

Toward a Probabilistic Acoustic Emission Source Location Algorithm: a Bayesian Approach

Thomas Schumacher^{a,*}, Daniel Straub^b, and Christopher Higgins^c

^a *Civil and Environmental Engineering, University of Delaware, Newark, DE, USA*

^b *Engineering Risk Analysis Group, Technische Universität München, Germany*

^c *School of Civil and Construction Engineering, Oregon State University, Corvallis, OR, USA*

Abstract

Acoustic Emissions (AE) are stress waves initiated by sudden strain releases within a solid body. These can be caused by internal mechanisms such as crack opening or propagation, crushing, or rubbing of crack surfaces. One application for the AE technique in the field of Structural Engineering is Structural Health Monitoring (SHM). With piezo-electric sensors mounted to the surface of the structure, stress waves can be detected, recorded, and stored for later analysis. An important step in quantitative AE analysis is the estimation of the stress wave source locations. Commonly, source location results are presented in a rather deterministic manner as spatial and temporal points, excluding information about uncertainties and errors. Due to variability in the material properties and uncertainty in the mathematical model, measures of uncertainty are needed beyond best-fit point solutions for source locations. This paper introduces a novel holistic framework for the development of a probabilistic source location algorithm. Bayesian analysis methods with Markov Chain Monte Carlo (MCMC) simulation are employed where all source location parameters are described with posterior probability density functions (PDFs). The proposed methodology is applied to an example employing data collected from a realistic section of a reinforced concrete bridge column. The selected approach is general and has the advantage that it can be extended and refined efficiently. Results are discussed and future steps to improve the algorithm are suggested.

Keywords: Acoustic Emission, Stress waves, Probabilistic source location algorithm, Parameter estimation, Bayesian analysis, Markov Chain Monte Carlo simulation.

* Corresponding author. *E-mail address:* schumact@udel.edu.

1 Introduction

1.1 Background on Acoustic Emission

Acoustic Emission (AE) is the ‘term used for transient elastic waves generated by the release of energy within a material or by a process’ [1]. A stress wave is generated, traveling from the source origin away to the surface of the body where it can be recorded by sensors. This process is irreversible and therefore not repeatable. Other terms include ‘stress wave emission’ or ‘nano-seismic activity’. Within the family of non-destructive testing methods, AE has its own special place because it is a passive technique. AE are stress waves produced by a spontaneous internal dynamic process such as crack initiation and propagation or internal material fracture. The source location and mechanism is therefore unknown and is the subject of investigation. An introduction to the many applications in AE research can be found for example in [2, 3]. Subsequently, the terms Acoustic Emission (or AE) and stress wave are used interchangeably.

In infinite elastic solids, two types of waves, sometimes referred to as modes, exist. The first is called compression, or primary (p -) wave where the particles move in the direction of the wave. The second is called shear, or secondary (s -) wave with particle motion perpendicular to the direction of the propagating wave. With the introduction of boundaries, a third wave type, the surface, or Rayleigh (R -) wave exists. Between the three wave mode velocities c_p , c_s , and c_R , the following relationship is true [4]:

$$c_p > c_s > c_R \quad (1)$$

The p -wave velocity in an isotropic, homogeneous and elastic body can be calculated as:

$$c_p = \sqrt{\frac{E}{\rho} \frac{1-\nu}{(1+\nu)(1-2\nu)}} \quad (2)$$

where E is the dynamic Modulus of Elasticity in N/m^2 , ρ the material density in kg/m^3 , and ν Poisson’s Ratio [4]. Typical average values for the p -wave velocity c_p and Poisson’s Ratio ν for vintage bridge reinforced concrete range from 3.5 to 4.5 m/ms (140 to 180 in./ms) and 0.25 to 0.3, respectively [5].

Stress waves emitted from AE sources are usually recorded by a network of piezo-electric sensors mounted to the surface of the specimen. The analog signal is intensified by an amplifier and digitized to enable hard disk storage. AE data is normally not acquired continuously, since that would produce enormous data files which would be difficult to analyze and interpret. Instead, pre-selected criteria are used to trigger the system for individual burst signals from which p -wave arrival times as well as qualitative wave form parameters are then extracted [2].

The first important step in quantitative AE analysis is the estimation of spatial and temporal parameters of the stress wave source. Typically only the p -wave arrival times are employed because they represent the first, undisturbed arrival of a stress wave and are thus the easiest to deduce. If at least four sensors detect a discrete stress wave signal it can be identified as an AE event and spatial and temporal parameters of the stress wave source can be estimated. An AE event is defined as the ‘physical phenomenon giving rise to acoustic emission’ [1]. Fig. 1 illustrates a typical recorded transient stress wave signal recorded at sensor k . The first arrival of the stress wave, or p -wave arrival time, is enlarged and stretched in time for better clarity. The source occurrence time is unknown and denoted as t_0 . For the arrival time of the p -wave, two values are shown: $t_{a[k],man}$ denotes the arrival time picked manually, and $t_{a[k],auto}$ by some automated picking method. A denotes the maximum signal amplitude. Depending on the method employed and the signal-to-noise ratio of the waveform, an off-set Δt from the manual pick will be inherent in the detected arrival times and is often unavoidable.

The first crucial step for any reliable AE source location estimation is the accurate deduction of the p -wave arrival times from each sensor in the network. In the present work, a picker based on the Akaike Information Criterion (AIC) was employed as described in [6]. This method has proven reliable and accurate in picking arrival times from stress waves [5, 7].

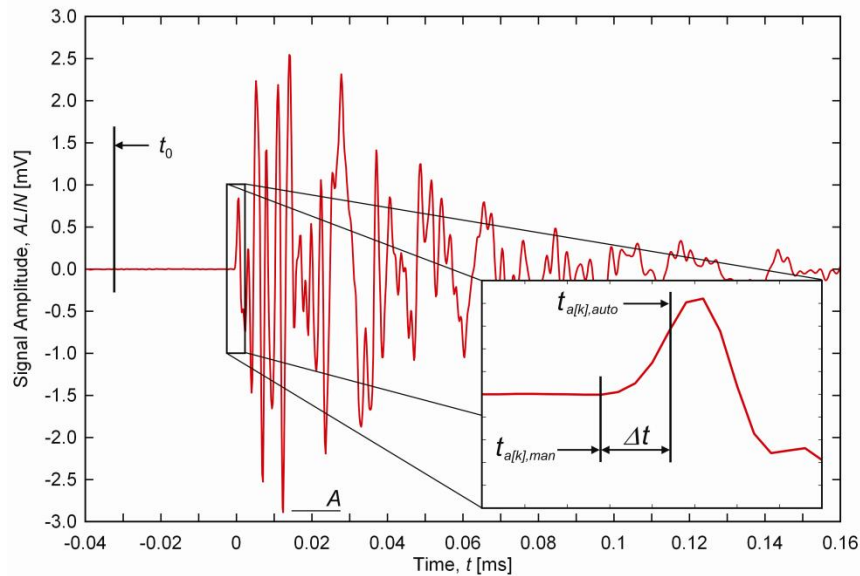


Fig. 1. AE burst signal example with enlarged p -wave arrival at sensor k

In sections 1.2 and 1.3, the basics of traditional source location algorithms as well as a short literature review of earlier probabilistic approaches are presented, respectively.

