

An inquiry into the lunar interior: A nonlinear inversion of the Apollo lunar seismic data

A. Khan

Department of Geophysics, Niels Bohr Institute, University of Copenhagen, Denmark

Département de Géophysique Spatiale et Planétaire, Institut de Physique du Globe de Paris, France

K. Mosegaard

Department of Geophysics, Niels Bohr Institute, University of Copenhagen, Denmark

Received 22 September 2001; revised 6 December 2001; accepted 6 December 2001; published 11 June 2002.

[1] This study discusses in detail the inversion of the Apollo lunar seismic data and the question of how to analyze the results. The well-known problem of estimating structural parameters (seismic velocities) and other parameters crucial to an understanding of a planetary body from a set of arrival times is strongly nonlinear. Here we consider this problem from the point of view of Bayesian statistics using a Markov chain Monte Carlo method. Generally, the results seem to indicate a somewhat thinner crust with a thickness around 45 km as well as a more detailed lunar velocity structure, especially in the middle mantle, than obtained in earlier studies. Concerning the moonquake locations, the shallow moonquakes are found in the depth range 50–220 km, and the majority of deep moonquakes are concentrated in the depth range 850–1000 km, with what seems to be an apparently rather sharp lower boundary. In wanting to further analyze the outcome of the inversion for specific features in a statistical fashion, we have used credible intervals, two-dimensional marginals, and Bayesian hypothesis testing. Using this form of hypothesis testing, we are able to decide between the relative importance of any two hypotheses given data, prior information, and the physical laws that govern the relationship between model and data, such as having to decide between a thin crust of 45 km and a thick crust as implied by the generally assumed value of 60 km. We obtain a Bayes factor of 4.2, implying that a thinner crust is strongly favored.

INDEX TERMS: 6250 Planetology: Solar System Objects: Moon (1221); 5430 Planetology: Solid Surface Planets: Interiors (8147); 3260 Mathematical Geophysics: Inverse theory; 5455 Planetology: Solid Surface Planets: Origin and evolution; **KEYWORDS:** Moon, inverse theory, seismology, interior structure, terrestrial planets

1. Introduction

[2] Using seismology to obtain information about the interior of the Moon saw its advent with the U.S. Apollo missions which were undertaken from July 1969 to December 1972. Seismic stations were deployed at five of the six locations (Apollo 17 did not carry a seismometer) as part of the integrated set of geophysical experiments called the Apollo Lunar Surface Experiment Package (ALSEP). Only four of these five stations (12, 14, 15, and 16), powered by radioactive thermal generators, operated concurrently as a four-station seismic array, which was operative from April 1972 until 30 September 1977, when transmission of seismic data was suspended. Each seismic package consisted of three long-period (LP) seismometers aligned orthogonally to measure one vertical (Z) and two horizontal (X and Y) components of surface motion. The sensor unit also included a short-period seismometer which was sensitive to vertical motion at higher frequencies (for more instrumental details,

see *Latham et al.* [1969]). Digital data were transmitted continuously from the lunar surface to receiving stations on Earth and were stored on magnetic tapes for subsequent analysis. All the seismic data were then displayed on a compressed timescale format from which lunar seismic signals were identified [*Nakamura et al.*, 1980].

[3] The lunar seismic network spans the near face of the Moon in an approximate equilateral triangle with 1100 km spacing between stations with two seismometers placed 180 km apart at one corner (see Figure 1) and covers most geological settings on the front side of the Moon. Since the first mission, more than 12,000 events have been recorded and catalogued over a period of 8 years, and it has taken several more to evaluate the data [*Nakamura et al.*, 1981]. Since the compilation of the recorded seismograms, it has been shown that the Moon is very aseismic compared to the Earth. In comparison to the Earth the energy released in seismic activity is 8 orders of magnitude less, being about $10^{10} \text{ J yr}^{-1}$, compared to $10^{18} \text{ J yr}^{-1}$ by earthquakes [*Goins et al.*, 1981a], although *Nakamura* [1980] has pointed out that the actual average lunar seismic energy release could be as high as $10^{14} \text{ J yr}^{-1}$. Most of the moonquakes are very small

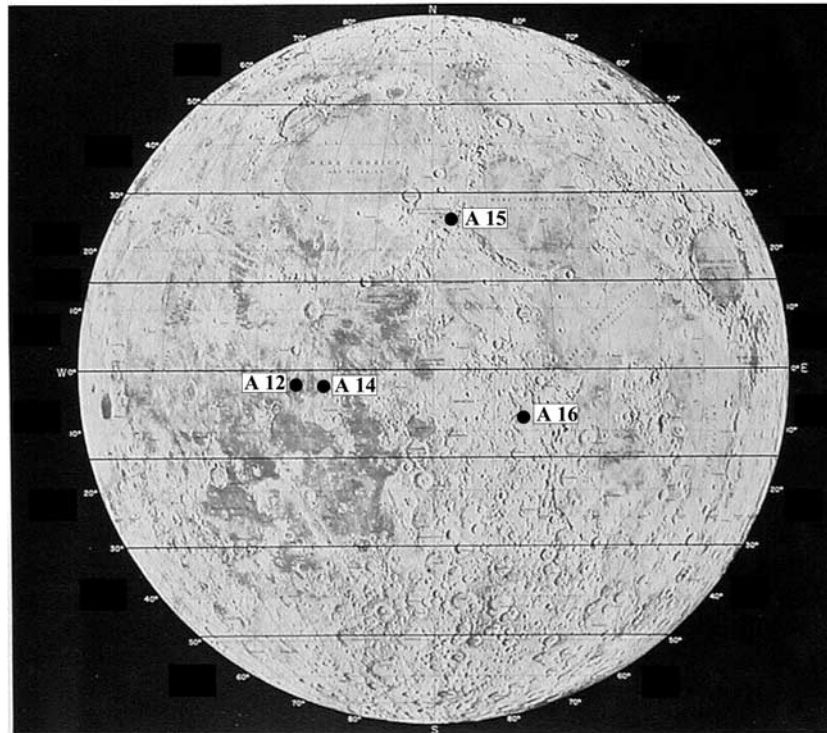


Figure 1. Location map showing the four seismic stations that operated simultaneously.

on the Richter scale with magnitudes ranging up to 4 for shallow events [Nakamura *et al.*, 1974a], whereas the deep events were generally of magnitude <2 [Goins *et al.*, 1981a]. Because of the high sensitivity of the seismometers and the low level of microseismic background noise, however, each station detected on the average between 650 and 1250 moonquakes per year.

