Rate Distortion Analysis of Wavefield Coding in Wireless Geophone Networks

Hamood ur Rehman Khan[1] and Farhan Khan[2(✉)]

[1] Sir Syed Center for Advanced Studies in Engineering, Islamabad, Pakistan
hamood.rehman@carepvtltd.com
[2] Electrical and Computer Engineering Program, Habib University, Karachi, Pakistan
farhan.khan@sse.habib.edu.pk

Abstract. Current and future trends in seismic acquisition point towards higher geophone densities (forecasted to be 1M nodes per survey). The geophones' high operating precision and a sampling rate of a few milliseconds leads to a huge aggregate data rate in the geophone array. To handle this large data rate a hierarchy of multiplexed lines (including fiber optic cables) are used, resulting in substantial deployment costs. This work considers wireless geophone networks to mitigate these costs. Because of limited bandwidth of the wireless medium, compression of acquired data is needed. We assess lossy source coding (signal compression) performance using a rate-distortion tradeoff based on a physical model of the earth. The distortion criterion used is mean squared error. The earth's physical model considered here consists of a randomly layered subsurface structure where the acoustic impedance of the earth varies as a homogeneous spatial random process. The rate distortion performance is assessed in terms of the parameters of this physical model e.g., the speed of sound underground and correlation length of acoustic impedance process. In comparison with previous work, this paper derives a closed form expression for the autocorrelation function for the reflection process for the 3D subsurface volume. We also show from the resulting rate-distortion curves that the compression performance improves both as the coding bock length increases and as the correlation length of the medium properties increases. The rate-distortion surface as a function of normalized MSE and sound of speed underground is also computed and validated through extensive simulations.

Keywords: Wireless Sensor Networks · Seismic Data Compression · Rate Distortion · Wave Equation · Green's Function · Random Differential Equation · Randomly Layered Medium

1 Introduction

In exploration seismology acquisition is the first stage of the signal processing pipeline. It consists of laying out a large number of acoustic sensors (geophones)

X. Li et al. (Eds.): SmartGift 2024, LNICST 600, pp. 54–77, 2025.
https://doi.org/10.1007/978-3-031-78806-2_4

in a network and injecting the ground with energy from an energy source (e.g. Vibroseis). The injected energy is reflected from subsurface layers of differing acoustic impedance as a seismic wave and is sensed at the geophones and converted into an electrical voltage that is digitized and transferred to a central fusion center. In this way, subsurface structures are mapped to investigate the occurrence of natural resources. Wavefield coding entails the use of data compression on the digitized amplitudes of the seismic wavefield and decompression of the data at the recorder unit.

Currently there is a shift towards large scale data acquisition. A 10 to 100-fold increase in sensor node count for a given area is underway [5]. Future node densities are forecasted to be even higher, with some surveys aiming to achieve a density of 1 million nodes [6]. The projected shift is from accurate acquisition to better statistical sampling of seismic data. A second trend is towards completely wireless data acquisition—which is the scenario considered in this paper. This presents a huge increase in the handled data rate. To get an idea of the problem, consider the following: A typical survey might cover an area of 2500 sq km. The line separation between geophone rows is typically on the order of 50 to 100 meters, while the receiver interval is around 10–25 m. This gives a total number of about $50-100$K geophones per survey. Given that the geophones operate at 2 ms sampling interval with a bit depth of usually 24 bits, with shot measurements lasting for 6 s, the total aggregate rate of the network is on the order of 1 Gbps. This is the data traffic of medium-sized city.

Since the wireless channel has limited bandwidth and the nodes have limited power owing to working on batteries or on solar power, there needs to be a careful tradeoff analysis between bandwidth (rate) and reproduction accuracy (alternatively known as distortion incurred in compression). Compression involves discretizing the field's continuous valued amplitude at each spatial location. Information is lost during this process. The problem is known as lossy source coding for sensor networks in information theory parlance.

In geoscience and remote sensing, rate-distortion analysis in the formualtion of compression and coding of signal fields has been done in a number of works (see [1–4] for recent contributions). Previous work done in relation to our problem includes the study of [13,14]. [13] studies the problem when the excitation signal (sound) is a two-dimensional random field emanating from a line source embedded in a stationary fluid in thermodynamic equilibrium supporting acoustic propagation of waves that is free from reflecting and/or scattering interfaces. The work done in [14] corresponds to the case of temperature fields instead of acoustic fields and is done for the geometry of rods and rings instead of a slab of layers (as in our subsurface model). In our work the propagation medium is randomly layered. The area of wave propagation in random media has been very actively researched during the last four decades [23,26] and is a thorough field as of now. The main motivation for studying wave phenomenon in random media is ostensibly to generate strategies for seismic inversion for complex media [23] whereas here we are concerned with obtaining a statistical model for seismic image sources so that they can be compressed.

In this work we carry out a performance analysis of possible compression algorithms employed in a wireless geophone networks. This paper analyzes the opposite scenario than that considered in [13]. In our case the exciting signal is a known pulse located at the surface (as opposed to being buried in the subsurface as in [13]) and the recorded wavefield is the result of interaction of the excitation signal with a layered medium, which is the case in seismic exploration experiments. As pointed out earlier, sensor nodes have limited resources including computational power, limited energy, memory and bandwidth. Source coding involves finding the optimal tradeoff between the use of these scarce resources and the quality of the reconstructed data at the central recording unit. The main tool we use to characterize this tradeoff is R(D) theory for random fields. We assume the existence of a distortion measure or measure of dissimilarity between the original source wavefield and the reconstructed wavefield at the decoder. Specifically, rate-distortion provides us with the answer to the following question: What is the minimum number of bits per sample that are needed to describe a source (the seismic image) at distortion level D, determined by the distortion measure. In our treatment, the distortion measure D is the squared error per digitized letter of the source wavefield. The theory assumes that a statistical description of the source to be coded is available for analysis.

Unfortunately for the seismic acquisition problem the exact statistics of the source are not known. To date there are no detailed statistical models for seismic images. To overcome this difficulty, we make use of the physics of the seismic acquisition process. That is, the seismic source is modelled through the process of injecting a known signal into a randomly layered medium in which the acoustic wave equation holds to a first order of approximation. The output signal measured at the surface through the sensors is then a random field. The Green's function for the random wave equation is obtained and its power spectrum is computed. The power spectrum for the output random field at the surface is then expressed in terms of the Green's function's spectrum. The output field is sampled in space and time along a line of geophones using periodic spatio-temporal sampling scheme, i.e., a sampling lattice. The sampled field is then coded (quantized) and R(D) bounds are obtained through the use of well-known parametrized formulas of the rate-distortion curves for a process with a known correlation structure [7].

To analyze the compression performance, we consider three source coding schemes. The first scheme is centralized coding in which all the data from all sensors is available at a single spatial location, free of communication costs. The second scheme is independent coding in which all sensors independently code their respective data streams and are independently decoded at the fusion center without paying regard to the correlation of information among different sensors. The third scheme is what is known as distributed coding scheme [9] in which all geophones independently (without internode communication) encode their data streams but are jointly decoded at the fusion center, exploiting the correlation within the data of the entire field. The distributed or multi-terminal coding scheme has historically been of much interest in sensor network engineering as

its performance is guaranteed to be as good as when the nodes can communicate with each other, without actual internode communication, thus alleviating the need for internode communication and freeing up precious resources. Recent advances [10] have given realizable (non-asymptotic) source codes that achieve the performance of distributed source coding as promised by theoretical consideration for asymptotically long block length regime. Together, these three coding schemes correspond to the best case, the worst case and a middle ground scenario respectively, in terms of rate-distortion performance. Thus, they in a sense, bound the compression performance achievable for coding of wavefields through a wireless network of geophones. One more point to be considered in our problem formulation is that we treat the channel coding and source coding aspects for the wireless channel as separate. Although it is well known that channel and source separation does not hold in general sensor networks [11], here we treat the two problems as separate, assuming that the wireless channel introduces no errors into the data stream or that at least that the errors introduced by the wireless channel are handled by separate channel coding block independent of the source code without any loss of optimality in performance. This assumption can be justified in light of wireless sensor network models considered in [12] (which are approximated by high geohpone-count networks), where the source-channel separation theorem holds as the number of sensor nodes goes to infinity.