1.2 Traditional Source Location Algorithms

Traditional methods shall be defined as deterministic methods that do not directly take into account variability of parameters, i.e. locations of sensors, p -wave velocity, p -wave arrival times, etc. are assumed as constants, i.e. deterministic. If the source hypocenter and sensors are modeled as dimensionless points and the travel path of the stress wave is assumed as a straight line between source and sensor, and the p -wave velocity c_p assumed as constant and known for all travel paths, the following general relationship, sometimes referred to as *arrival time function*, for one event can be derived:

$$f(x_0, y_0, z_0, t_0) = t_{a[k]} = t_0 + \frac{d_{[k]}}{c_p} \quad (3)$$

where t_0 is the event source time and x_0 , y_0 , and z_0 are the event source coordinates. The signal travel distance $d_{[k]}$ between source and sensor locations is calculated as:

$$d_{[k]} = \sqrt{(x_{[k]} - x_0)^2 + (y_{[k]} - y_0)^2 + (z_{[k]} - z_0)^2} \quad (4)$$

The Cartesian coordinates of sensor k are denoted with $x_{[k]}$, $y_{[k]}$, and $z_{[k]}$. To solve for the four unknowns x_0 , y_0 , z_0 , and t_0 , time arrivals $t_{a[k]}$ from at least four sensors must be detected. Most commonly, this non-linear inverse problem is solved by employing iterative algorithms based on Geiger's method [8]. A summary of existing deterministic schemes can be found for example in Ge [9, 10]. Examples of successfully implemented source location algorithms in the area of concrete research can be found in [11, 12]. These algorithms provide a covariance matrix of source coordinates as an estimate of uncertainty if arrival times from at least five sensors are available. From that, principal standard deviations can be deduced that give a measure of inconsistency in the several measured arrival times. This approach does not account for all intrinsic uncertainties and errors in the model and systematic errors can only be revealed when the true source location is known, but this is usually not the case for AE applications.

1.3 Bayesian Probabilistic Source Location Algorithms

In earthquake source location, probabilistic methods have been employed for many years [13-15]. Tarantola and Valette [13] presented a probabilistic formulation to solve nonlinear inverse problems with a finite number of parameters, such as the recovery of hypocenter source locations. A Bayesian approach was suggested to combine experimental observation of data, prior information of parameters, and the theoretical relationship. The idea is to incorporate errors and uncertainties by assigning probability density functions (possibly non-Gaussian) to parameters as well as theoretical relationships. An example with actual observed arrival time data is presented for 11 stations (or sensor locations). The spatial coordinates of the stations were modeled as deterministic and all parameters Gaussian distributed. The velocity model of the medium was assumed as uniform. The resulting depth plot showed that two local

maxima could be found. Reported were expected values and standard deviations of the source location coordinates.

In Lomax et al. [14], a Bayesian probabilistic approach to locate earthquake sources was presented, which is alternatively based on: 1) locally directed Metropolis-Gibbs sampling (a Markov Chain Monte Carlo (MCMC) technique) and 2) an exhaustive global grid search method. Both approaches are able to deal with non-uniform velocity models, i.e. layers with different material properties. Solutions are given as probability density functions in four dimensions, i.e. three spatial and the temporal. It is pointed out that only in this manner a complete understanding of the uncertainties of a location result can be gained. The performance of the proposed non-linear probabilistic source location algorithms are then compared to traditional linear schemes using synthetically generated arrival times.

Based on the above methods, a first approach to account for the probabilistic nature of AE source parameters was undertaken by Schechinger [7]. Therein, an adapted version of a computer program called *NonLinLoc* was used [16] which is based on the methods described in [14]. The mode of the posterior distribution corresponding to the point of maximum likelihood was used as estimator for the source location. Non-homogeneous material models that allow the existence of a nonlinear velocity field can be implemented as well. Model parameters cannot be estimated based on observations but must be specified by the user.

None of the previous work on probabilistic source location algorithms for AE has considered applying a Bayesian approach for estimating the model parameters of the algorithms using observed data. In the present paper, such an approach and its numerical implementation is presented. In chapter 2, a background on statistical modeling using Bayesian analysis is presented. Sources of uncertainty are explored and listed. Chapter 3 develops and explains the methodology for the proposed probabilistic source location algorithm. The framework is applied to an example and results visualized. A summary and conclusions are presented in chapter 4.

2 Statistical Modeling Using Bayesian Analysis Methods

2.1 Concepts of Bayesian Analysis

Bayesian analysis is the framework for assessing uncertainties and errors in the present study. Consider an event A that describes the parameters of interest and an event B that describes the observation (data). Bayes' rule describes how the probability of A changes upon observation of B :

$$\Pr(A|B) = \frac{\Pr(B|A)}{\Pr(B)} \Pr(A) \quad (5)$$

where $\Pr(A|B)$ is the (updated) posterior probability of event A given B ; $\Pr(B|A)$ is the probability of observing B given A and is known as the likelihood; $\Pr(A)$ is the probability of A before observing B , also called prior probability; $\Pr(B)$ represents a normalizing scalar. This rule can now be applied to a random variable \mathbf{X} having the PDF $f_{\mathbf{x}}(\mathbf{x}|\boldsymbol{\theta})$ with distribution parameters $\boldsymbol{\theta}$. In order to account for uncertainties in the parameters $\boldsymbol{\theta}$, they are modeled as random variables Θ . Once events are observed that provide information about the random variable \mathbf{X} , this new knowledge can be used to update the joint distribution of Θ by using Eq. 5 as follows:

$$f_{\Theta}(\boldsymbol{\theta}|\mathbf{x}_0) = \frac{L(\boldsymbol{\theta}|\mathbf{x}_0)}{c} f_{\Theta}(\boldsymbol{\theta}) \quad (6)$$

where $f_{\Theta}(\boldsymbol{\theta}|\mathbf{x}_0)$ represents the posterior distribution of Θ conditioned on observed outcomes \mathbf{x}_0 of the random variable \mathbf{X} , $L(\boldsymbol{\theta}|\mathbf{x}_0) = f_{\mathbf{x}}(\mathbf{x}_0|\boldsymbol{\theta})$ is the likelihood function of parameters $\boldsymbol{\theta}$ for observed realizations \mathbf{x}_0 , c is a normalizing constant that enforces the area underneath the posterior distribution to be equal to 1 and is defined as the m -fold integral:

$$\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} L(\mathbf{x}_0|\boldsymbol{\theta}) f_{\Theta}(\boldsymbol{\theta}) d\theta_1 \dots d\theta_m \quad (7)$$

where m is the number of parameters and $f_{\Theta}(\boldsymbol{\theta})$ is the prior distribution for Θ before additional knowledge is gained.

Once the posterior distribution $f_{\Theta}(\boldsymbol{\theta}|\mathbf{x}_0)$ for the parameters is obtained, it can be incorporated in a general predictive joint distribution of the form:

$$f_{\mathbf{x}}(\mathbf{x}|\mathbf{x}_0) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f_{\mathbf{x}}(\mathbf{x}|\boldsymbol{\theta}) f_{\Theta}(\boldsymbol{\theta}|\mathbf{x}_0) d\theta_1 \dots d\theta_m \quad (8)$$

to predict future events \mathbf{X} based on knowledge from past observed realizations \mathbf{x}_0 .

The difficulty in Bayesian inference lies in computing the integrals required for the evaluation of the constant c (Eq. 7) and the computation of the predictive distribution (Eq. 8) for which analytical solutions exist only in special cases. Numerical methods using Markov Chain Monte Carlo (MCMC) simulation techniques represent a powerful method to address this problem and are introduced in the following section.

2.2 The Numerical Approach

In the present study, MCMC simulation is employed to perform Bayesian inference. An introduction to these methods can be found in [17]. Using MCMC, it is possible to directly sample from posterior distribution (Eq. 6) without knowing the proportionality constant c . This is achieved by defining the samples as a Markov Chain whose stationary distribution is equal to the sought posterior distribution $f_{\theta}(\theta | \mathbf{x}_0)$. After a sufficient number of steps, the chain becomes independent of its initial state θ_0 and converges to this stationary distribution. The sampling of θ_{i+1} conditioned on θ_i is based on the Gibbs sampling technique, a special form of the Metropolis-Hastings algorithm [18]. The computer program utilized for the present study is called *WinBUGS* and available free from the internet [19]. In the present study, *WinBUGS* was employed to (1) estimate the model parameters based on observed experimental data and to (2) predict future events using newly available data as explained in sections 3.2 and 3.3, respectively.

