Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Sound and Vibration 268 (2003) 167–175
JOURNAL OF SOUND AND VIBRATION
www.elsevier.com/locate/jsvi

Determination of uncertainty in environmental noise measurements by bootstrap method

F.A. Farrelly*, G. Brambilla

Istituto di Acustica "O. M. Corbino", Consiglio Nazionale delle Ricerche, Via del Fosso del Cavaliere, 100, 00133 Rome, Italy

Received 13 August 2002; accepted 26 November 2002

Abstract

An effective method for real-time evaluation of confidence intervals associated to quantile (L_q) and equivalent (L_{eq}) levels in environmental noise measurements is presented. The non-parametric surrogate data (or *bootstrap*) method, is described here in its basic form, valid for independent and identically distributed data, but is readily extendible to the treatment of dependent data. Application to actual measurements are shown which illustrate the practical effectiveness of real-time error evaluation during environmental monitoring.

1. Introduction

In environmental noise measurements, a multitude of independent signals from different sources (transportation, industries, etc.), quite possibly located in different places, contribute to form the noise under analysis. Due to the uncorrelated nature of this multitude of acoustic sources a statistical description of this noise is appropriate; this is somewhat different from the situation in which a signal $S(t)$ is combined with noise $N(t)$. In this latter case, the aim of a measurement might be to determine the signal root mean square $\langle S(t)^2 \rangle$ and the noise is an unwanted nuisance. However, in environmental noise measurements, the situation is reversed as it is the noise component $N(t)$ whose parameters are of interest, while the signal component $S(t)$ is absent. This interpretation leads to a model in which the observed noise ($N_0(t)$) is a particular realization of a stochastic variable whose distribution can be characterized by some appropriate parameters. It is common practice to express the properties of environmental noise distributions through the use of evaluation indices such as *equivalent sound level* L_{eq} and *quantile levels* L_q. The

*Corresponding author. Tel.: +39-0649-934-480.
E-mail address: francis@idac.rm.cnr.it (F.A. Farrelly).

0022-460X/03/$ - see front matter © 2003 Elsevier Ltd. All rights reserved.
doi:10.1016/S0022-460X(03)00195-0

determination of the values of these parameters is the aim of this type of measurement; due to the statistical nature of the model a particular measurement can only lead to an estimate of these parameters whose true values are known only to Mother Nature. This statistical interpretation of environmental noise measurements is consistent with international standards and many national legislations which require quantitative estimates of the errors associated with a noise measurement. Specifically, it is important to quantify the range in which a parameter might vary with a specific likelihood; this is generally expressed by confidence intervals. Assigning a probability p, the upper $L(p)^{sup}$ and lower $L(p)^{inf}$ limits of the confidence interval are determined in such a way that the chance of obtaining a value for measurement of the parameter L, with a value within this interval $[L(p)^{inf}, L(p)^{sup}]$, is p.

Most of the current instrumentation for field sound level measurements do not evaluate the uncertainties inherent in the evaluation indices. These uncertainties intrinsic of the noise itself are worth investigating not only in that international standards require their assessment but also as this may lead to a deeper understanding of the structured features of the stochastic process depicted by the observed noise measurement. Furthermore, by integrating the noise measurement with real-time information on confidence limit intervals, this method may prove to be a valuable aid in practical measurements where the assessment of error-trends may enable one to reduce measurement times as well as revealing anomalous acoustic events.

Online quality test procedures have already been implemented for the determination of confidence level limits of L_{eq} and L_q indices [1], the approach being based on establishing a relationship between the variance of a partitioned subset of the distribution and the first two moments of the ‘‘crossing up’’ (and down) statistics obtained from the data-set values.

This paper describes a different approach, a simple bootstrap method, in which the data set itself is used as a model of the distribution from which the confidence intervals for L_{eq} and L_n are determined. This implementation facilitates a numerically efficient procedure suitable for real-time assessment of the intervals.