The paper is organized as follows. In Sect. 2 we describe the randomly layered earth model and the governing wave equation that describes propagation of acoustic pressure in the subsurface. The Green function for random wave equation is constructed and the spectral measure for this Green's function is obtained. In Sect. 3 we consider the spatio-temporal sampling of the Green's function. In Sect. 4 the finite order R(D) function is considered. The three cases of centralized, independent and distributed coding are handled. In Sect. 5 the previous analysis is simplified to the case of a one dimensional medium and the R(D) analysis for it carried is out using numerical techniques for non-stationary wavefields. In Sect. 6 numerical analysis and simulations are presented as results. In Sect. 7 conclusions are drawn.

2 The Random Medium Model and Wave Equation

In this section we develop the basic model that is used to analyze wave propagation in a randomly layered earth. A wave equation for the random medium is described and its Green's function is obtained. We express the action of this Green's function on the propagating wave as a random integral operator. The autocorrelation for this operator is obtained at the end. This autocorrelation function is required for the calculation of the R(D) function.

We are interested in assessing the compression performance tradeoff for seismic images generated by reflection seismology. This assessment requires a statistical model for general seismic images. To obtain such a statistical description we model the earth as a randomly layered semi-infinite medium, with the reflectivity varying randomly in the vertical direction. This formulation then effectively

allows us to perform information theoretic analysis on the resulting setup. In seismic inversion modelling, the reflectivity function, $\rho(x)$ has a step variation which we approximate by a smoothly varying curve that jumps along the discontinuities in $\rho(x)$ and is flat where the reflectivity is constant (between the random layers). The acoustic approximation for the medium is assumed where the subsurface obeys a scalar wave equation and elasticity of the medium is ignored (i.e., there is no shear wave propagation and the dominant waves are compressional or P waves).

The setting shown in Fig. 1 is considered. A receiver line at the surface measures the sensed field $V(x,t)$. The inverted triangles or wedges are the geophones. The Vibroseis source is to the left of the sensor line and is the original cause of excitation in the subsurface structure. Sound waves travel from the subsurface reflecting layers and travel to the receiver line while satisfying the following acoustic wave equation.

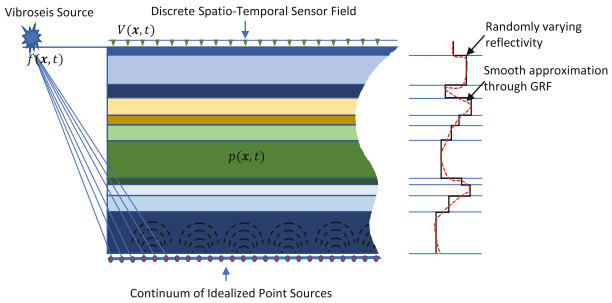

Fig. 1. The physical model of randomly layered earth being considered for the seismic acquistion process

$$\Delta p(t,\mathbf{x}) - \frac{1}{c^2(\mathbf{x})}\frac{\partial^2}{\partial t^2}p(t,\mathbf{x}) = f(t,\mathbf{x}) \tag{1}$$

where the subsurface structure is characterized by position dependent slowness $\frac{1}{c(\mathbf{x})}$, $p(t,\mathbf{x})$ is the acoustic pressure term and $f(t,\mathbf{x})$ is the source field at spatial location $\mathbf{x} = (x_1, x_2, x_3) \in \mathbb{R}^3$ and time $t \in [0, +\infty)$. Δ is the Laplacian operator, defined by $\Delta := \partial^2/\partial^2 x_1 + \partial^2/\partial^2 x_2 + \partial^2/\partial^2 x_3$. In our model the slowness is given by

$$\frac{1}{c^2(\boldsymbol{x})} \rightsquigarrow \frac{1}{c_0^2}\left(n_0^2(\boldsymbol{x}) + \rho(\boldsymbol{x})\right) \tag{2}$$

where c_0 is the reference wave speed in the medium, $n_0^2(\boldsymbol{x})$ is a smooth background index of refraction taken to be unity throughout the medium and $\rho(\boldsymbol{x})$ is the unknown reflectivity. We assume that $\rho(\boldsymbol{x})$ is a homogenous Gaussian random field. This is a key assumption that leads to the solution of the R(D)

problem in later analysis. The condition of homogeneity or shift-invariance is needed because calculation of the R(D) tradeoff for general random fields with arbitrary correlation structure or memory is not possible with current methods. Homogeneous Gaussian random fields enable the application of asymptotic distribution of eigenvalues of Toeplitz forms [15], which is a powerful tool for R(D) computation.

With the slowness thus modelled, the wave equation now reads as:

$$\Delta p(t,\boldsymbol{x}) - \frac{1}{c_0^2}\left(1+\rho(\boldsymbol{x})\right)\frac{\partial^2}{\partial t^2}p(t,\boldsymbol{x}) = f(t,\boldsymbol{x}) \tag{3}$$

together with the following initial conditions:

$$p(0,\boldsymbol{x}) = 0 \tag{4a}$$

$$\left.\frac{\partial p(t,\boldsymbol{x})}{\partial t}\right|_{t=0} = 0 \tag{4b}$$

The initial conditions ensure that the solution obtained to the above Cauchy problem is unique. To retrieve the solution to the hyperbolic PDE that relies only on the effect of the external seismic source initial conditions are assumed to be identically zero.

When the source field is modelled as an impulse at spatial location $\mathbf{y}$ and at time zero, the Green's function for our partial differential equation satisfies,

$$\Delta G(t,\boldsymbol{x},\boldsymbol{y}) - \frac{1}{c_0^2}\left(1+\rho(\boldsymbol{x})\right)\frac{\partial^2}{\partial t^2}G(t,\boldsymbol{x},\boldsymbol{y}) = \delta(\boldsymbol{x}-\boldsymbol{y})\delta(t) \tag{5}$$

where δ is the Dirac distribution. Equation (5) is a random differential equation because of the random term $\rho(x)$ as a co-efficient. If the medium is homogeneous ($\rho(\boldsymbol{x}) = 0$) the free space time-dependent Green's function is given by

$$G_0(t,\boldsymbol{x},\boldsymbol{y}) = \frac{1}{4\pi\|\boldsymbol{x}-\boldsymbol{y}\|}\delta\left(t - \frac{\|\boldsymbol{x}-\boldsymbol{y}\|}{c_0}\right) \tag{6}$$

The above is a spherical wave propagating at speed c_0 outward from the point $\boldsymbol{y}$. Define the temporal Fourier transform of the Green's function as

$$\hat{G}(\omega,\boldsymbol{x},\boldsymbol{y}) = \int G(t,\boldsymbol{x},\boldsymbol{y})e^{-i\omega t}dt \tag{7}$$

The free space Green's function in the frequency domain, corresponding to (6) is given by

$$\hat{G}_0(\omega,\boldsymbol{x},\boldsymbol{y}) = \frac{1}{4\pi\|\boldsymbol{x}-\boldsymbol{y}\|}e^{-ik\|\boldsymbol{x}-\boldsymbol{y}\|} \tag{8}$$

where $k = \omega/c_0$ is the spatial frequency or the wavenumber. Taking the Fourier transform of (5) we get:

$$\Delta\hat{G} + \frac{\omega^2}{c_0^2}\left(1 + \rho(\boldsymbol{x})\right)\hat{G} = \delta(\boldsymbol{x} - \boldsymbol{y}) \tag{9}$$

The free space Green's function satisfies the equation:

$$\Delta\hat{G}_0 + \frac{\omega^2}{c_0^2}\hat{G}_0 = \delta(\boldsymbol{x} - \boldsymbol{y}) \tag{10}$$

It can be shown that under the Born approximation (holds when the reflectivity $\rho(x)$ is either small or has small support), we can write the overall Green's function as:

$$\hat{G}(\omega, \boldsymbol{x}, \boldsymbol{y}) = \hat{G}_0(\omega, \boldsymbol{x}, \boldsymbol{y}) + \frac{\omega^2}{c_0^2}\int \rho(\boldsymbol{z})\hat{G}_0(\omega, \boldsymbol{x}, \boldsymbol{z})\hat{G}_0(\omega, \boldsymbol{z}, \boldsymbol{y})d\boldsymbol{z} \tag{11}$$

The first term in (10) is the direct wave emanating from the source and that reaches the geophone array without interacting with subsurface layers, while the second term accounts for the single scattered waves from source located at $\boldsymbol{y}$ to a subsurface point $\boldsymbol{z}$ and then back to the surface point $\boldsymbol{x}$ inside the geophone array. In usual seismic processing the term $\hat{G}_0(\omega, \boldsymbol{x}, \boldsymbol{y})$ is windowed and removed from the seismogram. We will from now on only consider the second term in our data model. The time domain Green's function is thus given as (with the direct wave removed):

$$G(t, \boldsymbol{x}, \boldsymbol{y}) = -\frac{1}{c_0^2}\frac{\partial^2}{\partial^2 t}\iint \rho(\boldsymbol{z})G_0(t - s, \boldsymbol{x}, \boldsymbol{z})G_0(s, \boldsymbol{z}, \boldsymbol{y})dsd\boldsymbol{z} \tag{12}$$