2.3 Experimental Data

The proposed methods are demonstrated using a calibration experiment of a large size reinforced concrete specimen as illustrated in Fig. 2 using pencil lead breaks (PLB) applied to the surface. Five replicate PLBs to initiate stress waves were performed on 76 x 76 mm (3 x 3 in.) grids on the front x - y and top x - z face, and the bottom of the pull-out rebar, and are shown as grey crosses in Fig. 2(a). Note that the majority of PLBs lie within or close by the sensor array. Sources far away from the sensor array are typically not well constrained and errors in the input parameters are amplified [5]. It can thus be expected that the variability of the prediction will be larger for sources outside the sensor array, but this uncertainty can be consistently quantified using the Bayesian method. PLBs represent a common means to produce highly repeatable stress wave sources in AE [1]. Eight piezo-electric resonant-type AE sensors with peak response frequencies of 150 kHz were deployed non-symmetrically on four faces of the specimen and are shown as yellow cylinders in Fig. 2(a). These sensors have a frequency range of approximately 50 to 500 kHz and have been found to work well for concrete specimens of specimens with this size [5, 20]. Sensor location coordinates are listed in Table 1.

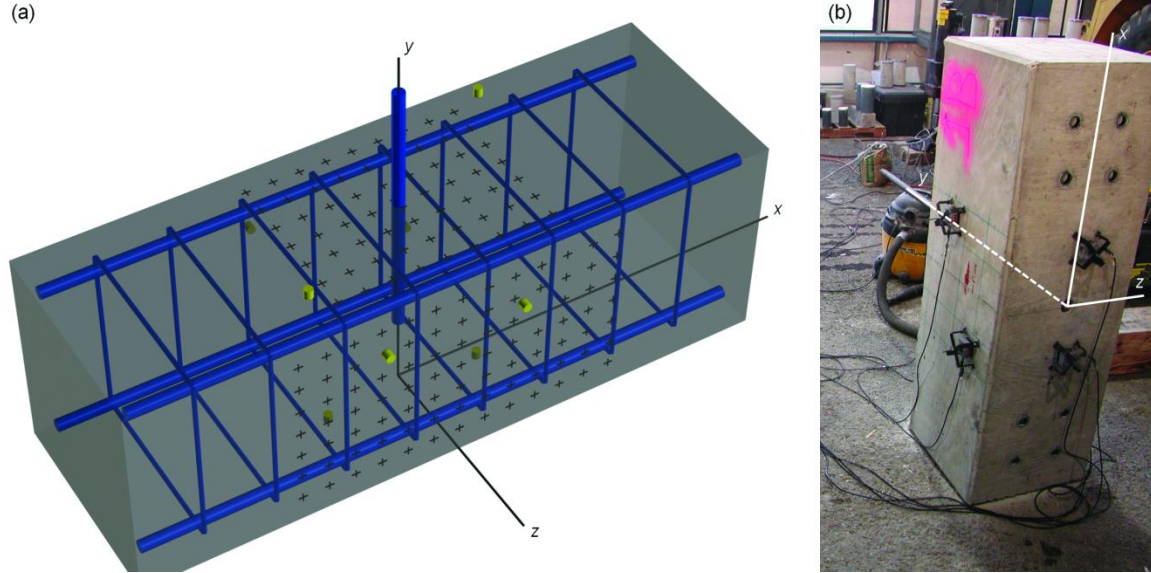


Fig. 2. View of reinforced concrete test specimen:

- (a) Graphical model (semi-transparent to visualize embedded reinforcing bars)
- (b) Photo showing sensors 3 and 4 (bottom x - z face) and 5 and 6 (back x - y face)

The test specimen represents a full-scale reinforced concrete section of a bridge bent cap column with dimensions $l_z \times l_y \times l_x = 610 \text{ mm} \times 610 \text{ mm} \times 1.83 \text{ m}$ (24 in. x 24 in. x 6 ft). This specimen was originally designed to determine the ultimate anchorage force of the embedded $\phi 35 \text{ mm}$ (#11) reinforcing bar aligned with the y -axis [21]. The embedment length of this bar was 406 mm (16 in.). Other steel reinforcing bars are shown in blue (Fig. 2(a)) for reference. All reinforcing bars were Grade 60 ($f_y = 414 \text{ MPa}$) and the average concrete compressive strength was found to be 28.0 MPa (4060 psi).

Table 1. Sensor locations

Sensor [-]	x		y		z		Location on face [-]
	[mm]	[in.]	[mm]	[in.]	[mm]	[in.]	
1	-216	-8.5	457	18.0	302	11.9	front x - y
2	178	7.0	432	17.0	302	11.9	front x - y
3	-229	-9.0	0	0.0	25	1.0	bottom x - z
4	229	9.0	0	0.0	25	1.0	bottom x - z
5	-254	-10.0	394	15.5	-311	-12.3	back x - y
6	216	8.5	178	7.0	-311	-12.3	back x - y
7	394	15.5	613	24.1	-254	-10.0	top x - z
8	-356	-14.0	613	24.1	178	7.0	top x - z

For computational efficiency, 31 PLB locations were selected and included in the regression analysis. The total number of observed p -wave arrival times t_a included was therefore: 31 locations x 5 PLBs per location x 8 sensors = 1240.

2.4 Model and Parameter Uncertainties

Any modeling of a physical phenomenon includes uncertainties on the true nature of a problem and model parameters are known only to a certain degree. This is also observed in the present problem. As an example, the specimen is assumed to be isotropic and homogeneous, but reinforced concrete is a composite material consisting of cement, aggregates, and embedded reinforcing steel, each constituent having significantly different wave propagation properties. In the present work, the specimen was modeled as a homogeneous isotropic elastic bulk solid. Homogeneity is assumed because the specimen size is large compared to the biggest constituents, i.e. the aggregates. The assumption of isotropy is justified by the fact that the reinforcing bars are local elements and rather small. Also, numerical simulation of wave propagation has shown that typical reinforcing bar diameters do not change the stress wave propagation characteristics significantly transverse to the reinforcing bar [7]. This was also investigated and verified experimentally by Lovejoy [20]. Along their length, however, reinforcing bars may act as wave guides and result in a faster apparent p -wave velocity in some cases. This is not yet accounted for in the present paper but will be investigated in the future. Furthermore, it is assumed that the p -wave mode is non-dispersive due to the specimen geometry. Although some dispersion may be present due to the concrete's constituents, it is reasonable to neglect this effect due to the fact that the wave travel path lengths for each sensor are comparable. The dispersive effect would have a similar distortion effect on all measured p -waves and therefore cancel out.

Fig. 3 shows a generic test specimen and visualizes the most important parameters involved in estimating stress wave source locations. Indices are assigned as follows and impose a distinct hierarchy in the model:

- i for pencil break locations (1 to 31)
- j for the event at each location (1 to 5)
- k for the sensors (1 to 8)

In Table 2, the most important parameters are listed and an attempt to classify their variability into epistemic (*epi*) or aleatory (*ale*) uncertainty is undertaken [22]. Epistemic uncertainties are those associated with limited knowledge of the problem. They can be decreased by for example refining the mathematical model or gathering more information about a process. Aleatory uncertainties arise because of the inherent randomness in a system and are irreducible. These include for instance material properties that were determined through sample tests. The parameters in Table 2 were ranked subjectively, according to their importance based on experience and observation.