[4] Upon examination of the first lunar seismic data returned to Earth in 1969, it became evident that their interpretation would be a somewhat more intricate process, owing to an apparent complexity inherent in lunar seismograms. It turned out that the lunar seismic signals were characterized by being very long, of high frequency, and of reverberating nature with small first arrivals and slowly building amplitudes [Latham *et al.*, 1972]. Unlike the Earth, where seismic pulses in general are of rather short duration, of the order of minutes, the most prominent feature of lunar signals is their anomalous long continuance. Strong signals, as those from the impacts of the upper stage of the Saturn rocket, last several hours. Moonquake and meteoroid impact signals typically continue for 30 min to 2 hours. The general picture that appears upon deciphering lunar seismograms from a meteoroid impact is the following [Lammlein *et al.*, 1974]: lunar signals have emerging beginnings, increase gradually to a maximum, and then slowly decay. Following the first one or two cycles of the *P* wave, ground motion is very complex, with little or no correlation between any two components. The onset of a shear wave from an impact signal is indistinct where it can be identified at all. Coherent surface wave trains displaying dispersion have not been recognized in any recordings to date, although it is believed that scattered surface waves undoubtedly contribute to the signals [Toksöz *et al.*, 1974]. It has been suggested that these signals are caused

by intense scattering of the waves in the uppermost layers of the lunar crust [e.g., Latham *et al.*, 1970]. Topographic features, lunar regolith, compositional boundaries, and especially joints and cracks in the crust become very efficient scatterers in the absence of water and the absence of damping. The seismic velocity increases markedly after the first 10–15 km, and deeper material is believed to be sufficiently homogeneous to transmit seismic waves with little scattering. In this lunar environment, seismic waves generated by an impact are intensively scattered near the impact point. Scattered energy gradually diffuses into the lunar interior, in which it propagates normally and undergoes further scattering where it reenters the lunar surface layer. As a result of this, a prolonged wave train with gradual rise and decay is observed at a distant seismic station [Lammlein *et al.*, 1974].

[5] Four distinct types of events have been identified. They are deep moonquakes, shallow moonquakes, and thermal moonquakes, all of which reflect the present dynamic state of the lunar interior, and of course meteoroid impacts. For a summary of the catalogued events detected on the LP seismograms during the operation of the network, see Nakamura *et al.* [1981]. The deep moonquakes, by far the most numerous of events, are usually ~ 1 on the Richter scale [Goins *et al.*, 1981a; Lammlein, 1977] and were found to be located halfway toward the center of the Moon in the depth range 700–1200 km [Nakamura *et al.*, 1982]. They consist of repetitive moonquakes that emanate from specific source regions, and many nearly identical wave trains have been observed [Lammlein *et al.*, 1974]. This important observation meant that the locations are fixed, and moreover, it allowed the summing of a large number of moonquake signals, improving the signal-to-noise ratio. At most hypocenters, one moonquake occurred for a period of a few days during a

fixed time in the monthly lunar tidal cycle, giving rise to peaks at 27-day intervals of the observed lunar seismic activity. In addition, a 206-day variation and a 6-year variation in the activity, also due to tidal effects, such as the solar perturbation of the lunar orbit, have been observed [Lammlein *et al.*, 1974; Lammlein, 1977]. These periodicities have been taken as evidence that the deep focus moonquakes are related to the tidal forces acting on the Moon and appear to represent merely a process of storage and release of tidal energy without a significant release of tectonic energy [Nakamura, 1978; Koyama and Nakamura, 1980].

[6] The shallow moonquakes are a manifestation of another type of natural lunar seismicity. These are the most energetic seismic sources observed on the Moon, although they are less abundant than the other types of seismic events, with an average of 4 events per year [Nakamura, 1977]. They are also known as high-frequency-telesismic (HFT) events owing to their unusually high frequency content and the great distances at which they are observed [Nakamura *et al.*, 1974a]. The estimation of the source depth at which HFT events occur has been inconclusive, although several lines of evidence, such as the variation of the observed amplitude of HFT signals with distance, suggested that they originated no deeper than a few hundred kilometers [Nakamura *et al.*, 1979]. It has been pointed out that there is no clear correlation between them and the tides, as is obvious for deep moonquakes [Nakamura, 1977]. This led to the conclusion that their origin is likely to be tectonic, given their similarity to intraplate earthquakes [Nakamura *et al.*, 1979, 1982]. Their mechanism, though, could not be explained by plate motion, because of the lack of concentration of events into narrow belts as is observed on the Earth.

[7] A large percentage of the events observed on the short period components are very small moonquakes, occurring with great regularity. It is believed that these events, also termed thermal moonquakes, are triggered by diurnal thermal variations [Duennebieer and Sutton, 1974].

[8] With the completion of processing of all the lunar seismic data collected during the Apollo seismic network operation, a set of arrival times was obtained, constituting a primary data set from which the interior velocity structure of the Moon could be inferred. The most recent and concise summary of the seismic velocity profile is based on the complete 5-year data set acquired when the four Apollo seismometers were simultaneously operative [Nakamura *et al.*, 1982; Nakamura, 1983]. Nakamura and coworkers have used arrival times from various seismic events for the elucidation of the seismic velocity profile of the lunar interior. Analysis of the man-made impacts led to a model of the shallow lunar structure, a model for the upper mantle was constructed on the basis of the shallow moonquakes and the meteoroid impacts, and modeling of the deep lunar interior was subject to deep moonquake data. A linearized least squares inversion technique using arrival time data from 41 deep moonquakes, 7 artificial impacts, 18 meteoroid impacts, and 14 shallow moonquakes as compared to only 24 deep moonquake sources used by Goins *et al.* [1981b] was applied in steps. These models established the Moon as a highly differentiated body, with a crust and a mantle whose lower parts were thought to be partially molten [Nakamura *et al.*, 1973]. However, velocity variations and the depths of possible discontinuities in the mantle could not be addressed,

although they were believed to be present [Nakamura, 1983]. The central part of the Moon could also not be ascertained from the seismic data owing to the distribution of seismic sources. All confirmed deep moonquakes occurred on the nearside, except for one source, A₃₃, which is located on the farside beyond the eastern limb. No big meteoroid impacts antipodal to the stations giving rise to unequivocal arrivals were detected, leaving the important question of the existence of a lunar core unanswered, although it has to be noted for the sake of completeness that a farside impact, almost diametrically opposite to station 15, gave tentative evidence for a low-velocity core with a radius around 400 km [Nakamura *et al.*, 1974b; Sellers, 1992].

[9] In the earlier studies mentioned above, the analysis of the nonlinear inverse problem was primarily centered on the construction of best fitting models, thereby obviating the analysis of uncertainty and nonuniqueness, which are important items when inferring scientific conclusions from inverse calculations. We [Khan *et al.*, 2000] presented a new *P* and *S* wave velocity structure for the Moon from a Monte Carlo inversion of the Apollo lunar seismic data, where we adopted a Bayesian viewpoint on inverse problems [Tarantola and Valette, 1982; Mosegaard and Tarantola, 1995] whose main feature is the use of probabilities to describe the model parameters (which are the ones we invert for) and what lends itself to their description is the a posteriori probability density in the model space which summarizes all information about the model we are studying supplied by data, a priori information, and the physical laws relating model and data. Given that for general inverse problems the shape of the posterior probability distribution is not known, it cannot simply be described by mathematical means and covariances. The Markov Chain Monte Carlo (MCMC) algorithm, on the other hand, samples a large suite of models from this probability distribution, thereby rendering us with a better representation. The main purpose of the present study is, on the one hand, to detail the method of analysis underlying that investigation and moreover to extend the Bayesian analysis carried out by Khan *et al.* [2000] (a detailed discussion of the results has already been given in that study and will not be reiterated here) and as an application of this to investigate the suite of models sampled via the MCMC algorithm as to lunar crustal thickness using Bayesian hypothesis testing [Bernardo and Smith, 1994]. In regard to the data set, no attempt was made to identify new events or arrivals, and the data used in this study are the same events as those considered in the study by Nakamura [1983], comprising first arrivals of *P* and *S* waves. However, seismograms from these events have been reviewed in order to assess the uncertainty and consistency on these arrivals. Finally, as regards the outline of the present manuscript, we have chosen first to present general ideas concerning (1) the use of MCMC algorithms to solve the general inverse problem and (2) the analysis of the posterior distribution. Upon this follows a detailed description of the application of these methods to the Apollo lunar seismic data set.

