left(\frac{1}{M} \sum_{i=1}^M 10^{0.1 L_{DEN,i}} \right), \quad (3)$$

$$L_N = 10 \log \left(\frac{1}{M} \sum_{i=1}^M 10^{0.1 L_{N,i}} \right), \quad (4)$$

forming a set of two indicators $L_{Aeq,LT} = \{L_{DEN}, L_N\}$, requires an access to the results of the whole year sound level monitoring.

The necessity of validation of the obtained results, which requires the analysis of uncertainty budget of estimation, is connected with the process of calculating the average long-term noise indicators determined by values of L_{DEN} and L_N .

An essential component of such budget is the standard type *A* uncertainty defined as the standard deviation of the mean from the inspections results. The rules given in the ISO/IEC Guide 98 [3] are used in the calculations. They are based on the classic variance estimators and on the condition of assigning the normal distribution to random results of the sampling inspections.

However, the application of its recommendations in respect of the estimation of the type *A* standard uncertainty of noise hazard indices in environment is rather dubious. As it results from [4–7], the assumption of a normal distribution of measurement results is difficult to be accepted. This is confirmed by the analysis of the measurement results of traffic noise, which in significant ma-

* corresponding author; e-mail: batko@agh.edu.pl

jority required the rejection of the hypothesis concerning the possibility of using the normal distribution for the description. Extra-statistical information in relation to the occurrence of certain noise expositions in environment, especially in night hours (more than one maximum) also discredit this assumption.

However, in practice, there is a necessity of estimating the average long-term noise indicators $L_{\text{Aeq},LT}$ on the basis of environment sampling inspections. It generates the need of estimation of the expected value and variances of the controlled noise hazard indicators from the highly limited random sample, which does not undergo the normal distribution. Therefore searching for non-standard procedures of the estimation of the average long-term noise indicators L_{DEN} and L_N and their variances, seems to be necessary.

The authors proposed the Bayesian inference for solving these problems. Inference about θ parameter is made on the basis of the conditional distribution of this parameter for the fixed observation x , often called the posterior distribution from the Bayes formula. An application of this idea, based on the Bayesian inference, constitutes a premise for increasing the estimation accuracy of the expected value and variances of the long-term noise indicators $L_{\text{Aeq},LT}$ on the bases of the results obtained by sampling inspections. Discussion of the method, together with the example illustrating its functioning, will be contained in the present paper. The reference base constitutes the results of the constant noise monitoring recorded in one of the main arteries of Kraków, Poland.

2. Bayesian inference

2.1. Bayes' theorem

Bayesian methods are currently regarded as non-classical statistical methods. Their difference compared to classical methods is the result of the adoption of other interpretations of probability. In the classical approach it is interpreted as an idealization of frequency, while in the Bayesian approach the probability is a measure of the researcher's degree of belief about the veracity of the event occurring. In consequence the model classical parameters are recognised as random variables.

Suppose we are trying to estimate a random parameter θ from data x . Then the associated conditional density $p(\theta|x)$ is called the posterior density because the estimate is conditioned "after the measurements" have been acquired. Estimators based on this a posteriori density are usually called Bayesian because they are constructed from Bayes' theorem, since $p(\theta|x)$ is difficult to be obtained directly. That is, Bayes' rule is defined [8–10]:

$$p(\theta|x) = \frac{f(x)}{p(x)} = \frac{p(x|\theta)p(\theta)}{\int_{\Omega} p(x|\theta)p(\theta)d\theta}, \quad (5)$$

where $p(\theta)$ is called the prior density (before measurement), $p(x|\theta)$ is called the sampling density or likelihood (more likely to be true), $p(x)$ is called the marginal data density or evidence (normalizes the posterior to assure its

integration to unity), Ω is called the parameters space. Bayesian methods view the sought after parameter as random possessing a "known" a priori density. As measurements are made, the prior is converted to the posterior density function adjusting the parameter estimates. Thus, the result of increasing the number of measurements is to improve the a posteriori density resulting in a sharper peak closer to the true parameter. The attractiveness of the Bayesian approach is also associated with the fact that the posterior density reflects information about unknown parameters in a purely probabilistic method and asymptotic approximation-free, which would entail the total inference for small samples [11, 12].

In this paper to generate samples from posterior density here was applied the random walk Metropolis–Hastings sampling method, discussed below.

2.2. Random walk Metropolis–Hastings sampling

The Markov chain simulation is essentially a general technique based on generating samples from the proposal distributions and then correcting (acceptance or rejection) those samples to approximate a target posterior distribution. Here we must know both the target $p(\theta|x)$, and the proposal $p_E(\xi)$ distributions. The samples are sequentially generated forming the Markov chain. Typically, in the Markov chain simulation, samples are generated from the transition kernel or distribution. The key, however, is not really the chain itself, but the fact that the approximate distribution improves sequentially as it converges to the target posterior.