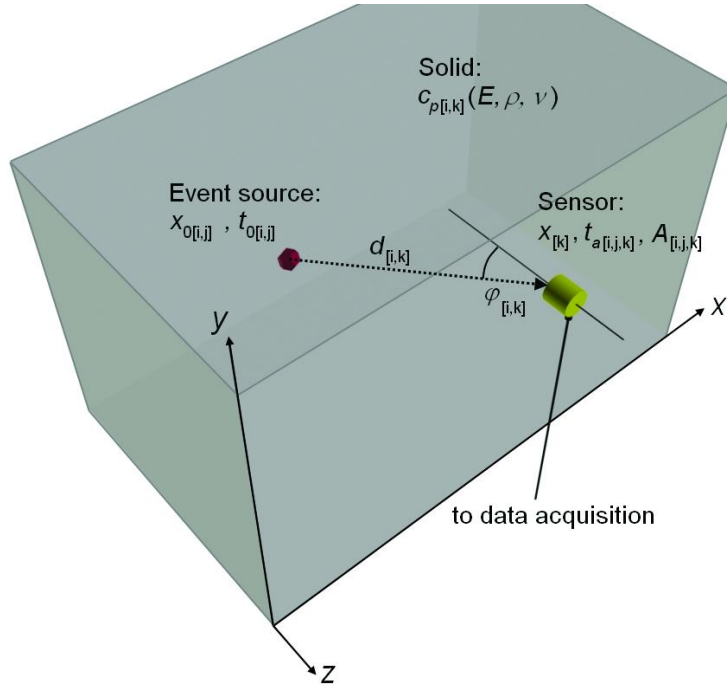


Fig. 3. Generic test specimen with parameters (only one sensor shown for clarity)

Table 2. Most important parameters present in source location estimations

Parameter	Type	Description
p -Wave arrivals, $t_{a[i,j,k]}$	<i>Ale</i>	Accuracy of determination dependent on: <ul style="list-style-type: none"> - Signal-to-noise ratio of recorded signal - Wave travel path and attenuation characteristics Arrival time picking errors Δt may be a function of: <ul style="list-style-type: none"> - Wave travel path length, $d_{[i,j,k]}$ - Wave incident angle, $\varphi_{[i,k]}$ - Maximum signal amplitude, $A_{[i,j,k]}$ (observed)
p -Wave velocity, $c_{p[i,k]}$	<i>epi</i>	Function of material properties, micro-cracks: <ul style="list-style-type: none"> - Modulus of Elasticity, E - Material density, ρ - Poisson's Ratio, ν - Anisotropic behavior, e.g. due to reinforcing steel - Dispersion due to the constituents
Wave travel path, $d_{[i,k]}$	<i>ale/epi</i>	Can deviate from straight line due to macro-cracks, voids, specimen geometry, etc.
Sensor locations, $x_{[k]}$	<i>epi</i>	Measurement errors, inexact placement
Source characteristics	<i>ale</i>	Type, spatial and temporal dimensions
Location scheme	<i>epi</i>	<ul style="list-style-type: none"> - Sensitivity of location with respect to sensors - Errors in the source location scheme
Sensor characteristics	<i>epi</i>	Sensitivity, resonant behavior

From a Bayesian point of view, there is no fundamental difference between these types of variable, i.e. they are all regarded and treated as random quantities. However, it is still important to know which model parameters are present and irreducible and which can be influenced by improving the mathematical model or including more observations.

3 Model Formulation and Implementation

The development of a probabilistic stress wave source location algorithm to predict source location parameters in full probabilistic form based on Bayesian analysis consists of the following steps:

1. Formulation of inference model
2. Estimation of model parameters using Bayesian inference
3. Implementation of predictive model
4. Model validation

In the succeeding four sections, these steps are described in detail. A simple probabilistic model is implemented and used to demonstrate the potential of the proposed framework.

3.1 Formulation of Inference Model

Based on the uncertainties described in section 2.4, a basic probabilistic model for the estimation of stress wave source locations was developed. The general relationship between source parameters and sensors is the arrival time function as stated in Eq. 3 and illustrated in Fig. 5. This function is now restated taking into account uncertainties in the model parameters and errors. The slowness s_p was used as the descriptor for the motion of the p -wave rather than velocity c_p and is defined as:

$$s_p = c_p^{-1} \quad (9)$$

Variability in the model is assumed to be associated with the basic source location parameters, which are the p -wave slowness s_p , the observed arrival times t_a , and the event time t_0 . Note that not all parameters vary with each index i, j, k . There is only one event time t_0 per AE event, i.e. for a group of eight sensors, and thus it does not vary with k . The arrival times $t_{a[i,j,k]}$ are normalized for each event, i.e. the arrival times for the first-hit-sensor are 0.0. The wave propagation characteristics were assumed to be the same for each PLB at a specific location and s_p and d are therefore not updated with j . The travel distance d is modeled as a deterministic parameter and calculated from the ‘known’ source and sensor locations using Eq. 4. The central model equation is based on Eq. 3 and can be stated as follows:

$$t_{a[i,j,k]} = t_{0[i,j]} + s_{p[i,k]} d_{[i,k]} + \varepsilon_{[k]} \quad (10)$$

where ε is an error term modeled as a Gaussian random variable with zero mean and standard deviation $\sigma_{\varepsilon[k]}$, i.e. it is assumed that the error is a function of the sensor. Alternatively, this probabilistic model can be written as:

$$T_{a[i,j,k]} \sim N\left(t_{0[i,j]} + s_{p[i,k]}d_{[i,k]}, \sigma_{\varepsilon[k]}\right) \quad (11)$$

The model parameters that were estimated and used to predict future events are the following:

- Mean of the p -wave slowness, μ_{s_p}
- Standard deviation of the p -wave slowness, σ_{s_p}
- Event time, $t_{0[i,j]}$
- Standard deviation of the observed arrival times, $\sigma_{\varepsilon[k]}$

The first two parameters are global parameters, i.e. they do not vary with any of the indices i , j , or k . In the MCMC algorithm, instead of the standard deviation σ , the precision τ is used and defined as the inverse of the variance σ^2 :

$$\tau = \sigma^{-2} \quad (12)$$

The software *WinBUGS* allows for a graphical representation of the probabilistic model which is illustrated in Fig. 4. Round nodes correspond to random variables and square nodes to constants. Observed nodes are shown in grey. The four model parameters listed above require specification of prior distributions as described in section 2.1; the selected distributions are listed in Table 3. Normal (N) distributions with low precisions are assigned as priors to random variables that represent means, and Uniform (U) distributions with extensive intervals as priors to random variables that represent standard deviations. The selected distributions are weakly informative priors.

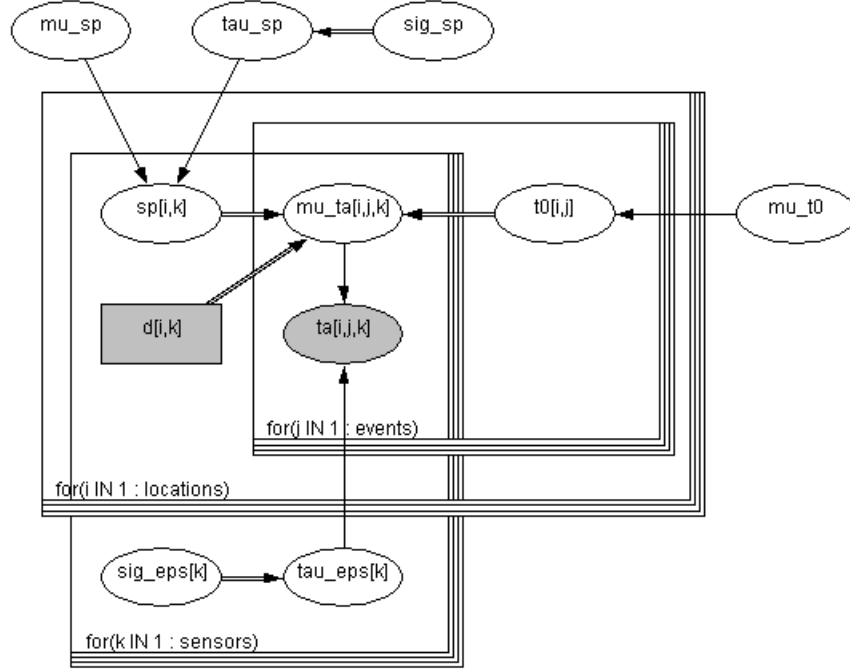


Fig. 4. Graphical representation of inference model (screenshot from *WinBUGS*)