2. Theory: General Ideas

2.1. Solving the General Inverse Problem Using a MCMC Algorithm

[10] It is customary to commence an investigation such as this one by delineating our physical system by a set of

model parameters, $\mathbf{m} = (m_1, m_2, \dots, m_s)$, which completely define the system. These parameters are not directly measurable. What we usually are in possession of, though, are certain observable data, $\mathbf{d} = (d_1, d_2, \dots, d_n)$, obtained through physical measurements. Now, what we could do next is to try to prognosticate the observable parameters by using any theory applicable to our particular predicament. This results in another set of parameters, termed calculated data \mathbf{d}_{cal} , which are dependent on the model parameters. This dependency can be depicted by a relation of the form

$$\mathbf{d} = g(\mathbf{m}), \quad (1)$$

g being a functional relation governing the physical laws that correlate model and data. What is meant by solving the forward problem, then, is the prediction of observable data given a set of model parameters, and conversely, solving the inverse problem is understood to be the inference of values of the model parameters, given observable data. Central to our method is the notion of a state of information over the parameter set. In concordance with the most general description of states of information over a given parameter set, as presented by *Tarantola and Valette* [1982] and *Tarantola* [1987], we shall be employing probability densities over the corresponding parameter space, describing the various states of information inherent to our system. This knowledge embodies the results of measurements of the observable parameters and the a priori information on model parameters as well as the information on the physical correlations between observable and model parameters. Solving the inverse problem, then, shall be formulated as a problem of combining all this information into an anterior state, termed the posterior probability density. The extensive use of probability densities for delineating any information has the advantage of presenting the solution to the inverse problem in the most generic way, thereby implicitly incorporating any nonlinearities [*Tarantola and Valette*, 1982].

[11] It is clear, then, that the most general method for solving nonlinear inverse problems needs an extensive exploration of the model space, since the posterior probability density in the model space contains all the information about the system being studied. Therefore, given probabilistic prior information on \mathbf{m} and a statistical description of the observational uncertainties of \mathbf{d} , the main idea is to design a random walk in the model space which samples the posterior probability distribution, that is, samples models which are consistent with data as well as prior information. To this end, we shall use a Markov Chain Monte Carlo algorithm of the following form (the basic premises underlying the use of MC algorithms to solve general inverse problems are reviewed by *Mosegaard* [1998]):

1. Propose a new model, \mathbf{m}_{pert} , by taking a step of a random walk to some current model \mathbf{m}_{cur} , with a probability proportional to $\rho(\mathbf{m})$, the prior probability density on the model parameters.

2. Calculate the likelihood function for the new model using $L(\mathbf{m}) = k \cdot \exp(-S(\mathbf{m}))$, where k is a normalization constant, $S(\mathbf{m})$ is the misfit function, and $L(\mathbf{m})$ is a measure of the degree of data fit.

3. Accept the new model with a probability $P_{\text{acc}} = \min(1, L(\mathbf{m}_{\text{pert}})/L(\mathbf{m}_{\text{cur}}))$.

4. If \mathbf{m}_{pert} is accepted, then $\mathbf{m}_{\text{cur}} = \mathbf{m}_{\text{pert}}$. If not, then reapply \mathbf{m}_{cur} and repeat the above steps.

[12] This algorithm will sample the posterior probability density

$$\sigma(\mathbf{m}) = \eta \rho(\mathbf{m}) L(\mathbf{m}), \quad (2)$$

η being a normalization constant, asymptotically. ‘‘Asymptotically,’’ in this case, implies that the statistical correlation between samples taken at times separated by n iterations will converge toward zero as n goes to infinity. The advantage of using the above scheme to sample the posterior probability density $\sigma(\mathbf{m})$ is that sampling is conferred to those parts of the model space where model parameters consistent with data and prior information exist.

2.2. Analysis of the Posterior Distribution

[13] Having designated the posterior probability distribution $\sigma(\mathbf{m})$ as the solution to our inverse problem, our main concern is now the analysis of this distribution. However, owing to its complex shape (it might be multimodal, contain infinite variances, etc.), it is not possible to directly access information from it. Instead, the information of most use is obtained by investigating the probability \mathcal{P} that a certain feature resides at a given depth or depth range (corresponding to the model parameters being contained within a given subset of the model space), which can be calculated from

$$\mathcal{P}(\mathbf{m} \in \Lambda) = \frac{\int_{\Lambda} \sigma(\mathbf{m}) d\mathbf{m}}{\int_{\Omega} \sigma(\mathbf{m}) d\mathbf{m}}, \quad (3)$$

where Λ denotes a subset of the model space Ω and the denominator is clearly identified as a normalization factor. This can easily be extended to the calculation of means and covariances.

[14] We could also adopt the point of view taken by *Mosegaard* [1998], who states that the main aim of resolution analysis is to aid in having to choose between differing interpretations of a given data set. In line here-with our principal purpose in analyzing the sampled models will basically be to conjure up queries addressing correlations between several model parameters. The one question that we intend to investigate using the sampled models concerns the lunar crust and could be formulated as follows:

- How likely is it, on the basis of the seismic data and their uncertainties as well as prior information, that the Moon has a discontinuity (suitably defined) in a certain depth range, marking for example the crust-mantle interface?

However, answering questions like these involves having to evaluate resolution measures of the form [*Mosegaard*, 1998]

$$\mathcal{R}(\Lambda, f) = \int_{\Lambda} f(\mathbf{m}) \sigma(\mathbf{m}) d\mathbf{m}, \quad (4)$$

where $f(\mathbf{m})$ is a given function of the model parameters \mathbf{m} and Λ is, as in (3), an event or subset of the model space Ω containing the models of current interest. The similarity between (3) and (4) is obvious.

[15] It is clear from the above discussion that in the case of the general inverse problem, (3) and (4) are inaccessible to analytical evaluation, since we do not have an analytical expression for $\sigma(\mathbf{m})$. However, the problem can be solved using the MCMC algorithm, as mentioned before, which samples a large collection of models, $\mathbf{m}_1, \mathbf{m}_2, \dots, \mathbf{m}_n$, from $\sigma(\mathbf{m})$, whereby the resolution measure, given by (4), in turn can be approximated by the following simple average [Mosegaard, 1998]:

$$\mathcal{R}(\Lambda, f) \approx \frac{1}{N} \sum_{\{n|\mathbf{m}_n \in \Lambda\}} f(\mathbf{m}_n), \quad (5)$$

where N normalizes the resolution measure. For the number of samples $N \rightarrow \infty$ the equality of (4) and (5) corresponds to the important ergodicity property used in Monte Carlo integration.

[16] However, we shall avail ourselves of a slightly different approach which is known as hypothesis testing. While the two analyses basically amount to the same, since a hypothesis corresponds to a question, the advantage in hypothesis testing lies in the fact that we are able to compare any two hypotheses against each other.