Statistical analysis of linear and non-linear time series through the use of bootstrap (also known as surrogate data) methods has been the subject of many recent efforts. In particular, the problem of determining confidence limit regions of time series data has gone through several refinements since its first acceptance [2], particularly for what concerns the treatment of data that are dependent. Initially, this was done by introducing subseries analysis in the block bootstrap [3] then with autoregressive modelling in the sieve bootstrap method [4] which is particularly adapt in capturing second order dependencies in the data set and effectively applied to confidence region estimation [5]. Perhaps the most versatile method [6] is that in which the bootstrap is performed in Fourier space so that the spectral components of the signal are preserved in the surrogate realizations. Although the straight bootstrap method described in this paper is not entirely appropriate when dealing with dependent data [7], it serves as an initial attempt at producing a numerically efficient system which can be quite easily modified in the future by using one of the aforementioned techniques.

2. Confidence intervals of sound level measurements

As described in the introduction, environmental noise very often occurs in the form of randomly fluctuating sound signals. To quantitatively describe this phenomenon, noise indices such as the equivalent sound pressure level L_{eq} and quantile levels L_q are widely used and these are usually expressed in decibels (dB) relative to a reference pressure of 20 μPa. In practice, when performing a measurement with a sound level meter, the acoustic sound pressure level $p(t)$ is transformed into a discrete set of equivalent levels $L_{st}(t_i)$ given by

$$L_{st}(t_i) = 20\log_{10}\frac{\langle |p(t_i)|\rangle_\tau}{p_0}, \tag{1}$$

where $\langle |p(t_i)|\rangle_\tau$ denotes the time average of the absolute value of $p(t)$ in the interval $[t_i, t_i + \tau]$ and p_0 is the reference pressure. The duration of the interval τ is one of the measurement parameters typically lasting at least $\frac{1}{8}$ of a second. Analogously, the equivalent level $L_{eq}(T)$ is the average energy over a period of time T and is determined by evaluating Eq. (1) over this period of time ($\tau = T$). In practice, this is approximated by evaluating the average of the n data-set elements $\langle E(t_i)\rangle_\tau$ which encompass time T ($T = n\tau$),

$$L_{eq} \approx 10\log_{10}\frac{1}{n}\sum 10^{L_{st}(t_i)/10}, \tag{2}$$

whereas, the quantile level $L_q(T)$ is the element in the set of $L_{st}(t_i)$ exceeded by a fraction q (often expressed as a percentage) of the elements which occur in T.

In general, an experiment designed to determine the value of a parameter L_{true} will do this by applying an appropriate transformation to the measured data set D_0. The value obtained for the parameter L_0 for this data set will probably differ from the true one due to the effect of errors throughout the experiment chain and in the physical phenomenon under study. In most physical experiments there will be a random component affecting the data set so that even repeating it under identical stationary conditions, which can be viewed as re-extracting from the distribution describing the physical measurement D, different data-set realizations D_i will be formed. This implies that L_0 is simply an extraction from a set of possible values associated to the other possible data-set realizations which could have been obtained. The distribution of $L_0 - L_i$ is an estimate of $L_{true} - L_i$, very often, the best one available.

Taken together, the elements $L_{st}(t_i)$ occurring in T form D_0, while the index being analyzed, L, is either $L_{eq}(T)$ or a specific $L_q(T)$. Note that the transformations applied to D_0 to obtain L are intimately different in the two cases; in the latter case to obtain $L_q(T)$ a selection of an element of D_0 is performed, while in the former Eq. (2) is applied.

The significance, from a statistical point of view, of the values determined for these indices is given in terms of confidence limits, and in order to evaluate these, the distribution of $L_{true} - L_i$ is needed. In the case in which D has a finite variance and its elements are independent, the central limit theorem assures that the distribution of their mean will tend to a Gaussian distribution as the number of elements in D_0 increases; this greatly simplifies the determination of the confidence limits of $L_{eq}(T)$ through the use of the standard error [8]. However, the selection procedure required to determine $L_q(T)$ will not necessarily converge quickly to a Gaussian distribution, particularly for values of q near its upper and lower bounds, so that a different approach is needed.

In order to tackle the problem of determining the confidence limits of quantile levels a method more sophisticated than the standard error approach should be made. If an accurate picture of the physical processes that occur during the measurement chain is available the latter can be determined through a Monte-Carlo simulation which synthesizes many runs of the experiment with the corresponding data sets D_i^s from which their respective indices L_i^s can be evaluated. However, this is in general a very difficult task to undertake.