Assuming a source located at the origin and considering the shift-invariance of the free space Green's function, the expression for the overall Green's function becomes

$$\begin{aligned} G(t, \boldsymbol{x}, \boldsymbol{0}) &= G(t, \boldsymbol{x} - \boldsymbol{0}) \\ &= -\frac{1}{c_0^2}\frac{\partial^2}{\partial^2 t}\iint \rho(\boldsymbol{z})G_0(t - s, \boldsymbol{z})G_0(s, \boldsymbol{z} - \boldsymbol{0})ds\boldsymbol{z} \\ &= -\frac{1}{c_0^2}\frac{\partial^2}{\partial^2 t}\iint \rho(\boldsymbol{z})G_0(t - s, \boldsymbol{x} - \boldsymbol{z})G_0(s, \boldsymbol{z})dsd\boldsymbol{z} \\ &= G(t, \boldsymbol{x}) \end{aligned} \tag{13}$$

For a general source $f(t, \boldsymbol{x})$ the resultant wavefield is given in terms of this Green's function by the linear superposition and time invariance.

$$\begin{aligned} p(t, \boldsymbol{x}) &= \iint G(t - s, \boldsymbol{x}, \boldsymbol{y})f(s, \boldsymbol{y})dsd\boldsymbol{y} \\ &= \int G(t, \boldsymbol{x}, \boldsymbol{y}) *_t f(t, \boldsymbol{y})dsd\boldsymbol{y} \end{aligned} \tag{14}$$

where $*_t$ is the time convolution operator. Define the operator H by as the action of this convolution operator

$$Hf(t,\boldsymbol{x}) = \int G(t,\boldsymbol{x},\boldsymbol{y}) *_t f(t,\boldsymbol{y}) ds d\boldsymbol{y} \tag{15}$$

Of course for a Dirac distribution source at the origin, i.e., $f(t,\boldsymbol{x}) = \delta(\boldsymbol{x})\delta(t)$, the action of H is given by

$$H \cdot \delta = G(t,\boldsymbol{x},\boldsymbol{0}) = G(t,\boldsymbol{x}) \tag{16}$$

Note that H is a random operator because of the sctochastic Green's function that corresponds to the random differential equation (5). In fact the valuation of H is Gaussian random field that it inherits from the Gaussianity of the reflectivity function ρ which is involved in the convolution expression (13) that gives rise to $G(t,\boldsymbol{x})$. We are interested in the ensemble averaged autocorrelation function of H. The autocorrelation function of H corresponds to the autocorrelation of the Green function field $G(t,\boldsymbol{x})$. This autocorrelation (which is equal to $G(t,\boldsymbol{x})$'s autocovariance) is needed in the calculation of the rate-distortion tradeoff. The source term used in the analysis is a Dirac distribution. Using (14) we get:

$$\begin{aligned} R_{Hf}(\tau,\boldsymbol{u}) &= \mathbb{E}[Hf(t,\boldsymbol{r})H^* f(t+\tau,\boldsymbol{r}+\boldsymbol{u})] \\ &= \mathbb{E}[H\delta(t,\boldsymbol{r})H^*\delta(t+\tau,\boldsymbol{r}+\boldsymbol{u})] \\ &= \mathbb{E}[G(t,\boldsymbol{r})G^*(t+\tau,\boldsymbol{r}+\boldsymbol{u})] \end{aligned} \tag{17}$$

Thus obtaining

$$\begin{aligned} R_{Hf}(\tau,\boldsymbol{u}) = \mathbb{E}[&-\frac{1}{c_0^2}\frac{\partial^2}{\partial^2 t}\iint \rho(\boldsymbol{z}_1)G_0(t-s_1,\boldsymbol{r}-\boldsymbol{z}_1)G_0(s_1,\boldsymbol{z}_1)ds_1 d\boldsymbol{z}_1 \\ &\times -\frac{1}{c_0^2}\frac{\partial^2}{\partial^2 t}\iint \rho(\boldsymbol{z}_2)G_0^*(t+\tau-s_2,\boldsymbol{r}+\boldsymbol{u}-\boldsymbol{z}_2)G_0^*(s_2,\boldsymbol{z}_2)ds_2 d\boldsymbol{z}_2] \\ = &\frac{1}{c_0^4}\iiiint \frac{\partial^2}{\partial^2 t}G_0(t-s_1,\boldsymbol{r}-\boldsymbol{z}_1)G_0(s_1,\boldsymbol{z}_1) \\ &\times \frac{\partial^2}{\partial^2 t}G_0^*(t+\tau-s_2,\boldsymbol{r}+\boldsymbol{u}-\boldsymbol{z}_2)G_0^*(s_2,\boldsymbol{z}_2)\mathbb{E}[\rho(\boldsymbol{z}_1)\rho(\boldsymbol{z}_2)]ds_1 ds_2 d\boldsymbol{z}_1 d\boldsymbol{z}_2 \end{aligned} \tag{18}$$

Clearly $C(\boldsymbol{z}_1,\boldsymbol{z}_2) = \mathbb{E}[\rho(\boldsymbol{z}_2)\rho(\boldsymbol{z}_1)]$ in the last expression is an autocorrelation function. Since we assume homogeneity for the random field $\rho(\boldsymbol{z})$, $C(\boldsymbol{z}_1,\boldsymbol{z}_2)$ depends only on the difference of $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$. We consider Gaussian correlation for ρ:

$$C(\boldsymbol{z}_1,\boldsymbol{z}_2) = C(\|\boldsymbol{z}_1-\boldsymbol{z}_2\|) = \sigma^2 e^{-\frac{\|z_1-z_2\|^2}{a^2}} \tag{19}$$

where σ^2 is the variance of the fluctuation in the reflectivity and a is the correlation length. We define the spectral measure of H via the Weiner-Kinchin theorem:

$$S_H(\Omega,\boldsymbol{\Phi}) = \frac{1}{(2\pi)^4}\int_{\mathbb{R}^4} e^{-j(\Omega\tau+\boldsymbol{\Phi}\cdot\boldsymbol{u})}R_{Hf}(\tau,\boldsymbol{u})d\tau d\boldsymbol{u} \tag{20}$$

where Ω is the temporal frequency,$\boldsymbol{\Phi} = (\Phi_x, \Phi_y, \Phi_z)$ is the three-dimensional spatial frequency and $\boldsymbol{\Phi} \cdot \boldsymbol{u} = \Phi_x u_x + \Phi_y u_y + \Phi_z u_z$ is the dot product between the wave vector $\boldsymbol{\Phi}$ and the space vector $\boldsymbol{u}$. We now turn to sampling of this random field in time and space.

3 Sampling the Green's Function

In this section we consider the sampling of the Green's function obtained in Sect.2. As described before, the geophone array consists of lines of acoustic sensors that are more or less regularly spaced and that acquire the seismic wavefield reflected/scattered off the subsurface structure. The main objective is to temporally and spatially sample and then encode (quantize) the random wavefield generated by the impulsive source. This corresponds to sampling of the stochastic Green's function given by (13). Since in the setting discussed in this paper the geophone array is a one-dimensional linear array this corresponds to a spatio-temporal sampling of the following 2D Green function:

$$(G_1(t, x_1) = G(t, x_1, 0, 0) \tag{21}$$

where G is the Green's function in (13). To avoid cumbersome notation, we drop the subscript of G_1 in the following discussion. The sampling scheme considered is periodic sampling through a lattice in the space-time plane $\{(x, t) \in \mathbb{R}^2\}$. $G(t, x)$ is an analog field which is digitized in space, time and amplitude. The spatio-temporally discretized field is Gaussian, as is $G(t, x)$. The discrete Gaussian field is then quantized using source coding and its Gaussianity facilitates the application of results from information theory for Gaussian random fields to obtain the R(D) tradeoff. We proceed along the lines of [16] which gives the analysis of multidimensional sampling for the case of deterministic signals and arbitrary geometries of the sampling lattice and which is extended in [13] to the case of random fields.