[17] The classical problem of hypothesis testing concerns itself with making a choice among different hypotheses, $\mathcal{H}_1, \mathcal{H}_2, \dots, \mathcal{H}_n$, on the basis of some observed data \mathbf{d} . Or formulated alternatively, we might be in the position of having to decide how the prior probability $\mathcal{P}(\mathcal{H})$, concerning any hypothesis \mathcal{H} , has been amended by taking into account observed data \mathbf{d} (henceforth abbreviated data); that is, we are interested in evaluating $\mathcal{P}(\mathcal{H}|\mathbf{d})$, which is usually termed the posterior probability distribution. The mathematical connection linking the prior and the posterior is given by the Bayes theorem

$$\mathcal{P}(\mathcal{H}|\mathbf{d}) = \eta \mathcal{P}(\mathbf{d}|\mathcal{H}) \mathcal{P}(\mathcal{H}), \quad (6)$$

where η is a normalization constant. Extending the analysis in order to compare any two hypotheses, we arrive at the Bayes factor, whose definition has been ascribed to Turing [e.g., Good, 1988].

2.2.1. Definition (Bayes factor)

[18] Given two hypotheses $\mathcal{H}_i, \mathcal{H}_j$ corresponding to different areas of the model space Ω , for data \mathbf{d} , the Bayes factor \mathcal{B}_{ij} in favor of \mathcal{H}_i (and against \mathcal{H}_j) is given by the posterior to prior odds ratio.

$$\mathcal{B}_{ij}(\mathbf{d}) = \frac{\mathcal{P}(\mathbf{d}|\mathcal{H}_i)}{\mathcal{P}(\mathbf{d}|\mathcal{H}_j)} = \frac{\mathcal{P}(\mathcal{H}_i|\mathbf{d})/\mathcal{P}(\mathcal{H}_j|\mathbf{d})}{\mathcal{P}(\mathcal{H}_i)/\mathcal{P}(\mathcal{H}_j)}, \quad (7)$$

$\mathcal{P}(\mathbf{d}|\mathcal{H}_i)$ being the probability distribution, usually termed the likelihood (see below). Thus the Bayes factor provides a measure of whether the data \mathbf{d} have increased or decreased the odds on \mathcal{H}_i relative to \mathcal{H}_j . Accordingly, if $\mathcal{B}_{ij}(\mathbf{d}) > 1$, \mathcal{H}_i is now more relatively plausible than \mathcal{H}_j in the light of \mathbf{d} ; on the other hand, if $\mathcal{B}_{ij}(\mathbf{d}) < 1$, it signifies that \mathcal{H}_j has now increased in relative plausibility [Bernardo and Smith, 1994]. It is clear that apart from how one defines the crust-mantle transition or any other hypothesis for that matter, this form of analysis is straightforward as seen from the

Bayesian viewpoint. This can be appreciated by rewriting the above equation as

$$\frac{\mathcal{P}(\mathcal{H}_i|\mathbf{d})}{\mathcal{P}(\mathcal{H}_j|\mathbf{d})} = \frac{\mathcal{P}(\mathbf{d}|\mathcal{H}_i)}{\mathcal{P}(\mathbf{d}|\mathcal{H}_j)} \times \frac{\mathcal{P}(\mathcal{H}_i)}{\mathcal{P}(\mathcal{H}_j)}, \quad (8)$$

that is, the ratio of posterior odds equals the integrated likelihood ratio times the ratio of prior odds. This equation shows explicitly how the ratio of integrated likelihoods plays the key part in providing the mechanism by which data transform relative prior beliefs into relative posterior beliefs. If we compare this to the form of the posterior probability distribution, equation (2), which we used in sampling solutions to our inverse problem, the connection is complete, in that (2) states in an analogous manner the role of the likelihood function $L(\mathbf{m})$, which contains information on data \mathbf{d} and on the physical theories linking data and model parameters, in changing the prior probability distribution on model parameters $\rho(\mathbf{m})$ into the posterior $\sigma(\mathbf{m})$ [Mosegaard and Tarantola, 1995; Mosegaard, 1998].

3. Analysis: Application of the MCMC Algorithm to the Apollo Lunar Seismic Data

3.1. Forward Problem

[19] As mentioned, the forward problem consists of calculating data, that is, travel times, given a set of model parameters, that is, a model of the subsurface velocity structure, among other things. In our model of the Moon the assumption of radial symmetry is made. The Moon is partitioned into 56 shells of variable size, with each layer being characterized by its physical extent and material parameters in the form of the velocity. To each shell is assigned a piecewise linearly varying P and S wave velocity of the form $v(r) = v_o + k \cdot r$, which is continuous at layer boundaries. In order to accommodate ray theory, certain limits are placed on the velocity gradients, resulting in a smoothness constraint limiting vertical resolution to roughly 5 km.

[20] Since the Moon is neither spherically symmetric nor geologically homogeneous, an additional asset was introduced to facilitate a more realistic modeling of the lunar interior. Our model of the Moon was stripped of a surficial layer, 1 km thick, which is known to be of very low velocity [Kovach and Watkins, 1973]. Because of the extreme velocity gradients encountered by the rays when impinging this layer, they will be almost vertically incident upon reaching the surface. The travel time for a ray in this layer will therefore to a good approximation be given by $T = 1 \text{ km}/v_{\text{surface}}$. So, instead of ray tracing in a sphere with a radius of 1738 km, we traced in one with a radius of 1737 km and added for every station and surface source a time correction to the travel time. Starting off with a surface velocity of, say, 0.5 km s^{-1} [Nakamura et al., 1982] means that we have to add a total time correction of 4 s to the travel time of a ray emanating from a surface source and 2 s for one originating within the Moon. Leaving the time corrections variable, this method has the added advantage of taking localized properties of the surficial material beneath each station into account. For consistency, all time corrections have been correlated in such a way that if a particular station has been assigned a given initial correction, this

same correction will be added to the travel times for the rays emitted from all sources and traveling to this particular receiver. In the same way, the corrections are correlated for the impacts; that is, all rays emanating from a particular impact have the same correction added to their total travel time.

3.2. Inverse Problem

[21] Three parameters as discussed in the previous section are used to describe our physical model of the Moon which included the position of the layer boundaries, r_i , their velocities, v_i , and the time corrections, t_k , in the surficial layer. In choosing how to parameterize our physical system, however, it has to be noted that the choice of which parameterization to employ is an ambiguous matter, in the sense that it is not unique [Tarantola and Valette, 1982]. So, instead of examining a number of different parameterizations, such as the velocity or the slowness, we shall avail ourselves of another approach, this being the invariance argument. As the name suggests, invariance implies the procurement of commensurate distributions upon transformation from any one invariant measure. The “log-velocity” possesses exactly this property [Tarantola, 1987]; that is, given a uniform distribution in $\log(v/v_o)$, for example, we will achieve a similar distribution whether it be the velocity or the slowness we are transforming to. We shall therefore adopt $\log(v/v_o)$ as the parameterization throughout this study. (In the following we shall continue to use v_i to describe the velocity parameters, but it is tacitly assumed that we actually mean the log-velocity. Throughout this study, $v_o = 1.0 \text{ km s}^{-1}$.)

[22] To fully characterize our physical system, we also need to incorporate parameters which describe the meteoroid impact as well as the shallow and deep moonquake locations, thus resulting in the addition of three other parameters in the delineation of our model. These parameters are the depth coordinate of every moonquake and the selenographical position of the epicenter, that is, longitude and latitude, which we shall label s_j and the latter two θ_j and ϕ_j , respectively. Our model is thus ultimately given by $\mathbf{m} = \{r_i, v_i, t_k, s_j, \theta_j, \phi_j\}$.