2.1. Bootstrap method implementation

When little is known of the underlying physical process involved in the measurement chain a modified version of the Monte-Carlo simulation technique, known as the bootstrap or surrogate data method, can be of great use [9]. The key aspect of this technique is to perform the Monte-Carlo simulation on the basis of the measured data set's (D_0) discrete distribution, this being an estimator of the actual data-set distribution D_{true}. An alternative description of this method can be given by considering the cumulative of the data set D_0 as the staircase function which is a best-fitting function $y(i)$ of D_0's distribution; the resultant chi-squared,

$$\chi^2 \equiv \sum_{i=1}^{n} \left(\frac{y_i - y(i)}{\sigma_i} \right)^2, \tag{3}$$

is of course 0. The mathematical model of D_{true} is this staircase function and it is used for the Monte-Carlo simulation. This formulation requires that the measured data set be composed of independent and identically distributed points (IID). In practice a synthesized data set D_i^s realized by extracting randomly n times from the n elements of D_0 so that each element can be drawn once, more than once, or never. This procedure is repeated for each realization until a sufficient number (m) of synthesized data sets are obtained for the Monte-Carlo simulation. For each data set D_i^s realization, the specific quantile level $L_q^i(T)$ (or equivalent level $L_{eq}^i(T)$) is computed so that this parameter's distribution can be determined and from this its confidence limits can be found.

Fig. 1 shows the logical structure of an acoustic monitoring measurement in which the gray colored elements represent the traditional phases and the white boxes are the additional phases used in the bootstrap method's implementation. The sound level meter (gray oval) feeds data to an acquisition system which records the time series data set D_0 (gray box) from which the measurement parameters L_0 are determined (gray box).

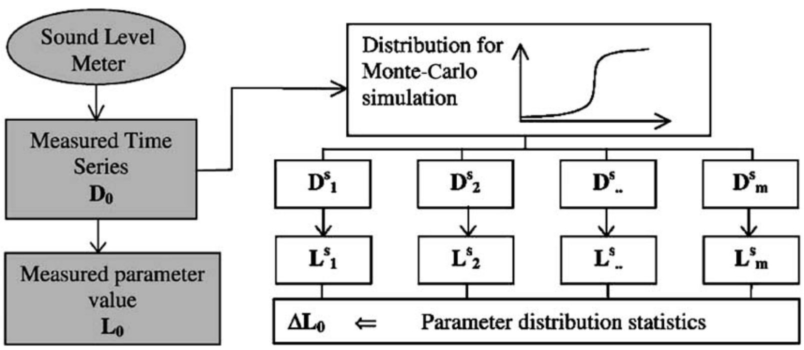

Fig. 1. The logical structure of an acoustic noise monitoring measurement (gray boxes) together with additional phases (white boxes) required to determine confidence limits using the bootstrap implementation.

The bootstrap method is used in order to calculate the confidence interval of the measurement by analyzing the n elements of the data set acquired so far. This data is used as the estimate of the distribution (white box) required by the Monte-Carlo method for creating the synthetic data sets D_i^s; an application written in the C programming language, implements the random extraction from the elements in D_0 creating m different data sets (small white boxes) each contains n elements. The number of data sets m is determined by the probability p associated to confidence limit interval so that

$$m = int(K/p), \tag{4}$$

where K is a pre-factor; as can be seen, the complexity of the algorithm grows inversely to the probability level chosen. For each of these synthesized data sets the value of parameter being analyzed L_i^s is determined resulting in a set of m elements (small white boxes) representing the outcome of the Monte-Carlo realizations. In the case in which the equivalent level is the parameter under analysis (L_i^s), Eq. (2) is computed with the $L_{st}(t_i)$ substituted by their synthesized analogues, whereas, for quantile levels, a selection procedure is applied to the elements of D_i^s so as to determine its value for the specific data set realization. In either case the complexity of the parameter determination is linear with respect to the number of elements ($O(n)$, see Ref. [10]). The final phase of the procedure (wide white box) determines the upper $L(p)^{sup}$ or lower $L(p)^{inf}$ confidence limit by selecting the kth highest or lowest element from the m-element parameter set L_i^s where k is determined by the probability p associated to confidence level $p = 2k/m$. The width of this interval ($\Delta L(p)$) is just the difference between these two: $\Delta L(p) = L(p)^{sup} - L(p)^{inf}$.