3.1 Spectral Measure of the Sampled Field

Since for bandlimited random process, sampling in time domain implies replication in the spectral domain, we derive the spectral measure of the sampled field in this subsection. A sampling lattice is characterized by sampling matrix V [16] whose columns form the basis which generates the lattice. The set of points $\Lambda_{\boldsymbol{V}} = \{\boldsymbol{V}\boldsymbol{n} : \boldsymbol{n} = (n_1, n_2)^T \in \mathbb{Z}^2\}$ form the vertices of the lattice where the analog field $G(t, \boldsymbol{x})$ is sampled to form a discrete parameter random field. The discrete parameter field or the sampled Green's function is given by

$$\tilde{G}[\boldsymbol{n}] = G(\boldsymbol{V}\boldsymbol{n}) = G(v_{11}n_1 + v_{12}n_2, v_{21}n_1 + v_{22}n_2) \tag{22}$$

where $\boldsymbol{v}_1 = (v_{11}, v_{21})^T$ and $\boldsymbol{v}_2 = (v_{12}, v_{22})^T$ are the first and second columns of the sampling matrix $\boldsymbol{V}$ respectively.

Under appropriate conditions (chiefly that the spectrum S_H [$= S_G$ for the case of an impulse source] in (20) has compact support) it is guaranteed by the spectral representation theorem [19] that the spectral measure $S_{\tilde{G}}$ of the discrete parameter random field $\tilde{G}$ is given by

$$S_{\tilde{G}}(\boldsymbol{\omega}) = \frac{1}{|\det \boldsymbol{V}|} \sum_{\boldsymbol{k} \in \mathbb{Z}^2} S_G(\boldsymbol{V}^{-T}(\boldsymbol{\omega} - 2\pi \boldsymbol{k})) \tag{23}$$

where $\boldsymbol{\omega} = \boldsymbol{V}^T \cdot (\Omega, \Phi)^T$, with Ω being the temporal frequency, Φ the spatial frequency and $\boldsymbol{V}^{-T}$ the transpose inverse of $\boldsymbol{V}$.

4 Finite Order Rate Distortion Analysis

In this setion we give the R(D) analysis. We begin with the finite order R(D) function. By finite order we mean that the block length of the symbols that are jointly encoded is finite and not in the asymptotic regime. The asymptotic R(D) function is considered in Setion V. Finite order allows us to use parametric formulas for the rate-distortion function that are independent of whether the wavefield is stationary or not. W using reverse water filling [24] to compute the R(D) function. We consider the three basic source coding schemes to arrive at three bounds for the R(D) function as is done in the analysis of [13]. The first is the centralized coding scheme. In the centralized scheme it is assumed that there is zero cost of communication between sensors: every sensor knows the entire data of every other sensor and encodes its data accordingly. In effect we have all the data available at a single "super" sensor. This is the best possible performance that can be achieved and thus forms a lower bound for the R(D) function. The second scheme is the one where all sensors encode their data independently without regard to the correlation between adjacent sensors. This forms an upper bound on the R(D) performance as this is the worst possible that can be done. The third scheme is a multi-terminal source doing scheme in which the sensors do not communicate with each directly, but inter-sensor correlation is exploited jointly by the decoder at the fusion center or recorder unit. This forms a tighter lower bound and is known is the Berger-Tung inner bound. It is known to be tight for some cases.

4.1 Simplification of the Autocorrelation Function for the Green's Function Field

In order to obtain the R(D) tradeoff for the wireless geophone network the autocorrelation function of the Green field $G(t, x)$ is needed. To that end, the expression for $G(t, x)$ as given in Eq. (13) is analyzed and simplified first.

$$\begin{aligned} G(t,\boldsymbol{x}) &= -\frac{1}{c_0^2}\frac{\partial^2}{\partial^2 t}\iint \rho(\boldsymbol{z})G_0(t-s,\boldsymbol{x}-\boldsymbol{z})G_0(s,\boldsymbol{z})dsd\boldsymbol{z} \\ &= -\frac{1}{c_0^2}\frac{\partial^2}{\partial^2 t}\iint \frac{\rho(\boldsymbol{z})}{\|\boldsymbol{z}\|\|\boldsymbol{x}-\boldsymbol{z}\|}\delta\left(t-s-\frac{\|\boldsymbol{x}-\boldsymbol{z}\|}{c}\right)\delta\left(s-\frac{\|\boldsymbol{z}\|}{c}\right)dsd\boldsymbol{z} \\ &= -\frac{1}{c_0^2}\frac{\partial^2}{\partial^2 t}\int \frac{\rho(\boldsymbol{z})}{\|\boldsymbol{z}\|\|\boldsymbol{x}-\boldsymbol{z}\|}\delta\left(t-\frac{\|\boldsymbol{z}\|+\|\boldsymbol{x}-\boldsymbol{z}\|}{c}\right)d\boldsymbol{z} \end{aligned} \tag{24}$$

Considering that the Green's Function Field is only two dimensional according to Eq. (21) and also that the reflectivity function ρ is one dimensional (constant over a given layer at a given depth), we see that the above expression becomes

$$\begin{aligned} G(t,x) = -\frac{1}{c_0^2}\frac{\partial^2}{\partial^2 t}\iiint & \frac{\rho(z_1)}{\sqrt{z_1^2+z_2^2+z_3^2}\sqrt{(x-z_1)^2+z_2^2+z_3^2}} \\ & \times\delta\left(t-\frac{\sqrt{z_1^2+z_2^2+z_3^2}+\sqrt{(x-z_1)^2+z_2^2+z_3^2}}{c}\right)dz_1dz_2dz_3 \end{aligned} \tag{25}$$

Equation (25) can be simplified to:

$$G(t,x) = \frac{4\pi}{x^2}\left[\rho\left(\frac{ct+x}{2}\right)+ct\dot{\rho}\left(\frac{ct+x}{2}\right)\right] \tag{26}$$

Since ρ is a Gaussian random process by assumption, the time derivative of ρ is also Gaussian owing to the fact that differentiation is a linear operation. The linear combination of two terms involving ρ and $\dot{\rho}$ is thus Gaussian, and hence so is $G(t,x)$.

The non-stationary autocovariance function for the Green Function field is given by:

$$\begin{aligned} & R_G(t_1,x_1;t_2,x_2) := \mathbb{E}[G(t_1,x_1)G(t_2,x_2)] \\ & = \frac{\sigma^2 e^{-\frac{(c\Delta t+\Delta x)^2}{4a^2}}(64a^4\pi^2+2a^2c^2(c^2t_1t_2-4c\pi\Delta t^2-4\pi\Delta t\Delta x)-c^4t_1t_2(c\Delta t+\Delta x)^2)}{4a^4x_1^2x_2^2} \end{aligned} \tag{27}$$

where $\Delta t = t_1 - t_2$, $\Delta x = x_1 - x_2$, a is the correlation length and σ^2 is the variation of the fluctuation in the reflectivity ρ. Obtaining the autocovariance function enables the construction of the autocovariance matrix Γ_G for the sampled field $\hat{G}$. Γ_G is used to compute the finite order R(D) tradeoff for a Gaussian random field as given in the work of Kolmogorov [17]:

$$\begin{aligned} R_n(\theta) &= \frac{1}{n}\sum_{k=0}^{n-1}\max\left(0,\frac{1}{2}\log_2\frac{\lambda_k^{(n)}}{\theta}\right) \\ D_n(\theta) &= \frac{1}{n}\sum_{k=0}^{n-1}\min\left(\theta,\lambda_k^{(n)}\right) \end{aligned} \tag{28}$$

where n is the block length of the source symbol string that is being coded. $\lambda_k^{(n)}$ is the kth eigenvalue of the $n \times n$-dimensional autocovariance matrix whos $(l,m)^{\text{th}}$ position is given by $\Gamma_G[l,m] := R_G(z_l; z_m)$ where $z_i \in \{jT_0 : j \in \mathbb{N}\} \times \{kX_0 : k \in \mathbb{N}\}$, for $i = 1, 2, \ldots, n$, with T_0 and X_0 being the temporal and spatial sampling intervals respectively. The advantage of using a finite order R(D) formulation is that it gives a lower bound to the optimal asymptotic performance of any code (bounds the R(D) curve from above). In practice the code length is always finite, so it makes sense to make use of finite order R(D) functions.