[23] In commencing the MCMC algorithm we started out by assuming some initial model \mathbf{m} , henceforth \mathbf{m}_{cur} , comprising a set of values. Now, by perturbing one of these parameters, that is, by either changing the position of a boundary layer, changing the value of the velocity at a given layer boundary, assigning a new time correction for either a source or a station, or amending the hypocentral or epicentral coordinates depending on the particular event, we obtain a new model. The random walker is then set out to sample the parameter space according to the prior information and using a set of random rules whose efficiency has been optimized through several numerical experiments. Let us outline these random rules and the prior information. For every iteration it is first decided which parameter is to be perturbed next. The decision is made whether to change a parameter pertaining to the lunar interior or one regarding the location of a seismic event, which are equally probable. Performing a velocity perturbation has the same probability (0.5) as performing a boundary layer perturbation. When it comes to perturbing the time corrections, these are set to be

amended every fifth iteration. When perturbing either the velocity or the position of a boundary, the layer is selected uniformly at random and so is the value of the parameter to be changed. It should be kept in mind that perturbations concerning the placement of boundaries have in a certain sense been restricted in order to accommodate ray theory. Every layer to be perturbed can be assigned any value, except in a range of 5 km within the layer immediately above or below; that is, any given layer i can assume a value within $r_{i-1} + 5 \text{ km} \leq r_i \leq r_{i+1} - 5 \text{ km}$. Concerning the time amendments, whether it is a source location or the location of a station, which is to be perturbed will be chosen uniformly at random, as are their values. Let us note that we have made the assumption of a uniform distribution in the “log-time” domain as in the case of the velocities, the reason being that the time is directly related to the velocities in the top layer. Concerning the hypocenter coordinates, the source depths are uniformly distributed from the surface to the center of the Moon, and the epicentral coordinates are likewise assumed to be uniformly distributed, in this case across the lunar surface. As regards the sampling of S wave velocities, it has to be remarked that in order to make the inversion tractable the problem was actually divided into two stages. The first entailed the inversion of the P wave arrival times, while the second considered the inversion of the S wave arrivals. Now, it is clear that the two parameters, v_p and v_s , are not independent if described in terms of the elastic moduli, ρ , κ , and μ , these being the density, bulk, and shear modulus, respectively. It could therefore be argued that this division of the problem might result in physically unrealizable models (for further discussion, see section 8). However, being aware of the connection between the two parameters, we did impose certain constraints on the sampling of S wave velocities, corresponding to the addition of prior information, by introducing a Gaussian distribution centered on the average $v_p/\sqrt{3}$ as obtained from the first inversion and with a standard deviation given by $\sigma_{v_p}/\sqrt{3}$. This prior information thus signifies that S wave velocities lying far from this distribution are less likely sampled.

[24] This body of information, then, serves as prior knowledge, in the sense that the random walker will be sampling the model space with a probability density describing exactly this information.

[25] At this point the reader might wonder about what happened to the estimation of parameters pertaining to the origin time of events, since these are also unknowns when it comes to event localization. These are also determined in this study; however, instead of including these in the model parameter vector \mathbf{m} above as another set of parameters to be determined, we adopted a slightly different approach.

[26] Generally, we have the following relation concerning the arrival times:

$$\mathbf{d}_a = \mathbf{m}_o + g(\mathbf{m}),$$

where \mathbf{d}_a is the vector of arrival times, \mathbf{m}_o is a model parameter vector of origin times corresponding to individual arrivals, and $g(\mathbf{m})$ is our set of calculated travel times. Now, for the present purposes we made the assumption that the a

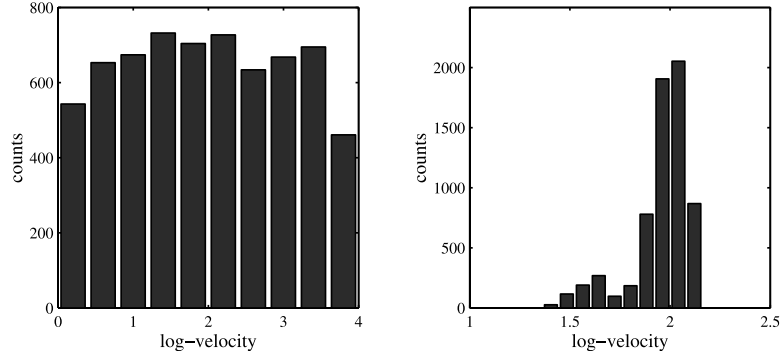


Figure 2. Prior and posterior marginal $\log(v/v_0)$ parameter distribution. Note how the prior is changed from sampling a uniform distribution by taking data into account so as to sample the posterior.

priori information concerning the origin time model parameters was normally distributed with mean values as determined by the Nakamura model [Nakamura *et al.*, 1976; Nakamura, 1983] and an a priori uncertainty, typically of a few seconds. This sets the stage for the following formal definition of new data:

$$\mathbf{d} = \mathbf{d}_a - \mathbf{m}_0.$$

To these were assigned a new standard deviation as the sum of the standard deviation of arrival times and the a priori dispersion on origin time model parameters, thus defining a composite uncertainty (to be dealt with later on).

[27] The reason that we ended up choosing what might be labeled rather loose prior knowledge on model parameters, by assuming that they were all uniformly distributed within a large range, was our main aim of wanting to investigate the full model variability inherent in the lunar seismic data.

[28] Having designated prior information, let us continue the MCMC algorithm and assume that the random walker wishes to sample the velocity at some depth. He currently resides at v_i and from there takes a step in some direction to a point v_i^{new} . Now, we do not restrict the random walker to wander within some specified range, but shall rather let him drift around the model space from zero to infinity, although every jump is in a sense constrained so as to assure that the random walker can take only small steps. Sampling the a priori distribution by this method basically corresponds to the process of diffusion, since the random walker can essentially, if given steps enough, explore the whole model space. Mathematically, we would write this condition as $v_i^{\text{new}} = v_i + \xi \cdot (2 \cdot \alpha - 1)$, where α denotes a uniformly distributed number in the interval $[0,1]$ and ξ is a constant, typically of the order of 0.2 km s^{-1} in this study.

[29] After having obtained a new model, \mathbf{m}_{pert} , our next step is to gauge the posterior distribution. This we do by first calculating a new set of arrival times using $g(\mathbf{m}_{\text{pert}})$ and then comparing these to the observed data in order to obtain the misfit function. This misfit is of importance, since we shall by the Metropolis rule use it to either accept or reject the new model, with probability $P_{\text{acc}} = \min\{1, L(\mathbf{m}_{\text{pert}})/L(\mathbf{m}_{\text{cur}})\}$. In words, we could state this by saying that the random walker essentially updates his knowledge of the current state of affairs, as prescribed by Bayes' theorem, by merging together prior information with data and theory. The revision of the prior into the posterior is illustrated in

Figure 2, which shows how a prior distribution for a given velocity parameter is being changed as data are taken into account. Moreover, we are led to the conclusion that the level of confinement is a direct measure of the knowledge we possess about the system; that is, the more the random walker is constrained when sampling the posterior distribution in the model space, the greater is our degree of knowledge.