In this subsection we discuss the random walk Metropolis–Hastings sampling method (random walk M–H). The fundamental idea is similar to the rejection method [10]. The random walk Metropolis–Hastings technique defines the Markov chain indexed by consecutive integers $i = 1, \dots, S, S+1, \dots, S+N$, where first S samples are treated as the cycles burned and discarded. The idea is to perturb the current sample x_i with an addition of random error, that is [10]:

$$x_{i+1} = x_i + \xi_i, \quad (6)$$

where ξ_i is a random error from the proposal distribution $p_E(\xi)$. A reasonable choice for this distribution is a symmetric Gaussian, that is $p_E(\xi) \sim N(0, \sigma_{\xi}^2)$.

In the first step, drawing a random error ξ_i from the proposal distribution $p_E(\xi)$, next generate "candidate" sample \hat{x}_i from (6) and drawing random sample u_i from the uniform distribution $U(0, 1)$. Now calculate the acceptance probability [10, 12]:

$$\rho(x_i, \hat{x}_i) = \min \left\{ \frac{f(\hat{x}_i)}{f(x_i)}, 1 \right\} \quad (7)$$

and then making a decision on the base [10, 12]:

$$x_{i+1} = \begin{cases} \hat{x}_i & \text{if } u_i < \rho(x_i, \hat{x}_i), \\ x_i & \text{otherwise} \end{cases} \quad (8)$$

whether this candidate should be accepted and retained

or rejected and discarded using the previous sample as the new. If accepted, \hat{x}_i replaces x_{i+1} otherwise the old sample x_i is saved. This is the core of the random walk M–H approach in its simplest form.

With this algorithm, we must use both the (known) proposal and target distributions to calculate the acceptance probability (7) and then generate samples (random walk) from the proposal. It is important to realize that a “good” proposal distribution can assure generating samples from the desired target distribution, although the samples must still be generated to “cover” its range. This method does not require knowledge of the marginal data density $p(x)$ from Eq. (5) [12].

2.3. Bayesian estimation

The random walk M–H method is able to generate correlated sample X from a continuous posterior distribution.

In the Bayesian inference the form of parameter θ depends not only on the prior and posterior distributions, but also on the loss function L . If the loss function is a quadratic function of form [13]:

$$L(\theta, d) = C(d - \theta)^2, \quad C > 0, \quad (9)$$

where d — decision. The Bayesian estimate of θ parameter is expected value of posterior distribution, which was determined as the mean from the Markov chain after removed cycles burned [8, 10]:

$$\bar{\theta}^{\text{BAY}} = \hat{E}[p(\theta|x)] = \frac{1}{N - S} \sum_{i=S+1}^N \theta_i^{\text{BAY}}, \quad (10)$$

where θ_i^{BAY} — elements of random sample X from posterior distribution.

Because of that we received correlated sample X as result of sampling. The autocorrelation function should be designated and there should be selected the elements that are not correlated. We obtain a simple random sample Y of size k on the base of which we can calculate the standard deviation θ parameter as follows [8, 10]:

$$\begin{aligned} \hat{s}_{\text{BAY}} &= \sqrt{\hat{E} \left\{ [p(\theta|x)]^2 \right\} - \left\{ \hat{E} [p(\theta|x)] \right\}^2} \\ &= \sqrt{\frac{1}{k} \sum_{i=1}^k \left(\theta_i^{\text{BAY}*} - \bar{\theta}^{\text{BAY}} \right)^2}, \end{aligned} \quad (11)$$

where $\theta_i^{\text{BAY}*}$ — elements of simple random sample Y .

3. Estimation of parameters of the long-term noise indicators

Point estimation of parameters of the long-term noise indicators $L_{\text{Aeq},LT}$ (an expected value and type A standard uncertainty) was performed on the bases of sampling inspections — when utilising the classical and the Bayesian inference.

The values determined from the data base of the results recorded in the year 2004 and 2005 by the constant

monitoring system operating in one of the main arteries in Kraków were assumed as the measured value of $L_{\text{Aeq},LT}$. This enabled the possibility of assessing the error of estimation of the expected values of the long-term noise indicators generated by solution proposed.

Type A uncertainty and expected value of the long-term noise hazard indicators was also determined with the application of two types estimators:

- classical, which assumes the normality of results of the long-term noise indicators,
- Bayesian, which does not generate limitations regarding forms and properties of the investigated statistics.

The investigated populations constituted the results of the 24 h average day A -weighted sound level and the 24 h average night A -weighted sound level determined on the basis of the constant monitoring. Out of those populations simple samples of sizes: $n = 5, 9, 12, 15$, simulating the number inspection days, were sampling. Those samples were the assessment basis for L_{DEN} and L_{N} . Utilising the application developed in the Matlab software packet the expected values as well as the type A uncertainty of the long-term noise indicators were assessed.