4.2 Centralized Compression

As discussed before the geophone array samples the analog field spatio-temporally and then performs data compression (quantization) on the sampled field. The analog field amplitudes are converted into binary data and transmitted over a digital wireless channel. In centralized compression all the field's sample are available at a single spatial location or equivalently the communication cost among sensors is zero. This effectively becomes a point-to-point compression problem. The analog field $G(t,x)$ is first discretized on its support as $\tilde{G}[m,n]$, $(m,n) \in \mathbb{Z}^2$, quantized and then reconstructed as $\tilde{G}_r[m,n]$ at the fusion center. The reconstructed analog field. $G_r(t,x)$ can be obtained from $\tilde{G}_r[m,n]$ using sampling representation given in [3]. The distortion criteria used is mean square error (MSE) as formulated in ([3], Sect. 4.6.3), extended here to the case of a 2D random field:

$$\begin{aligned} D &= \lim_{\substack{T\to\infty \\ X\to\infty}} \int_{-T}^{T}\int_{-X}^{X} [G(t,x) - G_r(t,x)]^2 dt dx \\ &= \lim_{\substack{N\to\infty \\ K\to\infty}} \frac{1}{2N+1}\frac{1}{2K+1} \sum_{m=-K}^{K}\sum_{n=-N}^{N} [\tilde{G}[m.n] - \tilde{G}_r[m,n]]^2 \end{aligned} \tag{29}$$

As pointed out in [13], [3], this is equivalent to finding the distortion normalized by unit time and unit length of the geophone sensor array.

The field to be compressed $\tilde{G}[m,n]$ is a homogeneous Gaussian random field. By simple transposition of the results of Sect. 4.6.3 in [3] for stationary, bandlimited sources to the case of homogeneous random fields, one obtains the following parametric form for the R(D) curve. The main tool used to obtain these expression is the Toeplitz distribution theorem of Szego [15].

$$D_\gamma = \left(\frac{1}{2\pi}\right)^2 \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \min[\gamma, S_{\tilde{G}}(\omega,\phi)] d\omega d\phi \tag{30}$$

and

$$R_\gamma = K\left(\frac{1}{2\pi}\right)^2 \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \max\left[0, \frac{1}{2}\log\frac{S_{\tilde{G}}(\omega,\phi)}{\gamma}\right] d\omega d\phi \tag{31}$$

where $S_{\tilde{G}}(\omega,\phi)$ is determined from Eqs. (18), (20) and (23), γ is a parameter of Eqs. (30) and (31) and K is a normalization constant that depends on the

temporal and spatial cutoff frequency in the support of $S_{\tilde{G}}$ and on the geometry of the sampling lattice.

4.3 Independent Coding

In independent coding all geophones encode and send their data streams directly to the fusion center without communicating with each other and without taking into consideration the correlation among data stream of neighboring sensors. Similarly, the data sent by each individual geophone is decoded/reconstructed separately by the fusion center. In the independent coding case, all geophones observe the same spectral measure of the underlying field $\tilde{G}[m,n]$ given by [13]:

$$S(\omega) = \frac{1}{2\pi}\int_{-\pi}^{\pi} S_{\tilde{G}}(\omega,\phi)d\phi \tag{32}$$

while the R(D) function is now given by

$$D_\gamma = \frac{1}{2\pi}\int_{-\pi}^{\pi} \min[\gamma, S_{\tilde{G}}(\omega]d\omega \tag{33}$$

$$R_\gamma = K\frac{1}{2\pi}\int_{-\pi}^{\pi} \max\left[0, \frac{1}{2}\log\frac{S_{\tilde{G}}(\omega,\phi)}{\gamma}\right] d\omega \tag{34}$$

4.4 Distributed Coding

In distributed source coding the correlation among data streams from different sensors is taken into account while encoding the source and also while performing centralized decoding. All this is done while there is no inter-sensor communication among the geophone network nodes, which makes it of high value for wireless sensor network applications as there is no internode communication cost. The complete rate-distortion region for distributed source coding has not been found to date, however several special cases and bounds for the rate region are known. For nearly lossless coding, the condition of no inter-sensor communication while encoding places no extra cost on the required rate [28]. In the lossy compression case, there is an incurred cost in terms of rate for not communicating between sensors [29]. In [29] a complete characterization of the rate region is also obtained under the assumption of high-resolution quantization. In [31] the rate-distortion region for the two-terminal case with Gaussian source is given. The case for arbitrary number of sources is yet unknown. Since the exact rate-distortion region for distributed source coding remains an open and difficult problem, we evaluate in this paper the inner bound developed by Tung in her PhD thesis, [30]. Under that formulation, the parametric form of rate-distortion function is given by

$$D_P = \left(\frac{1}{2\pi}\right)^2 \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \frac{S_{\tilde{G}}(\omega,\phi)P(\omega)}{S_{\tilde{G}}(\omega,\phi)+P(\omega)} d\omega d\phi \tag{35}$$

and

$$R_P = K\left(\frac{1}{2\pi}\right)^2 \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \frac{1}{2}\log\left(1 + \frac{S_{\tilde{G}}(\omega,\phi)}{P(\omega)}\right) d\omega d\phi \tag{36}$$

The fucntion $P(\omega)$ is a parameter of optimization (the minimum rate R_P subject to distortion level D_P) and can be found using Lagrangian optimization.

5 Analysis for a Plane Wave Scenario

In what follows we analyze a specialized, one dimensional version of the model presented in Sect. 2. A plane wave impinging on a one dimensional slab of length L consisting of randomly layered medium is considered. We focus on this simplified model for illustrative purposes, as the analysis is more straightforward and the results more elucidating. The plane wave analysis also allows us to obtain the asymptotic R(D) function as opposed to the finite order R(D) function obtained in Sect. 4. The model and its various generalizations have been extensively considered in [23] for seismic analysis. The main idea used in [23] is to split the total propagation into upward going and downward going waves and represent the total wave as their sum. Using stochastic calculus, the reflected wave from the random medium interface can be shown to be a locally stationary Gaussian process, whose time-varying power spectral density can be calculated in closed form, provided the medium has a constant background density and bulk modulus. This reflected Gaussian process is an asymptotic limit for the case of infinite number of layers in the slab, that is, this "ϵ"-process converges to a Gaussian process, as $\epsilon \to 0$, where ϵ is the width of the plane-wave pulse propagating from the source into the medium. The pulse width is assumed to be large compared to the width of the layers yet small with repsect to L, the width of the slab. The width of layers is taken to be on the order of ϵ^2.

We consider the anlaysis given in [32]. Let $\mathbf{u}(t,\mathbf{x})$ denote the velocity and $p(t,\mathbf{x})$ denote the pressure wiithing the subsurface. The linearized acoustic wave equations are then given by

$$\begin{aligned} \rho(z)\frac{\partial \mathbf{u}}{\partial t} + \nabla p &= \mathbf{F}(t,\mathbf{x}) \\ \frac{1}{K(z)}\frac{\partial p}{\partial t} + \nabla\cdot\mathbf{u} &= 0 \end{aligned} \tag{37}$$

where $\rho(z)$ is the density and $K(z)$ is the bulk modulus. $\mathbf{F}(t,\mathbf{x})$ is the forcing term due to source at the surface. For simplicity we assume that in the one dimensional case the source term is a plane wave given by

$$\mathbf{F}(t,\mathbf{x}) = f(t)\delta(x)\delta(y)\delta(z)\mathbf{e} \tag{38}$$

where $\mathbf{e} = (0, 0, 1)$ is a unit directivity vector, $(0, 0, 0)$ is the position of the source and $f(t)$ is a pulse function with compact support. The random medium occupies a slab of thickness L in the lower half space $z \leq 0$. The parameters ρ and K are constant throughout this half space region. We specify the boundary conditions for the system (24) and rewrite it as (in the 1-D case)

$$\begin{aligned} \rho(z)\frac{\partial u}{\partial t} + \frac{\partial p}{\partial z} &= 0 \\ \frac{1}{K(z)}\frac{\partial p}{\partial t} + \frac{\partial u}{\partial z} &= 0 \end{aligned} \tag{39}$$