Table 3. Description of nodes used in inference model

Parameter	<i>WinBUGS</i> identifier	Unit	Node type	Description/ Prior distributions
μ_{s_p}	mu_sp	[ms/m]	Stochastic	$N(0.25, 1.0)$
σ_{s_p}	sig_sp	[ms/m]	Stochastic	$U(0.0, 100)$
τ_{s_p}	tau_sp	$[(\text{ms}/\text{m})^{-2}]$	Logical	$\tau_{s_p} = \sigma_{s_p}^{-2}$
μ_{t_0}	mu_t0	[ms]	Stochastic	$N(-0.05, 1.0)$
$\sigma_{\varepsilon[k]}$	sig_eps[k]	[ms]	Stochastic	$U(0.0, 100)$
$\tau_{\varepsilon[k]}$	tau_eps[k]	$[\text{ms}^{-2}]$	Logical	$\tau_{\varepsilon[k]} = \sigma_{\varepsilon[k]}^{-2}$
$t_{0[i,j]}$	t0[i,j]	[ms]	Stochastic	$N(\mu_{t_0}, 1.0)$
$s_{p[i,k]}$	sp[i,k]	[ms/m]	Stochastic	$N(\mu_{s_p}, \tau_{s_p})$
$d_{[i,k]}$	d[i,k]	[m]	Constant	Calculated using Eq. 4, observed
$\mu_{t_a[i,j,k]}$	mu_ta[i,j,k]	[ms]	Logical	Eq. 10
$t_{a[i,j,k]}$	ta[i,j,k]	[ms]	Stochastic	$N(\mu_{t_a[i,j,k]}, \tau_{\varepsilon[k]}), \text{observed}$

3.2 Estimation of Model Parameters Using Bayesian Inference

The model described in section 3.1 is implemented in *WinBUGS* and the analysis executed to estimate the posterior distributions of all model parameters. The number of model simulation updates was selected based on convergence criteria presented in Congdon [23]: a total of 220,000 were generated of which the first 20,000 samples were discarded to account for the so-called ‘burn-in time’. Of the remaining 200,000 samples, every 10th sample was picked, producing a final set of 20,000 samples. This process is called thinning and ensures minimal autocorrelation between the samples. Fig. 5 shows kernel density estimates for the model parameters.

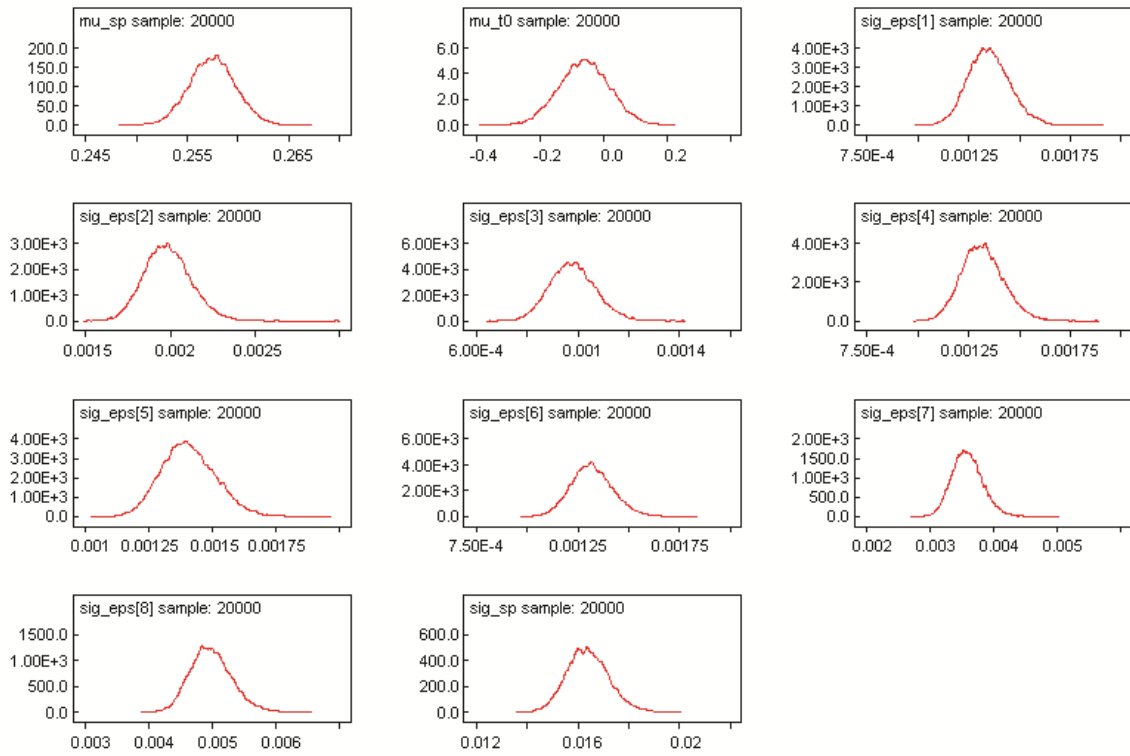


Fig. 5. Kernel density estimates for model parameters, units see Table 4

The estimated average p -wave velocity c_p (determined from node μ_{s_p} by Eq. 9) was 3.89 m/ms (153 in./ms) which is physically meaningful and comparable to values found in earlier studies by the first author [5]. The mean event time μ_0 was estimated to be approximately -0.06 ms which is reasonable as well, considering the size of the test specimen. Also, standard deviations for errors in time arrivals $\sigma_{\varepsilon[k]}$ range from approximately 0.001 to 0.005 ms. Finally, R - and P -matrices were computed containing the correlation coefficients ρ and p -values, respectively, to test for correlation between the model parameters.

No significant correlation was found, i.e. all model parameters were essentially statistically independent which indicates that the selected model is not over-parameterized.

Continuous probability distributions were subsequently fitted to each estimated parameter based on the highest log-likelihood value and are summarized in Table 4. Also, listed are simulation sample means \bar{x} , standard deviations s , coefficients of variation $C.o.V. = \frac{s}{\bar{x}}$, and the Monte Carlo standard error of the mean (MC error). As can be observed, the mean μ_{s_p} and the precision τ_{s_p} of the p -wave slowness were found to be well described by the Normal (N) and Gamma (Gam) distributions, respectively. The mean event time μ_{t_0} was found to be Normal (N) distributed and the precisions of the arrival time measurements $\tau_{e[k]}$ followed a Gamma (Gam) distribution. These fitted probability distributions can now be employed for the prediction of source location parameters once new observed time arrivals become available.

Table 4. Statistics of estimated (posterior) model parameters and fitted posterior distributions

Parameter	Unit	Posterior sample statistics				Fitted distributions
		\bar{x}	s	CV	MC error	
μ_{s_p}	[m/ms]	0.257	0.00227	0.009	5.61E-5	N(0.257, 194,000)
τ_{s_p}	[(m/ms) ⁻²]	3,744	372.0	0.099	2.79	Gam(102, 0.0272)
μ_{t_0}	[ms]	-0.0607	0.0796	1.31	6.24E-4	N(-0.0607, 158)
$\tau_{e[1]}$	[ms ⁻²]	557,000	84,700	0.152	571	Gam(43.6, 7.83E-5)
$\tau_{e[2]}$	[ms ⁻²]	257,000	35,400	0.138	284	Gam(52.7, 2.05E-4)
$\tau_{e[3]}$	[ms ⁻²]	1.07E+6	197,000	0.184	1400	Gam(30.1, 2.82E-5)
$\tau_{e[4]}$	[ms ⁻²]	585,000	90,600	0.155	641	Gam(42.2, 7.22E-5)
$\tau_{e[5]}$	[ms ⁻²]	513,000	76,900	0.150	554	Gam(45.1, 8.76E-5)
$\tau_{e[6]}$	[ms ⁻²]	585,000	89,900	0.154	600	Gam(42.8, 7.31E-5)
$\tau_{e[7]}$	[ms ⁻²]	79,300	10,400	0.131	70.9	Gam(57.7, 7.28E-4)
$\tau_{e[8]}$	[ms ⁻²]	40,900	5,270	0.129	39.2	Gam(60.3, 0.00148)

It should be understood that this probabilistic model is a simplified model because phenomena such as wave scattering, attenuation, and changing boundary conditions that will increase and complicate the travel path are modeled as one common uncertainty in the system, which limits the predictive power of the model. The authors intend to include additional model parameters and error terms in the future that can address each phenomenon individually. One way to approach the issue of changing boundary

conditions is by updating the model parameters through additional estimations from time to time, e.g. when deterioration occurs such as cracking, which can be detected by this algorithm.