The expected values of indicators — in the classical approach — were determined by the equation

$$\bar{L}_{\text{Aeq},LT} = 10 \log \left(\frac{1}{n} \sum_{i=1}^n 10^{0.1 L_{\text{Aeq},LT,i}} \right), \quad (12)$$

where n — sample size, $L_{\text{Aeq},LT,i}$ — index level for the i -th sample, dB.

The type A uncertainty of the long-term noise hazard indicators were determined by equation [3]:

$$s(\bar{L}_{\text{Aeq},LT}) = \sqrt{\frac{\sum_{i=1}^n (L_{\text{Aeq},LT,i} - \bar{L}_{\text{Aeq},LT})^2}{n(n-1)}}. \quad (13)$$

In this experiment posterior distributions of the long-term noise indicators are designated on the basis Eq. (5). Prior and sampling density determined make use of kernel density estimation presented in [14–17]. The samples of sizes: $n = 5, 9, 12, 15$, were the assessment basis to determined sampling density in the Bayesian inference. For sampling density from year 2004, prior density was determined on the basis results from year 2005 whereas, for sampling density from year 2005, prior density was determined on the basis results from year 2004. For the purposes of this simulation experiment $N = 70,000$ samples from posterior distribution were generated, while the first $S = 20,000$ were treated as burned. To generate samples from posterior distribution used the random walk M–H, discussed in Sect. 2.2. After considering the comments made in Sect. 2.3 the Bayesian estimate of long-term noise indicators were defined as

$$\bar{L}_{\text{Aeq},LT}^{\text{BAY}} = \frac{1}{N-S} \sum_{i=S+1}^N L_{\text{Aeq},LT,i}^{\text{BAY}} \quad (14)$$

where $L_{\text{Aeq},LT,i}^{\text{BAY}}$ — index level for the i -th sample from X , dB, N — total number of samples, S — number of samples burned, however the Bayesian estimates type A standard uncertainty were calculated

$$s_{\text{BAY}}(\bar{L}_{\text{Aeq},LT}^{\text{BAY}}) = \sqrt{\frac{1}{k} \sum_{i=1}^k \left(L_{\text{Aeq},LT,i}^{\text{BAY}*} - \bar{L}_{\text{Aeq},LT}^{\text{BAY}} \right)^2}, \quad (15)$$

where $L_{\text{Aeq},LT,i}^{\text{BAY}*}$ — index level for the i -th sample from Y , dB, k — sample size of Y .

4. Experimental results

The long-term environmental noise hazard indicators i.e. the day-evening-night level L_{DEN} and the night level L_{N} were determined on the basis of the results recorded in the year 2004 and 2005 by the constant acoustic monitoring station installed at one of the main arteries in Kraków.

The estimated results of the long-term environment noise hazards $L_{\text{Aeq},LT} = \{L_{\text{DEN}}, L_{\text{N}}\}$ as well as the type A uncertainty — describing them — obtained by application of the mentioned above methods are presented in Table I and Table II.

There are certain discrepancies between the known values of L_{DEN} and L_{N} , and the estimated values (classical and Bayesian), eight Bayesian estimate had a significant error (marked bold in Table I). On the basis of this experiment, it is difficult to say unequivocally which of the presented methods provided results with smaller error (the difference between the measured value and the estimated). The solution of this problem requires further research.

The estimation results of the standard type A uncertainty of the long-term environment noise hazard indicators determined for the same random samples (as the expected values shown in Table I) are presented in Table II.

It can be noticed that the type A uncertainties estimated by the Bayesian method are higher than the ones estimated by the classic method (marked bold in Table II). They have been estimated on the basis of distributions, not just random samples. This can indicate a more accurate estimation of the long-term noise indicators carried out by means of the Bayesian method.

5. Conclusions

The results described in the paper, concerning the application of the Bayesian estimation for the assessment of the mean of the long-term noise indicators and their uncertainties, indicate that they can constitute the new and promising calculation tool. They enrich the existing calculating algorithms. It should be mentioned that assumptions related to their application are free from limitations corresponding to the classic estimation analysis.

TABLE I

Estimates of the expected values of long-term noise indicators.

Indicator	Measured value [dB]	Sample size n	Classic estimate [dB]	Bayesian estimate [dB]
year 2004				
L_{DEN}	77.1821	5	77.0095	76.6370
		9	77.2291	77.1242
		12	77.3112	77.2514
		15	77.2424	77.2096
L_{N}	69.4956	5	68.9590	69.0768
		9	69.4655	69.3502
		12	69.4687	69.3287
		15	69.2699	69.2170
year 2005				
L_{DEN}	76.5461	5	75.6662	76.9996
		9	77.2036	77.3000
		12	76.1675	77.0731
		15	76.3431	77.0216
L_{N}	68.9467	5	69.7324	69.5002
		9	67.7056	69.2195
		12	68.1616	69.1306
		15	68.6730	69.1887

TABLE II

Estimates of the type A uncertainty of long-term noise indicators.