The material properties $\rho(z)$ and $K(z)$ fluctuate randomly around background smooth functions $\rho_0(z)$ and $K_0(z)$ governed by following relationship

$$\begin{aligned} \rho(z) &= \rho_0(z)\left(1 + \eta(\frac{z}{\epsilon^2})\right) \\ K(z) &= K_0(z)\left(1 + \nu(\frac{z}{\epsilon^2})\right) \end{aligned} \tag{40}$$

where $\eta(\frac{z}{\epsilon^2})$ and $\nu(\frac{z}{\epsilon^2})$ are zero-mean ergodic processes whose fluctuations are the scale ϵ^2. The effective medium assumption [23] is in force within the random medium so that $\rho_0(z) = \rho_0$ and $K_0(z) = K_0$ are constant background material properties. Continuity conditions at $x = 0$ and $x = L$ are imposed. The boundary conditions that are needed to completely specify the system (24) are given by

$$\begin{aligned} A_\epsilon(z, t) &= u(z, t) + \frac{p(z, t)}{\sqrt{\rho_0 K_0}} \\ B_\epsilon(z, t) &= u(z, t) - \frac{p(z, t)}{\sqrt{\rho_0 K_0}} \\ A_\epsilon(0, t) &= \frac{1}{\sqrt{\epsilon}} f\left(\frac{t}{\epsilon}\right) \\ B_\epsilon(L, t) &= 0 \end{aligned} \tag{41}$$

where A_ϵ is the wave traveling into the ground and B_ϵ is the reflected wave. We are primarily interested in the random process $B_\epsilon(0, t)$. It can be shown [32] that $B_\epsilon(0, t + \epsilon\sigma)$ converges in distribution, to a zero-mean, locally stationary Gaussian process $R_{f,t}(\sigma)$ as $\epsilon \to 0$. We are interested in the correlation function of this reflected process. Accordingly, we carry out the following analysis. The windowed correlation function is given by

$$C_f^\epsilon(t, \sigma) = \mathbb{E}[B_\epsilon(0, t - \frac{\epsilon}{2}\sigma) B_\epsilon(0, t + \frac{\epsilon}{2}\sigma)]$$

Then as discussed, we have the result that

$$C_f^\epsilon(t, \sigma) = \lim_{\epsilon \to 0} C_f^\epsilon(t, \sigma) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{i\omega\sigma} |\hat{f}(\omega)|^2 \Lambda(\omega, t) d\omega \tag{42}$$

where $\hat{f}(\omega)$ is the Fourier transform of the source pulse $f(t)$ and $\Lambda(\omega, t)$ is the local power spectral density of the reflected (Gaussian) process from the randomly layered medium. In our case of constant background material properties, it is given by [32]

$$\Lambda(\omega, t) = \frac{\alpha\omega^2}{(1 + \alpha\omega^2 t)^2} \tag{43}$$

where α depends on the matrial properties and on the speed of sound $c_0 = \sqrt{\frac{K_0}{\rho_0}}$ in the subsurface. It is given by the following expression [32],

$$\alpha = \frac{1}{4}\int_0^\infty \mathbb{E}[(\eta(0) - \nu(0))(\eta(z) - \nu(z)]dx \tag{44}$$

We consider a Dirac impulse as the source so that $\hat{f}(\omega) = 1$. The autocorrelation function of the process can be obtained in closed form from the local power spectral density $\Lambda(\omega, t)$by Fourier inverting it with respect to the ω variable and treat t as the center of the time window with edges at $-\frac{\sigma}{2}.\frac{\sigma}{2}$. The expression thus obtained for the non-stationary autocorrelation function is thus given as

$$R(t_1, t_2) = \mathbb{E}[B(0, t_1)B(0, t_2)] = \frac{e^{-\frac{\sqrt{2}|t_1 - t_2|}{\sqrt{\alpha(t_1 + t_2)}}}\left(\sqrt{2\alpha}(t_1 + t_2) - 2|t_1 - t_2|\sqrt{t_1 + t + 2}\right)}{2\alpha(t_1 + t_2)^{5/2}} \tag{45}$$

This autocorrelation function can be used to find the R(D) function for the process. To that end, we used the approach of [8] in which the R(D) pair given parametrically by [17] is extended to the non-stationary case. Accordingly, we have

$$\begin{aligned} D_\gamma &= \int_0^\infty \min[\gamma, \lambda(f)]df \\ R(D_\gamma)) &= \int_0^\infty \max[0, \log(\sqrt{\lambda(f)/\gamma})]df \end{aligned} \tag{46}$$

where $\lambda(f)$ is the power spectrum given by defined as [8]

$$\lambda(f) = \lim_{\substack{k\to\infty, b-a\to\infty \\ k/(b-a)\to f}} \lambda_k(a, b) \tag{47}$$

where $\lambda_k(a, b)$ is the kth eigenvale of the Karhunen-Loeve (KL) integral equation

$$\lambda g(t) = \int_a^b R(t, s)g(s)ds \tag{48}$$

Up till now we have assumed an infinite receiver line with an infinite number of geophones on it. In practical seismic surveys this is not valid. To model this scenario, we iintroduce the use of a windowing function on the sensed field. A rectangular window in the spatial domain is applied to the green function. In this

way the spectrum of the green function is a convolution of the original spectrum and the sinc function. This is computed numerically and then the computations are carried out for the R(D) function, the results of which follow in the next section.

6 Numerical Analysis and Simulations

The results obtained through our analysis are presented next. The results are a result of numerical analysis and simulations of the equationns of rate distortion function and autocorrelation functions of the reflected random field at the suface that is sampled by the sensor network in space and time.

In Fig. 2, the time varying power spectral density of the reflected Gaussian process emanating from the randomly layered medium is plotted. Different snapshots are given, it is seen as that time progresses, or as the center of window of the spectral density moves forward, the power decays. The value of the parameter α that is used is equal to 1.

Figure 3 gives the autocorrelation function of the reflected random process. This autocorrelation function is given in closed form by Eq. (45) and is derived from the power spectral density given by Eq. (43). The autocorrelation function is symmetric in both of its arguments and has a singularity at the origin.

The power spectrum as the solution of the eigenvalue problem for the Karhunen-Loeve equation is plotted in Fig. 4. The frequency variable f is the ratio of the eigenvalue index to the length of the interval on which the KL expansion is sought.

In Fig. 5, the asymptotic R(D) curve is given for $\alpha = 1$. Complete finite order R(D) tradeoff is obtained for a scenario where the wave propagation speed underground is 400 m/s and the correlation length a is 10 m.

It can be seen from Fig. 6 that in the limit of very low bit rate compression (on the order of 0.05 bits per sample) the distortion is practically infinite. However, for moderate compression ratios (12 bits per sample) the distortion is in the order of .001 MSE. While these results seem pessimistic somewhat, it should be noted that they are for a moderate block length of 625. In actual compression larger block lengths are to be expected.

In Fig. 7, the three bounds of independent, centralized and distributed source coding are given. The bounds have an intersection at a certain point along the distortion axis. Figure 8 gives the effect of finite aperture length on coding performance. The number of sensors are now limited and the sensor density increases with the number of total sensors as the receiver line is of finite length. We see that as we reduce the number of sensors the rate-distortion tradeoff detiorates.

The effect on compression performance for different block lengths can be seen in Fig. 9. Here there is a marked improved in compression performance as

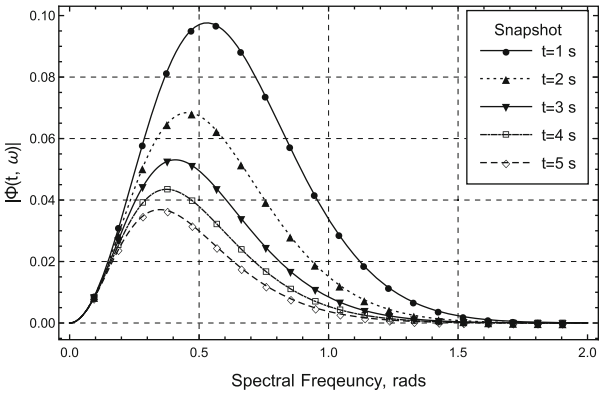

Fig. 2. Locally stationary windowed power spectral density of the reflected Gaussian process with $\alpha = 1$ at various time snapshots t which index the center the of the window. The excitation function is a Gaussian pulse