3. Results

The potential of the surrogate data method for determining the confidence limits of the equivalent and quantile levels is illustrated by showing examples both from synthesized data and from actual measurements of road traffic noise. Once the confidence level probability p is chosen, the upper (*sup*) and lower (*inf*) limits of the confidence interval are determined (by the procedure outlined above) in such a way that the chance of finding the parameter, with a value within this interval, is p.

In Table 1 the results of a Gaussian noise test are given in which 100 runs of synthesized noise, were generated having a mean of 70 dB, a relative standard deviation of 0.1 and containing 500 samples. This data is converted into its L_{st} equivalent and processed as though it had been acquired from a sound level meter. For each parameter its value, lower (*inf*) and upper (*sup*) confidence limits, associated to a probability $p = 80\%$ are given along with their standard deviation (Δ) for this sample set of runs. These results show excellent agreement between the bootstrap method technique and the values expected by statistical theory.

These simulations were performed on a PC with an Athlon XP processor™ running at 1533 MHz on a Linux operating system using the gcc 2.96 compiler. Timings were obtained from the average of 10 consecutive runs in which L_{st} values were read from a text file and four indices together with their respective lower and upper limits were calculated ($L(p)_5$, $L(p)_{eq}$, $L(p)_{50}$ and $L(p)_{95}$ with $p = 80\%$). In the case of 500 points, 62 ms were required per run, whereas for 5000 points 597 ms were needed. In all cases, these values were obtained from the unix *time* command

and indicated the total of *user* and *system* time for the process. Even in the case of a relatively large set of points (5000) this modest computational time is compatible with real-time operation.

Data taken from real applications are analyzed next; the first case is taken from a time series with fairly stationary conditions while the second will illustrate widely fluctuating noise. In Fig. 2 a typical environmental noise measurement is presented in the bottom graph (c) showing the short-time equivalent level L_{st} acquired by the sound level meter (at a rate of 1 s^{-1}) together with the calculated running L_{05}, L_{eq} and L_{95} levels. The relatively small variations in these parameters after the first 50 or so seconds is an indication of the time-independent nature of this data set.

Table 1
Gaussian noise test summary (100 runs)

Parameter	Value $\pm \Delta$	*inf* $\pm \Delta$	*sup* $\pm \Delta$	Predicted
L_{95}	69.22 ± 0.06	69.16 ± 0.06	69.28 ± 0.05	69.22
L_{eq}	70.00 ± 0.02	69.98 ± 0.02	70.03 ± 0.02	70.00
L_{50}	70.01 ± 0.02	69.97 ± 0.03	70.04 ± 0.03	70.00
L_{05}	70.67 ± 0.03	70.63 ± 0.04	70.72 ± 0.04	70.66