Indicator	Sample size n	Classic estimate [dB]	Bayesian estimate [dB]
year 2004			
L_{DEN}	5	0.7708	1.0979
	9	0.1702	0.5242
	12	0.1578	0.5345
	15	0.2551	0.7320
L_{N}	5	0.4073	0.9347
	9	0.1883	0.5229
	12	0.1914	0.6144
	15	0.2391	0.6858
year 2005			
L_{DEN}	5	1.1331	0.8873
	9	0.6212	0.5352
	12	0.7246	0.8546
	15	0.4051	0.6213
L_{N}	5	0.3062	0.6230
	9	0.7563	1.0009
	12	0.5494	0.7688
	15	0.5968	0.8459

They neither require knowing the size and character of the statistics nor the access to large bases of samples, which makes them more likely in an application.

The simulation experiment of the determination of the expected value and the type *A* uncertainty was carried out by means of two methods: classical and Bayesian, and its results presented. The results that were obtained allow to state that:

- it is difficult to say unequivocally which of the presented methods obtained results with smaller error (the difference between the measured value and the estimated); the solution of this requires further research,
- the type *A* uncertainty determined by the Bayesian method is of a higher value, because they have been estimated on the basis of distributions, which indicates that the estimation of the expected value of the long-term noise indicators is more accurate.

Due to the limited base of simulation experiments, the selection of the estimation method of the long-term noise indicators can be conditioned by the adopted criterion of the error of estimation of these indicators. The decision-making process can be conditioned by the estimation accuracy of the analysed noise indicators, or by the required uncertainty interval related to this process. However, settlement of this problem requires broader investigations.

Acknowledgments

This work was supported by the Polish Ministry of Science and Higher Education from means on science in 2009–2011 (research project no. NN504345636).

References

- [1] *Directive 2002/49/WE of the European Parliament and of the Council of 25 June 2002*, relating to the assessment and management of environmental noise, Official Journal of the European Communities 18.07.2002.
- [2] *Guidelines for Developing of Acoustic Maps*, Ed. R.J. Kucharski, Institute of Environmental Protection, Warszawa 2006 (in Polish).
- [3] *Guide to the Expression of Uncertainty in Measurement*, Central Office of Measures, Warszawa 1999 (in Polish).
- [4] C.G. Don, I.G. Rees, *J Sound Vibration* **100**, 41 (1985).
- [5] T. Wszolek, M. Kłaczyński, *Arch. Acoust.* **31**, (supplement), 311 (2006).
- [6] W. Batko, B. Stępień, in: *35th Winter School on Vibroacoustical Hazards Suppressions*, Ed. M. Roczniak, Upper Silesian Division of the Polish Acoustical Society, Gliwice 2007, p. 5.
- [7] M. Gałuszka, in: *Monitoring of Environment 2010*, Kraków 2010, full text on CD-ROM.
- [8] D. Gamerman, H.F. Lopes, *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*, Chapman & Hall/CRC, Boca Raton 2006.
- [9] M. Pipień, *Bayesian Inference in Financial Econometrics*, Wydawnictwo Akademii Ekonomicznej w Krakowie, Kraków 2006 (in Polish).
- [10] J.V. Candy, *Bayesian Signal Processing: Classical, Modern, and Particle Filtering Methods*, Wiley, Hoboken 2009.
- [11] J. Osiewalski, *Applications of Bayesian Econometrics*, Wydawnictwo Akademii Ekonomicznej w Krakowie, Kraków 2001 (in Polish).
- [12] M. Kostrzewski, *Bayesian Analysis of Financial Time Series Modeled by Diffusion Processes*, Uczelniane Wydawnictwa Naukowo-Dydaktyczne AGH, Kraków 2006 (in Polish).
- [13] Cz. Domański, K. Pruska, W. Wagner, *Statistical Inference for Non-Classical Assumptions*, Wydawnictwo Uniwersytetu Łódzkiego, Łódź 1998 (in Polish).
- [14] M. Rosenblatt, *Ann. Math. Statist.* **27**, 832 (1956).
- [15] E. Parzen, *Ann. Math. Statist.* **33**, 1065 (1962).
- [16] P. Kulczycki, *Kernel Estimators for Analysis of Systems*, WNT, Warszawa 2005 (in Polish).
- [17] W. Batko, B. Stępień, *Arch. Acoust.* **34**, 295 (2009).