3.3 Implementation of Predictive Model

Based on Eq. 7, a predictive model was implemented. It is based on the same probabilistic model as presented in section 3.1 and visualized in Fig. 6. For this model, subscript i corresponds to the predicted location and k to the sensor. The estimated model parameters presented in Table 4 were assigned as priors. Observed nodes, or known parameters are shown in grey.

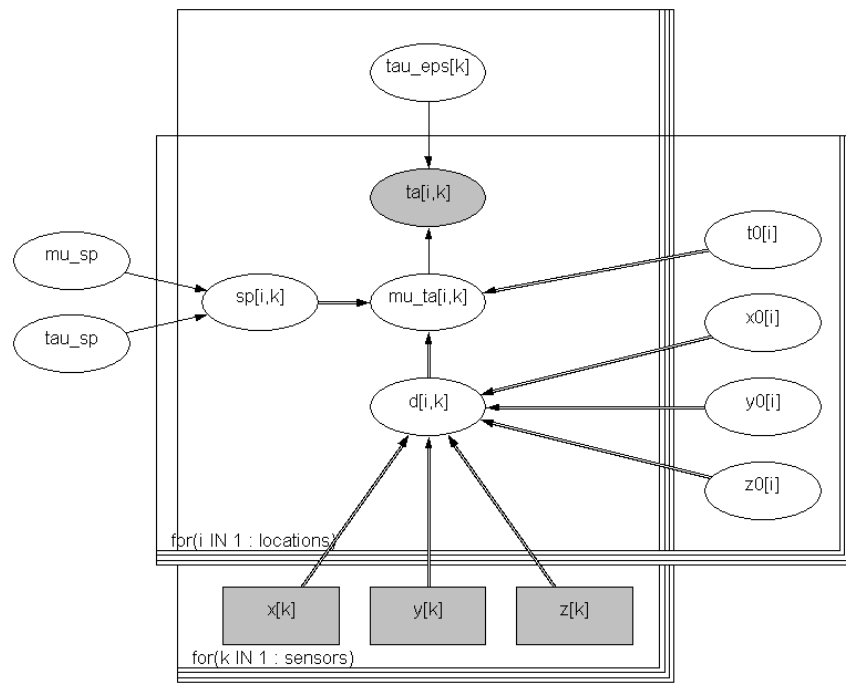


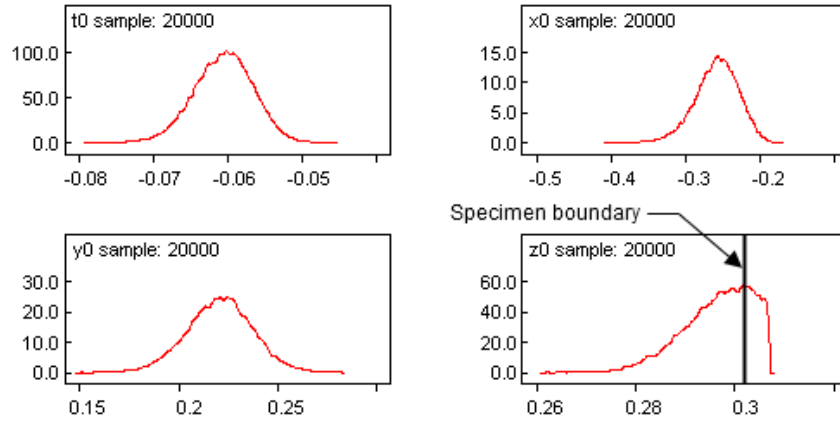
Fig. 6. Graphical representation of prediction model (screenshot from *WinBUGS*)

Once new p -wave time arrivals $t_{a[i,k]}$ are observed and available, the source location parameters t_0 , x_0 , y_0 , and z_0 can be estimated using this predictive model. The specified priors used in this study are listed in Table 5. The prior for the event time t_0 is readily available from the parameter estimation (see Table 4) and the priors for the spatial coordinates x_0 , y_0 , and z_0 were selected based on the geometrical boundaries of the specimen and the knowledge that the events lie on the front x - y face. A distance of 5 mm is added to each boundary coordinate to account for measurement uncertainties. This value is based on the authors' observation and experience in the laboratory on how accurate a specimen of this size can be constructed and its dimensions subsequently determined.

Table 5. Description of source location parameters and priors

Parameter	WinBUGS name	Unit	Node type	Specified priors	Lower, upper bounds
t_0	t0[i]	[ms], [ms ⁻²]	Stochastic	N(-0.06, 160)	-
x_0	x0[i]	[m]	Stochastic	U(-0.92, 0.92)	-
y_0	y0[i]	[m]	Stochastic	U(-0.005, 0.618)	-
z_0	z0[i]	[m], [m ⁻²]	Stochastic	N(0.302, 10,000)	(-0.316, 0.307)

In order to validate the predictive model, a series of 22 PLBs that had not previously been used in the model parameter estimation, were employed and are presented in detail in section 3.5. An example of the solution obtained for one selected event (number 5) is illustrated in Fig. 7: all four source location parameters are available with their PDFs.

**Fig. 7.** Example of kernel density estimates for predicted source location parameters, units see Table 5

For the updating process, the same principles as described in section 3.2 were employed. It should be noted that the solution does not need to follow any conventional parametric PDF but can be quite complex, or sometimes multi-modal. In section 3.4, means are presented to efficiently visualize such solutions and determine an optimal point solution.

3.4 Visualization of Source Location Parameters

The visualization of the simulated source location parameters is based on methods used in data mining for multi-dimensional density estimations. A summary of the concepts can be found in [24]. In order to enhance interpretation of the complex solution data, a 3-D scatter plot was generated with each data point in a color according to the density of the region in which that point is located. The density value for each

data point was estimated by evaluating a Gaussian-type kernel at the Euclidian distances to all other spatial points in the data set. The width of that kernel was defined in terms of standard deviations of the Gaussian- type kernel σ_k and set to the following value:

$$\sigma_k = \frac{\hat{s}}{3} \quad (13)$$

where \hat{s} is the mean sample standard deviation of the three spatial coordinates. This value was determined by trial and error. Smoothed and weighted intensity values for each data point were computed and then employed into a 3-D scatter plot to greatly enhance interpretation and understanding of the potentially complex solution space. A color map was assigned that shows low- and high-density regions in blue and red, respectively. These intensities can be converted into probability densities by applying the normalizing constant c as introduced in section 2.1. An example for the data set presented in Fig. 7 is illustrated in Fig. 8(a). As suggested by other researchers, the mode of the predicted source location parameters may be selected as optimal point solution representation [7, 14]. The 3-D (or spatial) mode can readily be determined by picking the sample point with the highest computed intensity. In addition to that, the temporal data can be included. In order to estimate the 4-D (or hyperspace) mode, the so-called mean-shift algorithm was implemented [25]. With this general and non-parametric method, modes and clusters of complex feature spaces can be explored and a common application is in image processing for detecting and selecting similar features (shapes, colors) in photos. In this study, the mean-shift method was used to estimate the optimal point for all solution parameters (x_0 , y_0 , z_0 , and t_0) simultaneously, i.e. the 4-D mode, as is shown by the white pentagram in Fig. 8. For reference, the 3-D spatial mode is shown as well as a white circle. As can be observed, the difference between the two optimal solutions is minor for this particular event.