[30] Now, the strategy set forth here of perturbing only one parameter at a time is sure to preserve most of the characteristics of the current model which may have resulted in a good data fit. While being efficient, in the sense that we are interested in sampling models with a good data fit, this strategy will result in a sequence of samples, that is, models, that tend to be correlated. This is somewhat unfortunate, since analyses of error and resolution require a collection of statistically independent models from the posterior distribution. The way to proceed is to choose fewer samples from the set of accepted models in such a way that they constitute a set of independent models. This can be done by introducing an elapse time (number of iterations) between retention of samples which was found by analyzing the fluctuations of the likelihood function as the algorithm proceeded to be 100. Inspection of the autocorrelation function for these fluctuations showed that accepted models separated by 100 iterations were unlikely to be correlated. On the other hand, the limited number of iterations performed may not be sufficient to allow the algorithm to visit enough extrema in the model space. To circumvent this problem, we chose to commence at different places in the model space for every 10,000 iterations performed by the algorithm. This was done by restarting the MCMC algorithm with different initial values at those intervals. This procedure should guarantee that the samples are less correlated as well as leading to a better coverage of the probability distribution. (It should be noted that although this method is more efficient in detecting most of the extrema, its downside is the fact that it is biased toward an approximately equal coverage of them all. The global extremum, corresponding to the most likely solution, might end up being sampled on a par with secondary extrema, thereby conferring them with equal weight.) To ensure that the posterior probability density was adequately sampled in our analysis, we monitored the time series of all output parameters from the algorithm to verify that these were indeed stationary over the many iterations performed.

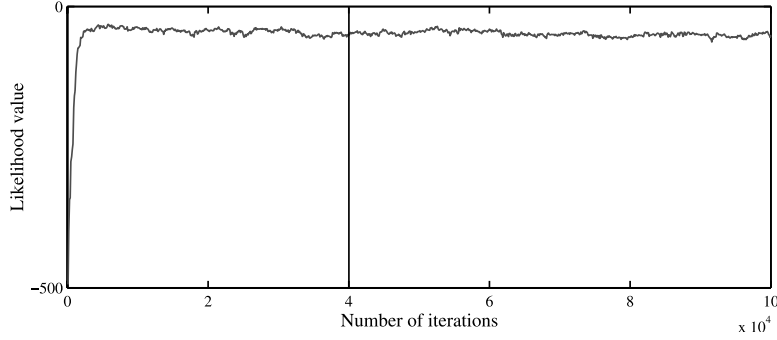


Figure 3. Convergence of the MCMC algorithm. The value of the likelihood function is a rough measure of how well calculated data fit the observables. As convergence has been reached, denoted by the vertical line, sampling of the posterior distribution is initiated.

In addition to showing the convergence (The issue of convergence presents a so far unresolved difficulty in practice, since it is not easy to decide how many samples are actually enough to constitute a good representation of the posterior probability distribution.) of the MCMC algorithm, Figure 3 also depicts an example of how it was decided when to initiate sampling from the distribution, by observing the values of the likelihood function from the time of commencement of the algorithm and as it proceeds. It is customary to start keeping samples only after the likelihood function has stabilized around some value. (It is clear that if our starting model is far removed from any extrema in the model space the longer time it may take, in terms of the sequence of updatings, before the algorithm can actually start to sample the distribution near the extrema where most of the contribution to the posterior distribution ought to come from, rendering the issue of when to start sampling rather important.)

[31] Returning to the sampling algorithm, the next step would be to perturb another model parameter, as explained above, and accept this model with the probability $P_{\text{acc}} = \min\{1, L(\mathbf{m}_{\text{pert}})/L(\mathbf{m}_{\text{cur}})\}$. Continuing along this line, we would assemble a suite of models which are distributed in accordance with the posterior distribution. The models thus gathered constitute our main output.

[32] Let us now turn to a point which is of importance when dealing with data suspected of containing outliers. The introduction of an outlier(s) is most probably the result of an erroneously read arrival time, and given the ingrained complexity of the lunar seismograms, as commented earlier, it only seems too natural to suspect inconsistencies in the readings of at least a couple of seismic phases. Now, the central question is, how do we detect these outliers, if there are any present, and furthermore, does their presence in a data set ensue in any form of distortion of the posterior probability density? According to *Tarantola* [1987], an outlier will have a proclivity to “translate” the posterior probability density if we are employing the l_2 norm, that is, if we are assuming independent, identically distributed Gaussian uncertainties, which is what we have done until now, the amount of “translation” necessarily depending on how much the outlier(s) is displaced from the rest of the data points. If a given data set is suspected of harboring an outlier(s), *Tarantola* [1987] advocates the use of the l_1 norm instead, since a contortion of the posterior probability

distribution is less prevalent in this case. The l_1 norm assumes that the errors can be modeled using an exponential probability density; that is, instead of having the misfit function given by $S(\mathbf{m}) = \frac{1}{2} \sum_{i=1}^N (g^i(\mathbf{m}) - d_{\text{obs}}^i)^2 / \sigma_i^2$, we shall assume the following form for the misfit function: $S(\mathbf{m}) = \sum_{i=1}^N |g^i(\mathbf{m}) - d_{\text{obs}}^i| / \sigma_i$, where σ_i denotes the uncertainty on the i th arrival time reading. In order to investigate as to whether there are any outliers present in our data set and additionally to check the relative robustness of the Gaussian and exponential hypotheses, we undertook an inversion of a more uncertain data set, namely, the meteoroid impacts in combination with the shallow moonquakes. We included in the inversion the artificial impacts, this being mainly for technical reasons. The reason that we have been relying on the l_2 norm hitherto is the fact that the artificial impacts constitute a more reliable data set, since origin times and impact locations are known parameters (except for the 16 SIV-B impact, for which origin time and location are not known parameters because of the loss of tracking of this spacecraft prior to impact [*Latham et al.*, 1972]). Figure 3 also shows how it can be used as an indicator of the presence of outliers in the data set. An estimate of the likelihood value when convergence has been reached using the l_1 norm is roughly given by “minus the number of values in the data vector.” If the MCMC algorithm converges on a value differing from this, it is attributed to the presence of conflicting data points in the observed data set. In order to identify outliers, we generated a set of samples of the order of 10^4 . The data, that is, the arrival times as calculated by each iterated model, were checked minutely, and 20 of the lunar seismic arrivals (not to be confused with events) were identified as outliers. The rejected outliers deviated significantly from the expected values by more than 4 and up to 15 standard deviations of data noise. Figures 4a and 4b depict actual histograms of calculated arrival time differences for a meteoroid impact and a shallow moonquake as registered at stations 15 and 16, respectively, which deviated by more than 4 standard deviations. The outliers were subsequently removed and the process was initiated all over again until the misfit attained the appropriate value. It should be noted that this method has the added benefit of leading to a reevaluation of the uncertainty on the individual arrival times, which amount to 1 s for the artificial impacts, 4–26 s for the