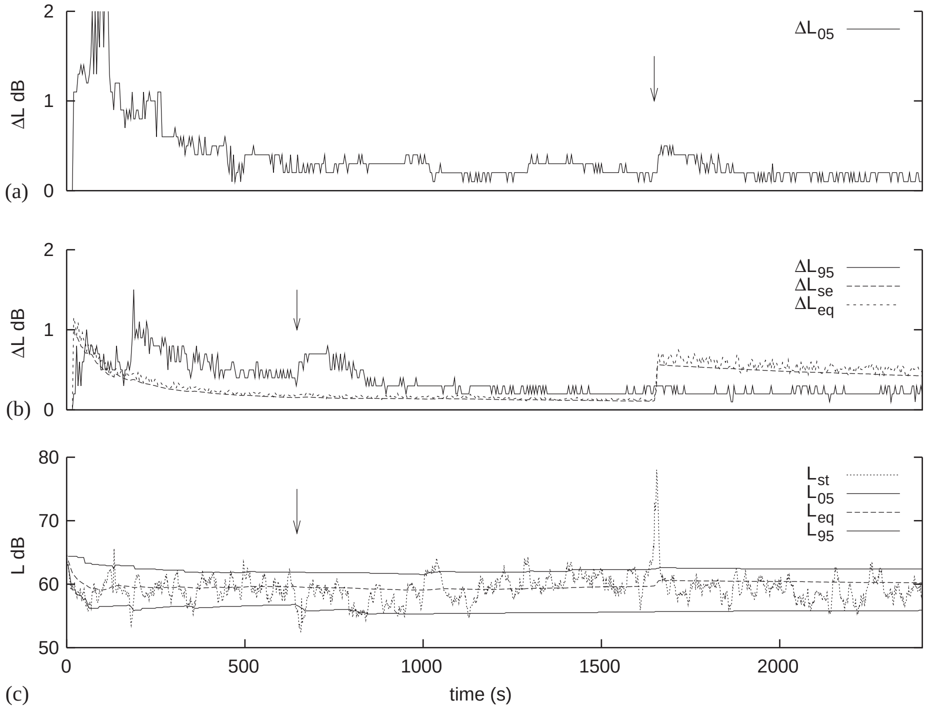

Fig. 2. (c) Short-time equivalent level L_{st} data set acquired by a sound level meter (at a rate of 1 s^{-1}) together with the calculated running L_{05}, L_{eq} and L_{95} levels. (b) Confidence interval for L_{eq} determined by bootstrap and by standard error ΔL_{se} computed once the first 32 values are available; (a) the same shown for top 5% ΔL_{95}.

A pronounced peak in the L_{st} is visible at $\sim$1650 s while the arrow points to a noticeable dip at $\sim$650 s.

The application of the bootstrap method for the determination of the confidence limit width ΔL_{95} ($p = 80\%$) of the 95th quantile level of this data, once 32 data points are available, results in the solid line in the middle graph (b). The arrow indicates the moment in which the dip in the L_{st} abruptly lowers the L_{95} level and noticeably increases the confidence interval width of this level. It can be seen that this solid line assumes discrete values in ΔL_{95} due to the fact that the acquired L_{st} are also discrete and the selection procedures used to determine the quantile confidence limits in the bootstrap method guarantees that both $L(p)^{sup}$ and $L(p)^{inf}$ are elements of the original data set.

The dashed line, also in this middle graph, shows the equivalent level's confidence interval determined directly by way of the standard error ΔL_{se}. For comparison, this same interval, calculated with the bootstrap method, is shown by the dotted line ΔL_{eq}; clearly, the two curves are very similar apart from a small noise effect in the latter. After the peak at $\sim$1650 s, the width as calculated by the bootstrap method is slightly larger than that determined through the standard error; as mentioned above, the latter method requires the data to be distributed normally and the presence of this peak represents an outlier which is not correctly accounted for under this hypothesis.

The top graph (a) in Fig. 2 corresponds to the confidence interval breadth of the parameter measuring the level of the highest 5% of the signals. It is unperturbed by the dip in the L_{st} which affected the ΔL_{95} but correctly changes at the moment in which the sharp peak occurs.

The effect caused by greatly fluctuating time series is considered next. The measurement site is near a one-way throughway and is characterized by a quiet background with occasional vehicle noise which show up as isolated peaks in the L_{st}.

In the lower graph of Fig. 3(b), the time series of the short-time equivalent levels L_{st} acquired by the sound level meter, at a rate of 8 s^{-1}, are shown together with the calculated running L_{eq}, L_{95} and L_{05} levels. The upper graph (a) in this figure portrays the width of the confidence limit interval ΔL_{eq} corresponding to the running L_{eq} in (b) and is calculated once the first 32 values of L_{st} (4 s) are available. An abrupt increase to 2.5 dB shows up after 29 s due to a peak in the L_{st} time series and it decays slowly to 0.8 dB when a second, louder event presents itself at 97 s. The short decay which follows, lasts just 10 s before a further event is encountered, from this point on ΔL_{eq} settles to a value of about 1.8 dB from where it decays very slowly reaching 0.7 dB at 360 s. Even though 17 other loud peaks can be counted they do not show up as distinct decays the way the first two did. Referring to part (b) of the figure, the running L_{eq} which is about 43 dB at 28 s, increases as each peak is encountered and reaches 66.2 dB at the end of the 375 s over which the measurement was made. The changes in L_{95} are small during the entire period and its final value of 39 dB coincides with its first one; likewise, L_{05} remains practically stationary to close to its final value of 74 dB, once the 5th peak is reached.