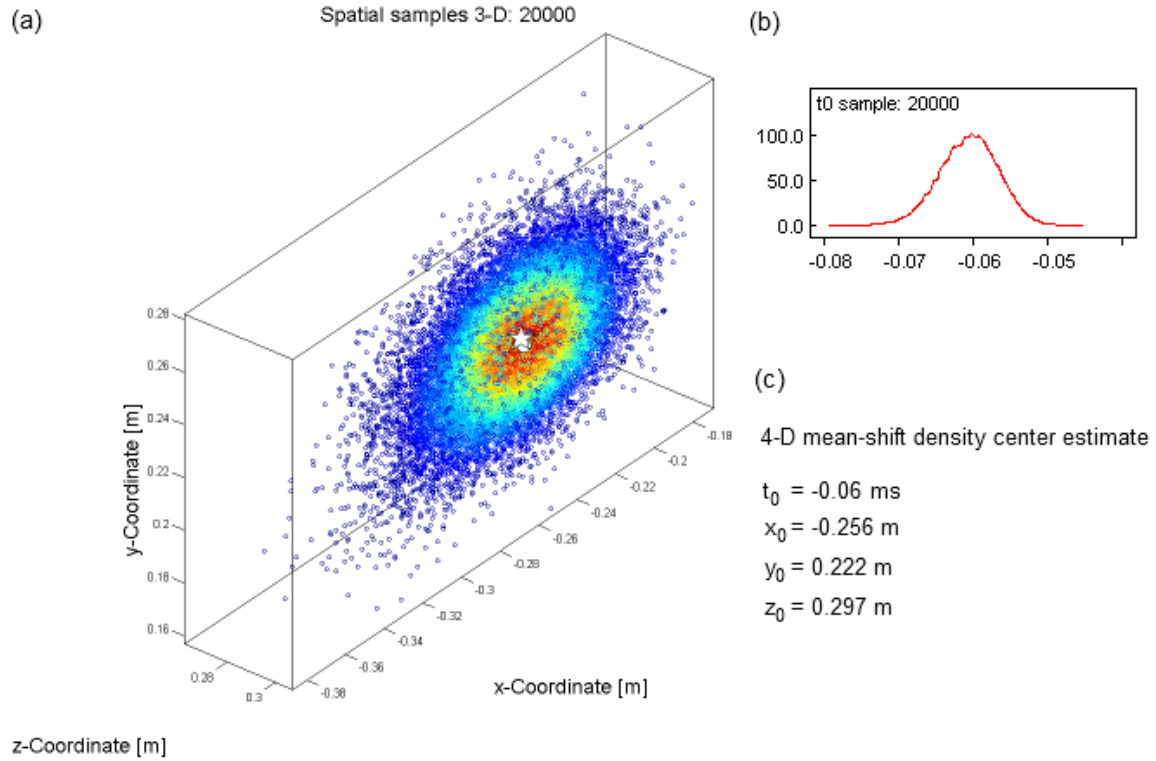


Fig. 8. Example of visualization of source location parameter results:

- (a) 3-D scatter plot of spatial samples with 3-D (white circle) 4-D mode (white pentagram)
- (b) Kernel density estimate for t_0
- (c) Tabulated results for 4-D mode estimated by mean-shift algorithm

3.5 Model Validation

In order to validate the proposed probabilistic prediction model, a set of 22 PLB events on the front x - y face of the specimen that had not previously been used in the model parameter estimation were evaluated. The obtained probabilistic source location predictions are illustrated in Fig. 9 wherein 2000 samples are shown for each solution. Each true PLB location is shown with a black '+'. The 3-D and 4-D modes are shown as white circles and pentagrams, respectively. It can be observed that the predicted locations are generally accurate, i.e. the estimated 3-D and 4-D modes are close to the true PLB location for most of the PLBs. Deviations occur for locations that are towards the boundaries of the sensor array. Also, it can be observed that the predictions are usually not Normal (N) distributed, which is revealed by the present source location algorithm.

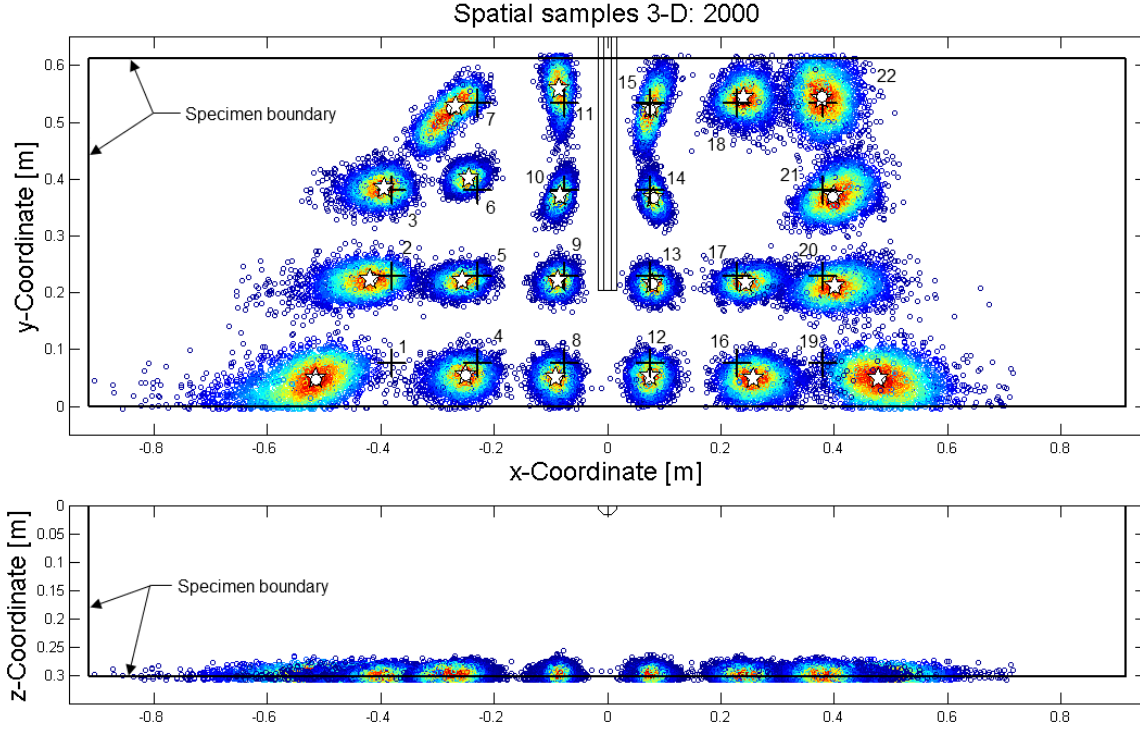


Fig. 9. Probabilistic source location predictions for 22 PLBs on x - y front face
 (+ = PLB locations, O = 3-D mode, ★ = 4-D mode)

Fig. 10 illustrates the prediction deviations in terms of the Euclidian distance between predicted (i.e. modes of the posterior PDF) and true PLB location. The results for a traditional prediction algorithm as presented in section 1.2 are also included for comparison. The deterministic parameters used for the traditional algorithm were 3.89 m/ms for the wave velocity c_p , the observed mean arrival times $t_{a[i,j,k]}$, and the measured sensor location coordinates (see Table 1). The total computation time for one event, including sampling, kernel evaluation, and visualization, is on the order of 50 s on a standard PC with an Intel I7 core processor. It can be observed that, generally, the mean solution of the traditional algorithm produces similar results to the mode of the probabilistic solution. In a majority of cases, the probabilistic predictions are slightly better but also provide insight into the source parameters. The larger errors for points 1 and 19 show that the present probabilistic algorithm does not provide improved predictions in the mean for all cases, in particular if the deterministic prediction uses the mean of the posterior distribution obtained with the presented method, as done here. However, the proposed probabilistic method has the benefit of providing an improved measure for the uncertainty of the prediction.