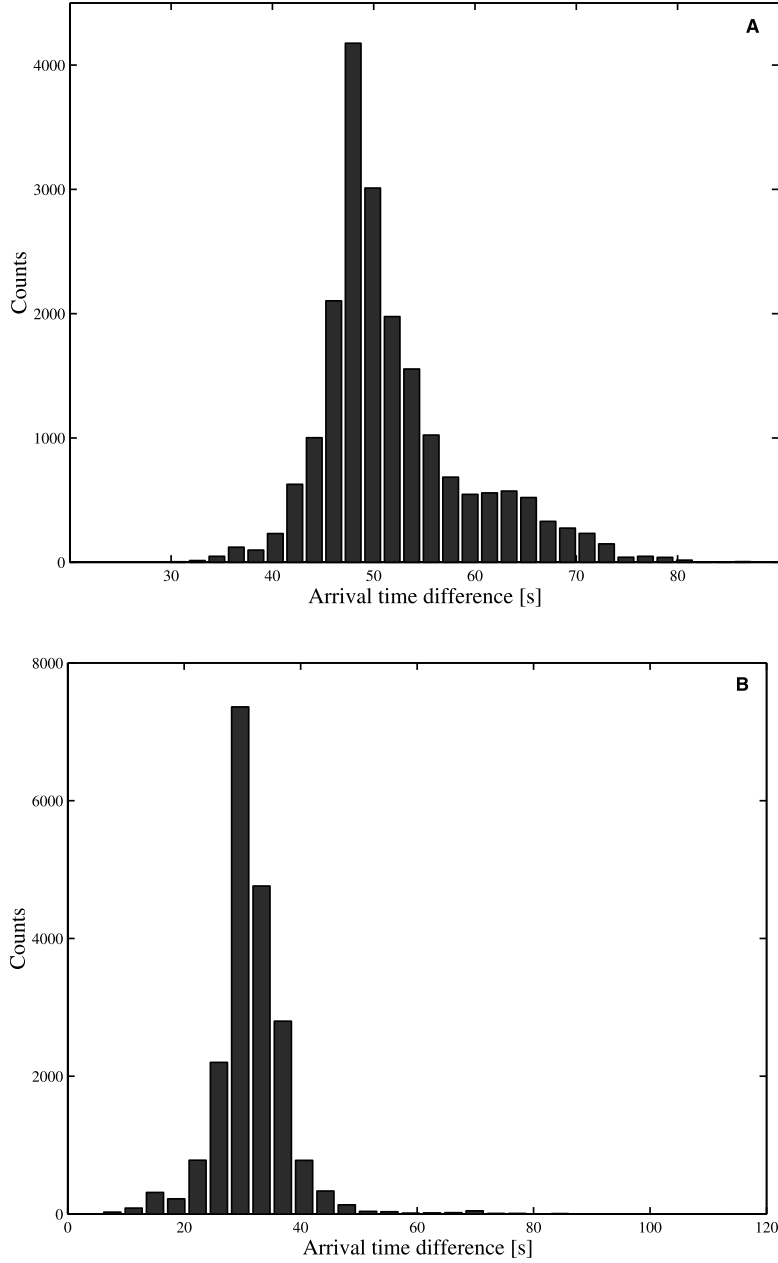


Figure 4. Identification of outliers: (a) the arrival time difference between calculated and observed P wave arrival for meteoroid impact on day 349 1974 as recorded by station 15 and (b) the case of a shallow moonquake recorded at station 16 on day 4 1976. The uncertainties on these two readings were assessed to be 7 and 5 s, respectively.

shallow moonquakes and the meteoroid impacts, and 4–7 s for the deep moonquakes. In order to safeguard the results from further inconsistencies in the data set, we simply chose to use the l_1 norm throughout this study. This method of elucidating outliers in a data set corresponds to the statistical technique known from robust M type estimation [Barnett and Lewis, 1984; Hampel et al., 1986].

[33] To recapitulate, we started off by randomly generating models which were distributed according to the posterior. In order to enhance the coverage of the model space by the random walker, we perturbed him randomly after a fixed number of iterations. This corresponds to the case in which

the random walker, after having sampled some local or global extremum for a preset amount of steps, suddenly finds himself displaced to another region of the model space around which he would steadily peregrinate in the usual manner. This subsequent wandering about leads to the equilibration at yet another extremum, thereby sampling models belonging to a specific class. From Figure 5 we clearly see that the models generated by sampling this particular extremum in the model space are different from the ones produced at other points of the model space, in that they accentuate contrastive properties. On display in Figure 5 are exactly 10 distinct models, and these are each

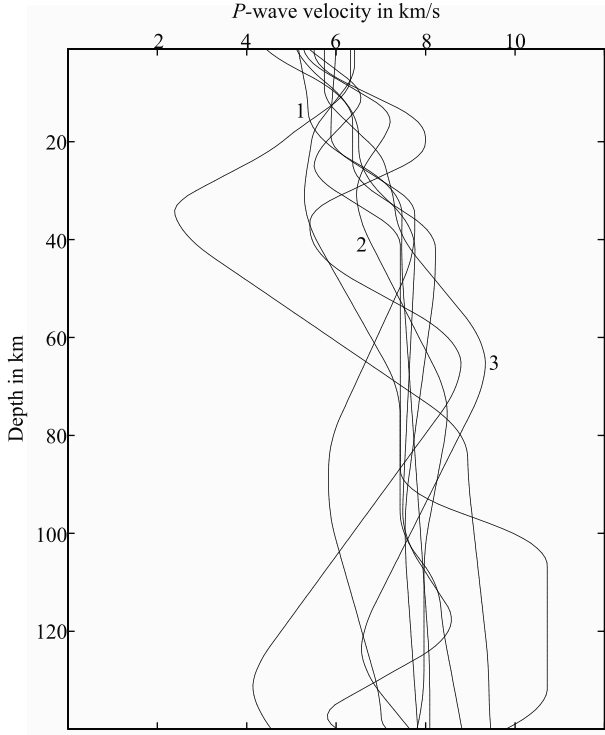


Figure 5. A smoothed version of 10 samples from the posterior distribution, each one obtained by sampling a different extremum in the model space. The models shown here emphasize only the crustal and upper mantle part. The three models numbered 1–3 have been taken as examples of models which individually highlight seemingly different structures but nonetheless are models that all produce a fairly large likelihood value, that is, produce a good fit to the observed data (see Table 1).

representative of a class of models obtained by gathering samples from 10 distinct extrema. Whether 10 reshufflings of the random walker actually result in 10 distinctive families of models is something we clearly cannot be aware of from the outset, and indeed, had we ended up with only three different classes of models, this would have been a revelation in itself, disclosing information about the degree of nonlinearity of the problem. The models juxtaposed in Figure 5 are also displayed to highlight the ingrained ambiguity contained in the data. Although the classes of models presented here are characterized by a fairly large likelihood value, that is, they produce arrival times which are in close agreement with the observed ones, being well within the specified error bounds, they clearly show divergencies among themselves. Figure 5 simply emphasizes that models with differing attributes are accountable for the same data, which is clearly evidenced by Table 1. All three models are seen to be capable of fitting the observational data within error bounds (± 1 s). This excellent agreement between observed and calculated data reinforces the method employed here of elucidating the lunar interior from a great number of models rather than just adopting the maximum likelihood model and inferring the velocity structure straight from this.

4. Analysis: Estimating the Crustal Thickness Using Bayesian Hypothesis Testing

[34] Having presented the general theory, we are now in the position to employ the Bayes factor in order to specifically analyze the outcome as to whatever feature we might be interested in. From (5) it is apparent that the number of times a model has been sampled is proportional to its posterior probability. The Bayes factor is thus easily evaluated as the ratio of the number of samples from the posterior distribution to the ratio of the number of samples from the prior distribution for any two hypotheses.