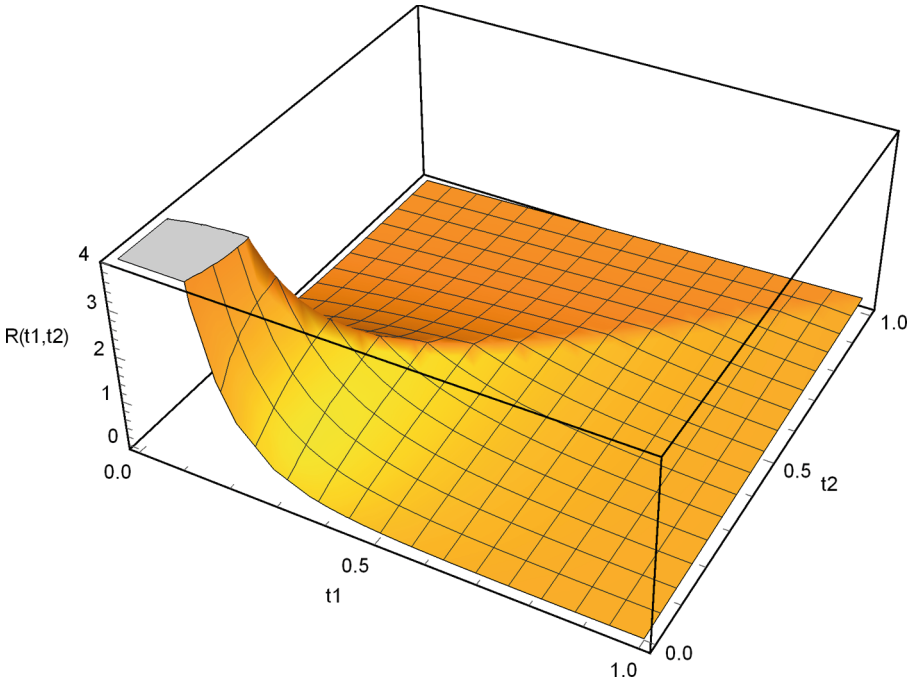

Fig. 3. Plot of the autorcorrelation surface of the reflected process with time indices t_1 and t_2 as argument, plotted for $\alpha = 1$

the block length increases with a distortion of .0001 MSE with 5 bits/sample allocation for a block length of 2401, which is quite good.

In Fig. 10 there is a plot of R(D) function for various correlation lengths between the random layers. The rate disotortion curves for a 100th order $R(D)$

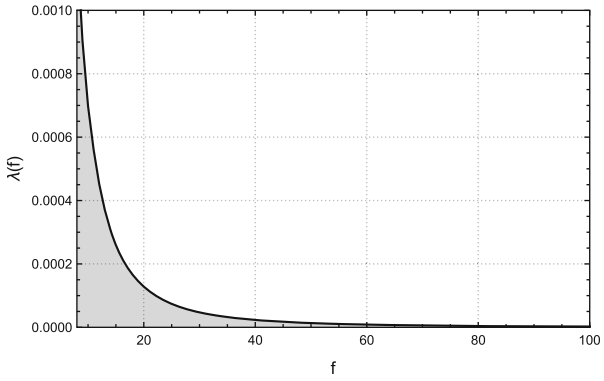

Fig. 4. Plot of the power spectrum $\lambda(f)$ for the reflected Gaussian process. The power spectrum is the eigen-solution to the KL equation

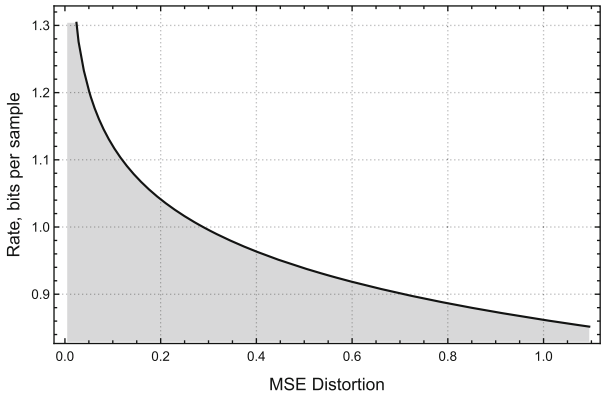

Fig. 5. Complete asymptotic R(D) curve for the reflected Gaussian process

function are given. The curves corresponds to a correlation length ranging from 10m (highly uncorrelate) to 1000m (highly correlated case). It is seen that as correlation increases the R(D) performance improves by a factor of almost 4 across the range for correlation lengths.

Figure 11 plots the R(D) surface as a function of sound speed under ground. Two surfaces are given, one for centralized coding and the other for distributed coding. As the sound speed increases the R(D) tradeoff improves.

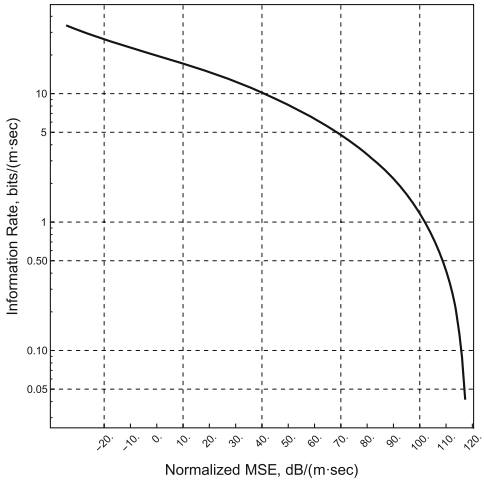

Fig. 6. Complete finite-order R(D) tradeoff for wavespeed $c_0 = 400$ m/s, correlation length $a = 10$ m and $\sigma^2 = 1$. The order of the R(D) function n, is 625.

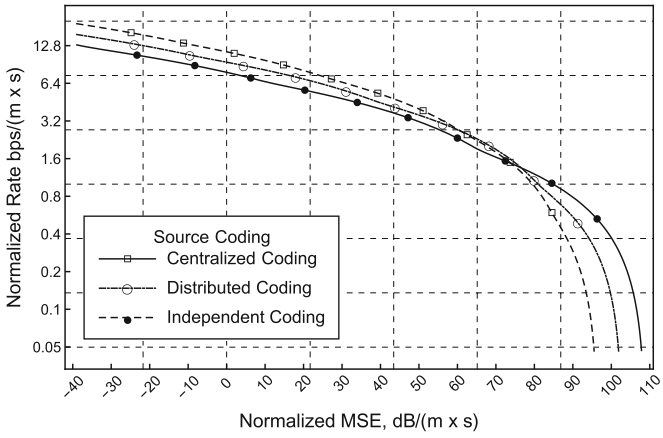

Fig. 7. R(D) bounds for indepdent, centralied and distributed source coding, the sensor line is assumed to be infinte, sound speed is c_0 is 1000 m/s

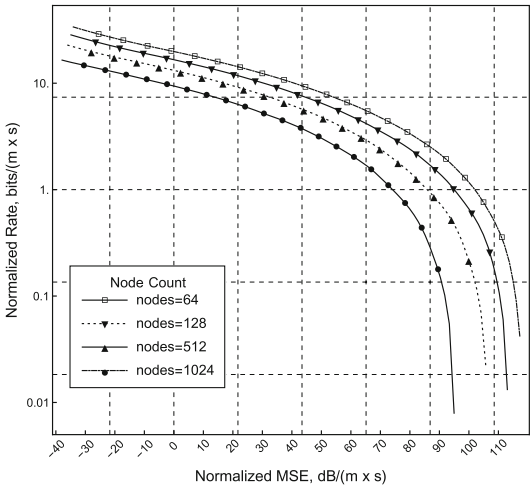

Fig. 8. R(D) for a finite receiver line. The node counts are given and vary between 64 to 1024 nodes.

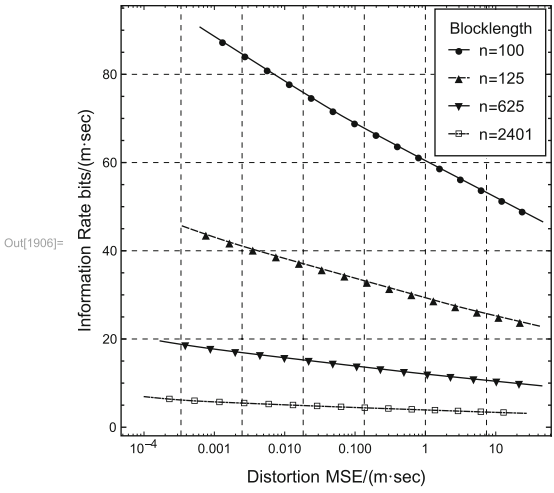

Fig. 9. Rate-distortion functions for various orders n or block lengths, $c_0 = 1000$ m/s, $a = 10$m