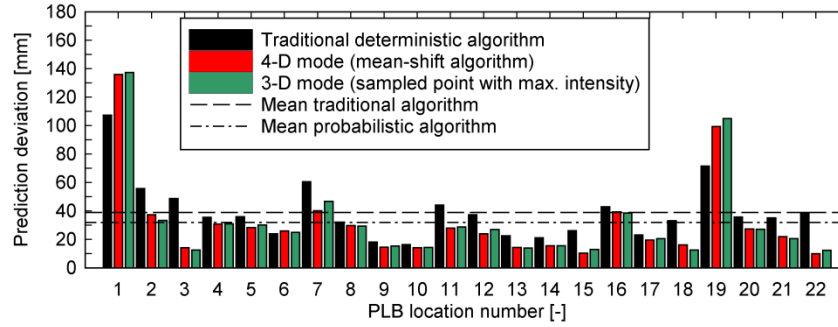


Fig. 10. Prediction deviations, absolute distance between predicted and true locations

The mis-predictions are assumed to be caused by local effects that depend on the individual signal travel paths and therefore are difficult to account for. Such local effects can be associated to voids in the concrete matrix or nearby steel reinforcing bars that interfere with the propagating stress wave. One way to address this issue could be to include only part of the observed arrival times in the prediction, i.e. arrival times that appear to be outliers are omitted. This could be done by trial and error. Another way could be to incorporate a correction term δ in the probabilistic model that can account for path dependency. This is planned for future work and may be used to employ additional information such as the signal strength or the angle at which the stress wave arrives at the sensor to further enhance the probabilistic model.

4 Summary and Conclusions

A novel framework for the development of a probabilistic algorithm to estimate stress wave (or AE) source locations is presented. Bayesian analysis methods in conjunction with Markov Chain Monte Carlo simulation were employed to first estimate model parameters using observed data from an experiment, and then to predict source location parameters of future observations in fully probabilistic form. The implemented probabilistic model accounts for variability in the wave propagation characteristics, the measured arrival times, and the event time. Bayesian inference was found to be a powerful and flexible method that can estimate parameters based on a large amount of observed data. All distribution and other parameters were estimated with their full posterior probability distributions. It was shown that predictions based on estimated parameters employing newly observed arrival times can be performed and produce acceptable results, even with a relatively basic model. Temporal and spatial source location parameters are available in full probabilistic form and 3-D or 4-D modes determined by kernel density estimation methods can be used as optimal point location results. Extension of the presented probabilistic model is possible by including a correction term to account for other uncertainties. The authors are currently working on improving the presented model formulation and include additional information in form of the

maximum signal amplitude (as shown in Fig. 1). First preliminary results are promising as they indeed produce slightly more accurate predictions.

5 Acknowledgements

The authors would like to thank Oregon Department of Transportation for the sponsorship of this study. Christian Baumberger's insight in the field of multi-variate density estimation (section 3.4) is greatly appreciated.

6 References

1. EN, *1330-9: Non-destructive testing-Terminology-Part 9: Terms used in acoustic emission testing*, 2000, European Committee for Standardization (CEN): Brüssel, Belgium.
2. Grosse, C.U. and M. Ohtsu, *Acoustic Emission Testing - Basics for Research-Applications in Civil Engineering* 2008, Berlin & Heidelberg, Germany: Springer Verlag.
3. Miller, R.K., E.v.K. Hill, and P.O. Moore, *Nondestructive Testing Handbook - Volume 6 - Acoustic Emission Testing* 2005, Columbus, OH: American Society for Nondestructive Testing (ASNT).
4. Graff, K.F., *Wave Motion in Elastic Solids* 1991, New York, NY: Dover Publications, Inc.
5. Schumacher, T., *Acoustic Emission Techniques Applied to Conventionally Reinforced Concrete Bridge Girders*, in *Final Report SPR 6332* 2008, Oregon Department of Transportation (ODOT): Salem, OR.
6. Kurz, J.H., C.U. Grosse, and H.-W. Reinhardt, *Strategies for reliable automatic onset time picking of acoustic emissions and of ultrasonic signals in concrete*. *Ultrasonics*, 2005. **43**(7): p. 538-546.
7. Schechinger, B., *Schallemissionsanalyse zur Ueberwachung der Schädigung von Stahlbeton*, in *Institut für Baustatik und Konstruktion* 2006, Eidgenössische Technische Hochschule Zürich: Zurich, Switzerland.
8. Geiger, L., *Herdbestimmung bei Erdbeben aus den Ankunftszeiten*. *Nachrichten von der Königlichen Gesellschaft der Wissenschaften zu Göttingen*, 1910. **4**: p. 331-349.
9. Ge, M., *Analysis of Source Location Algorithms, Part I: Overview and non-iterative methods*. *Journal of Acoustic Emission*, 2003. **21**: p. 14-28.
10. Ge, M., *Analysis of Source Location Algorithms, Part II: Iterative methods*. *Journal of Acoustic Emission*, 2003. **21**: p. 29-51.
11. Ohtsu, M., et al., *Determination of crack location, type and orientation in concrete structures by acoustic emission*. *Magazine of Concrete Research*, 1991. **43**(155): p. 127-134.
12. Grosse, C.U., H.-W. Reinhardt, and T. Dahm, *Localization and classification of fracture types in concrete with quantitative acoustic emission measurement techniques*. *NDT & E International*, 1997. **30**(4): p. 223-230.
13. Tarantola, A. and B. Valette, *Inverse Problems = Quest for Information*. *Journal of Geophysics*, 1982. **50**: p. 159-170.
14. Lomax, A., et al., *Probabilistic Earthquake Location in 3D Layered Models*, in *Advances in Seismic Event Location*, C.H. Thurber and N. Rabinowitz, Editors. 2000, Kluwer Academic Publishers: Dordrecht, The Netherlands. p. 101-134.
15. Weber, Z., *Probabilistic local waveform inversion for moment tensor and hypocentral location*. *Geophysical Journal International*, 2006. **165**(2): p. 607-621.

16. Lomax, A. *The NonLinLoc Software Guide*. 2008 September 2008 [cited 2009 September 4]; 5.00:[<http://alomax.free.fr/nlloc/>].
17. Gilks, W.R., S. Richardson, and D.J. Spiegelhalter, *Markov Chain Monte Carlo in Practice - Interdisciplinary Statistics*1996, Boca Raton, FL: Chapman & Hall/CRC.
18. Hastings, W.K., *Monte Carlo Sampling Methods Using Markov Chains and Their Applications*. Biometrika, 1970. **57**(1): p. 97-109.
19. Imperial College and Medical Research Council. *The BUGS Project - WinBUGS*. 2007 [cited 2009; <http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/contents.shtml>].
20. Lovejoy, S.C., *Development of Acoustic Emissions Testing Procedures Applicable to Conventionally Reinforced Concrete Deck Girder Bridges Subjected to Diagonal Tension Cracking*, in *Mechanical Engineering*2006, Oregon State University: Corvallis, OR.
21. Koester, C.C., *Testing and Evaluation of Flexural Reinforcing Bar Anchorages Terminating in Columns*, in *School of Civil and Construction Engineering*2007, Oregon State University: Corvallis, OR.
22. Der Kiureghian, A., *Analysis of Structural Reliability under Model and Statistical Uncertainties: a Bayesian Approach*. Computational Structural Engineering, 2001. **1**(2): p. 81-87.
23. Congdon, P., *Applied Bayesian Modelling*2003, West Sussex, England: John Wiley & Sons Ltd.
24. Scott, D.W. and S.R. Sain, *Multi-dimensional Density Estimation*, in *Handbook of Statistics, Vol. 24*, C.R. Rao, E.J. Wegman, and J.L. Solka, Editors. 2004, Elsevier B. V.: Amsterdam.
25. Comaniciu, D. and P. Meer, *Mean Shift: A Robust Approach Toward Feature Space Analysis*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002. **24**(5): p. 603-619.