4. Conclusions

A method for establishing the confidence limit interval of L_{eq} and L_q parameters in real-time environmental noise monitoring systems has been presented. Simulations using synthesized IID

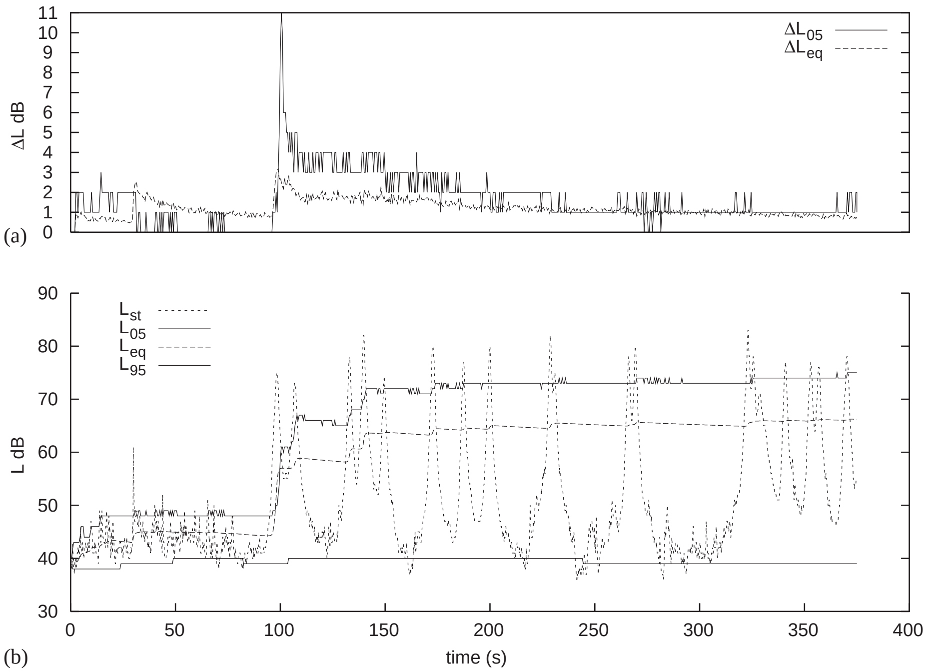

Fig. 3. Data set acquired at a rate of 8 s^{-1}, exhibiting numerous outliers; (b) L_{st} and calculated running L_{05}, L_{eq} and L_{95} levels. (a) ΔL_{eq} and ΔL_{eq} computed once 32 values are available.

data have been performed, and performance characteristics were tested, indicating that it is possible to determine the confidence intervals of four different indices of a 500 point data set at a rate of about 16 times/s. Application to real-life data sets shows that the onset of infrequent outliers abruptly increases the confidence interval of the appropriate quantile levels, unless their occurrence has been sufficiently numerous.

The implementation of the bootstrap method described in the paper is, in a strict sense, capable of dealing correctly only with IID data sets; however several methods of overcoming this limitation are known and will be the subject of further investigation. Specifically, block, sieve and Markov bootstrap methods can be considered as valid candidates as well as spectral surrogate data selection techniques.

References

[1] A. Heiss, A supplement of sound measurement technique by an online quality test procedure for the indices L_{EQ} and $L_{x\%}$, Journal of Sound and Vibration 241 (1) (2001) 29–39.
[2] B. Efron, R. Tibshirani, Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy, Statistical Science 1 (1986) 54–77.
[3] A.C. Davison, P. Hall, On studentizing and blocking methods for implementing the bootstrap with dependent data, Australian Journal of Statistics 35 (1993) 215–224.
[4] P. Bühlmann, Sieve bootstrap for time series, Bernoulli 3 (1997) 123–148.

[5] E. Choi, P. Hall, Bootstrap confidence regions computed from autoregressions of arbitrary order, Journal of the Royal Statistical Society B 62 (Part 3) (2000) 461–477.
[6] T. Schreiber, A. Schmitz, Surrogate time series, Physica D 142 (2000) 346.
[7] K. Singh, On the asymptotic accuracy of Efron's bootstrap, Annals of Statistics 9 (1981) 1187–1195.
[8] N.C. Barford, Experimental Measurements: Precision, Error and Truth, 2nd Edition, Wiley, New York, 1990, p. 38.
[9] B. Efron, Bootstrap methods: another look at the jackknife, Annals of Statistics 7 (1979) 1–26.
[10] R. Sedgewick, Algorithms, 2nd Edition, Addison-Wesley, Reading, MA, 1988.