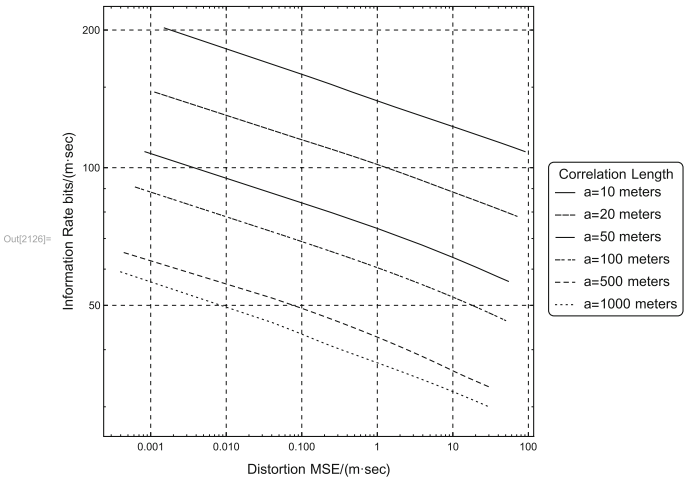

Fig. 10. Rate-distortion functions for various correlation lengths with a block length of 100m and $c_0 = 1000$m/s

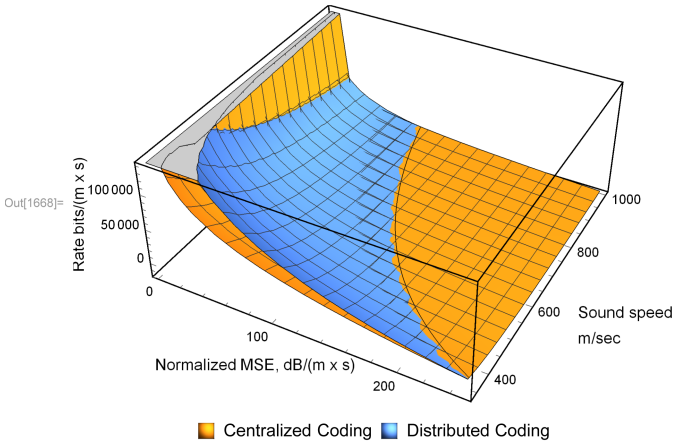

Fig. 11. R(D) surface as function of sound speed under ground

7 Conclusion

In this work we developed a detailed R(D) analysis for signal compression of seismic wavefields in a wireless geophone network. The analysis was based on a physical model of a randomly layered subsurface earth. The acoustic impedance of in the subsurface was assumed to vary according to a homogeneous Gaussian random field, with a given autocorrelation structure. Energy propagation in the earth was governed by a random wave equation involving the acoustic pressure. The Green's function was obtained for this random wave equation and its action on an impulse was seen to be a result of a random integral operator.

The R(D) analysis was built upon the autocorrelation function of this random operator. A finite order R(D) analysis for a two-dimensional subsurface medium and an asymptotic R(D) analysis for a one-dimensional medium (plane wave case) was carried out. The finite order R(D) analysis gives an indication of the of the compression performance for finite source code block lengths, whereas the asymptotic R(D) curve describes the absolute optimum performance achievable by any source code. We have presented and compared both. The final results obtained consisted of the R(D) function on an infinite receiver line, the R(D) function on a finite line (fixed number of nodes), R(D) function variation vs. the length of the coded block and the R(D) function vs. the physical model parameters like correlation length of the acoustic impedance process and the speed of sound underground. It is see that the compression performance improves as the coding block length increases and as the correlation length of the underlying random process that describes the earth's impedance increases. The rate distortion performance also improves as the speed of the sound in the subsurface increases. The study has particular significance as a guideline to the design of practical signal compression algorithms for seismic acquisition.

References

1. Gleich, D., Planinsic, P., Gergic, B., Cucej, Z.: Progressive space frequency quantization for SAR data compression. IEEE Trans. Geosci. Remote Sens. **40**(1), 3–10 (2002)
2. Gueguen, L., Datcu, M.: Image time-series data mining based on the information-bottleneck principle. IEEE Trans. Geosci. Remote Sens. **45**(4), 827–838 (2007)
3. Valsesia, D., Magli, E.: A novel rate control algorithm for onboard predictive coding of multispectral and hyperspectral images. IEEE Trans. Geosci. Remote Sens. **52**(10), 6341–6355 (2014)
4. Huang, K., Dai, D.: A new on-board image codec based on binary tree with adaptive scanning order in scan-based mode. IEEE Trans. Geosci. Remote Sens. **50**(10), 3737–3750 (2012)
5. Savazzi, S., Spagnolini, U., Goratti, L., Molteni, D., Latva-aho, M., Nicoli, M.: Ultra-wide band sensor networks in oil and gas explorations. IEEE Commun. Mag. **51**(4), 150–160 (2013)
6. Savazzi, S., Spagnolini, U.: Compression and coding for cable-free land acquisition systems. Geophysics **76**(5) (2011)
7. Berger, T.: Rate Distortion Theory: A Mathematical Basis for Data Compression, 1st edn. Prentice-Hall, Englewood Cliffs (1971)
8. Burger, T.: Information rates of Wiener processes. IEEE Trans. Inf. Theory **16**(2), 134–139 (1970)
9. Dragotti, P., Gastpar, M.: Distributed Source Coding: Theory, Algorithms, and Applications, 1st edn. Elsevier Inc., Burligton (2009)
10. Pradhan, S., Ramachandran, K.: Distributed source coding using syndromes (DISCUS): design and construction. In: Proceedings of the 1999 IEEE Data Compression Conference, Snowbird (1999)
11. Gastpar, M., Rimoldi, B., Vetterli, M.: To code, or not to code: lossy source-channel communication revisited. IEEE Trans. Inf. Theory **49**(5), 1147–1158 (2003)

12. Gastpar, M., Dragotti, P., Vetterli, M.: The distributed Karhunen-Loève transform. IEEE Trans. Inf. Theory **52**(12), 5177–5196 (2006)
13. Konsbruck, R., Telatar, E., Vetterli, M.: On sampling and coding for distributed acoustic sensing. IEEE Trans. Inf. Theory **58**(5), 3198–3214 (2012)
14. Beferull-Lozano, B., Konsbruck, R.: On source coding for distributed temperature sensing with shift-invariant geometries. IEEE Trans. Commun. **59**(4), 1053–1065 (2011)
15. Grenander, U., Szego, G.: Toeplitz Forms and Their Applications, 1st edn. UC Berkley Press, Berkeley and Los Angeles (1958)
16. Dudgeon, D., Mersereau, R.: Multidimensional Digital Signal Processing, 1st edn. Prentice-Hall, Upper Saddle River (1984)
17. Kolmogorov, A.: On the Shannon theory of information transmission in the case of continuous signals. IRE Trans. Inf. Theory **2**(4), 102–108 (1956)
18. Adler, R.: The Geometry of Random Fields, 1st edn. Wiley, Chichester (1981)
19. Doob, J.: Stochastic Processes, 1st edn. Wiley, Chichester (1953)
20. Grigoriu, M.: Evaluation of Karhunen-Loeve, spectral, and sampling representations for stochastic processes. J. Eng. Mech. **132**(2), 179–189 (2006)
21. Gersho, A., Gray, R.: Vector Quantization and Signal Compression. Kluwer, Boston (1992)
22. Gray, R., Neuhoff, D.: Quantization. IEEE Trans. Inf. Theory **44**(6), 2325–2383 (1998)
23. Fouque, J.-P., Garnier, J., Papanicolaou, G., Solna, K.: Wave Propagation and Time Reversal in Randomly Layered Media. Springer, New York (2007)
24. Cover, T.M.: Elements of Information Theory. Wiley-Interscience, Hoboken (2006)
25. Lippmann, B.A., Schwinger, J.: Variational principles for scattering processes. I. Phys. Rev. **79**(3), 469–480 (1950)
26. Bellman, R.: Stochastic process in mathematical physics and engineering. In: Proceedings of Symposia in Applied Mathematics, American Mathematical Society, Providence (1963)
27. Evans, L.C.: Partial Differential Equations. American Mathematical Society, Providence (2002)
28. Slepian, D., Wolf, J.K.: Noiseless coding of correlated information sources. IEEE Trans. Inf. Theory **19**(4), 471–480 (1973)
29. Zamir, R., Berger, T.: Multiterminal source coding with high resolution. IEEE Trans. Inf. Theory **45**(1), 106–117 (1999)
30. Tung, S. Y.: Multiterminal Source Coding. Ph.D. thesis Cornell University, Ithaca (1978)
31. Wagner, A.B., Tavildar, S., Viswanath, P.: Rate region of the quadratic Gaussian two-encoder source-coding problem. IEEE Trans. Inf. Theory **54**(5), 1938–1961 (2008)
32. Asch, M., Kohler, W., Papanicolaou, G., Postel, M., White. B.: Frequency content of randomly scattered signals. SIAM Rev. **33**, 526–629 (1991)