[35] As noted in section 1, our primary concern is the depth to the lunar Moho, and we shall here want to distinguish between the following two hypotheses given the seismic data and prior information:

Hypothesis 1 (\mathcal{H}_1): The lunar crustal thickness lies in the range 35–45 km.

Hypothesis 2 (\mathcal{H}_2): The lunar crustal thickness lies in the range 50–70 km.

[36] Of a more technical character is the definition of the crust-mantle interface. A model subject to the following two conditions is defined as accommodating a discontinuity:

1. Velocity gradients should reach $0.1 \text{ km s}^{-1} \text{ km}^{-1}$.
2. The velocity beneath the designated thickness of the crust should attain a value of at least 7.6 km s^{-1} .

[37] This interpretation of the lunar Moho has been constructed so as to agree with the characteristics of all three models discussed in section 1, \mathcal{H}_1 and \mathcal{H}_2 corresponding to the crustal depths as proposed by (1) Khan *et al.* [2000] and (2) Toksöz *et al.* [1974] and Nakamura [1983], respectively. This implies searching through the entire set of sampled posterior and prior models and by (5) counting the number of prior and posterior models, respectively, satisfying the above two definitions of a crust-mantle boundary and simply taking their ratio. As a further technical note, it should be mentioned that the MCMC algorithm is designed to sample models with the highest possibly achievable resolution. However, given the well-known trade-off that exists between resolution and uncertainty, high resolution implies a rather large dispersion in model parameters and vice versa. Since we are presently concerned about searching for velocity discontinuities, small uncertainties on model parameters are paramount. This is easily achieved by smoothing the individual models, although of course at the expense of resolution. We have used the median in filtering the models over a range of eight values around a given velocity point,

Table 1. A Comparison of Observations With Calculations^a

Impactor	Station	Observed Travel Times, s	Calculated Travel Times, s		
			Model 1	Model 2	Model 3
13 SIV-B	12	27.3	27.2	27.2	27.4
14 SIV-B	12	34.3	34.5	34.2	34.6
14 LM	12	23.7	23.0	23.2	24.0
14 LM	14	16.5	16.1	16.3	16.0
15 SIV-B	12	53.7	53.9	54.1	53.3
15 SIV-B	14	35.3	36.3	34.9	36.2
15 LM	15	20.7	21.3	19.8	20.1
17 SIV-B	12	54.7	55.4	55.3	55.5
17 SIV-B	14	30.7	30.0	31.2	30.5
17 SIV-B	15	149.7	149.1	150.0	150.3
17 SIV-B	16	121.8	122.4	121.1	122.0

^aThe three models refer to Figure 5.

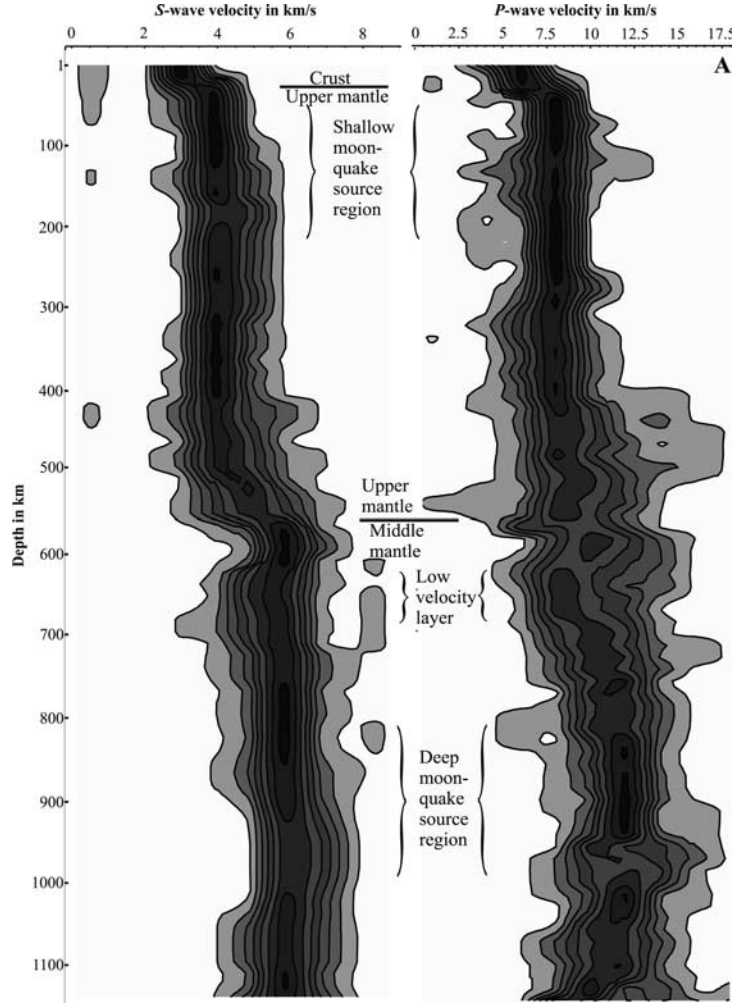


Figure 6. The marginal posterior velocity distributions depicting the velocity structure of the Moon. A total of 50,000 models have been used in constructing the two results. For each kilometer a histogram reflecting the marginal probability distribution of sampled velocities has been set up. By lining up these marginals, the velocity as a function of depth is envisioned as contours directly relating their probability of occurrence. The contour lines define nine equal-sized probability density intervals for the distributions. The uncertainties on the results are in part due to the large uncertainty in arrival time readings. It should be kept in mind that the velocity models are depicted using marginal probability distributions, and as such a model incorporating velocities of maximum probability does not necessarily correspond to the most likely model. See color version of this figure at back of this issue.

since the median preserves any discontinuities rather than smoothing them.

[38] The posterior and prior odds ratios were found to be $\mathcal{P}(\mathcal{H}_1|\mathbf{d})/\mathcal{P}(\mathcal{H}_2|\mathbf{d}) = 2.5$ and $\mathcal{P}(\mathcal{H}_1)/\mathcal{P}(\mathcal{H}_2) = 0.6$, respectively, resulting in a Bayes factor of

$$B_{12} = \frac{\mathcal{P}(\mathcal{H}_1|\mathbf{d})/\mathcal{P}(\mathcal{H}_2|\mathbf{d})}{\mathcal{P}(\mathcal{H}_1)/\mathcal{P}(\mathcal{H}_2)} = \frac{2.5}{0.6} = 4.2,$$

whereby \mathcal{H}_1 is favored to \mathcal{H}_2 given data and prior information.

5. Interpretation of the Results

[39] As emphasized previously, the solution to the Bayesian formulation of general inverse problems results in a

posterior probability distribution and not single estimates, although these can be obtained from the posterior distribution by numerical integration. What this signifies is that we are solely working with solutions in terms of probability distributions, and moreover, given the multidimensional nature of the posterior distribution, it is not accessible for direct display and only marginals can be exhibited. Therefore care must be exercised in interpreting the outcome, since false conclusions and overinterpretations are easily reached.

[40] Now, from Figure 6, depicting the final velocity model, it would seem very natural to assume the model belonging to the region of highest probability as the prevalent one and therefore pick it as the main lunar velocity structure. However, this particular model will usually not correspond to the most likely model (which is typical of strongly nonlinear problems and is clarified further below), since it has to be remembered that the

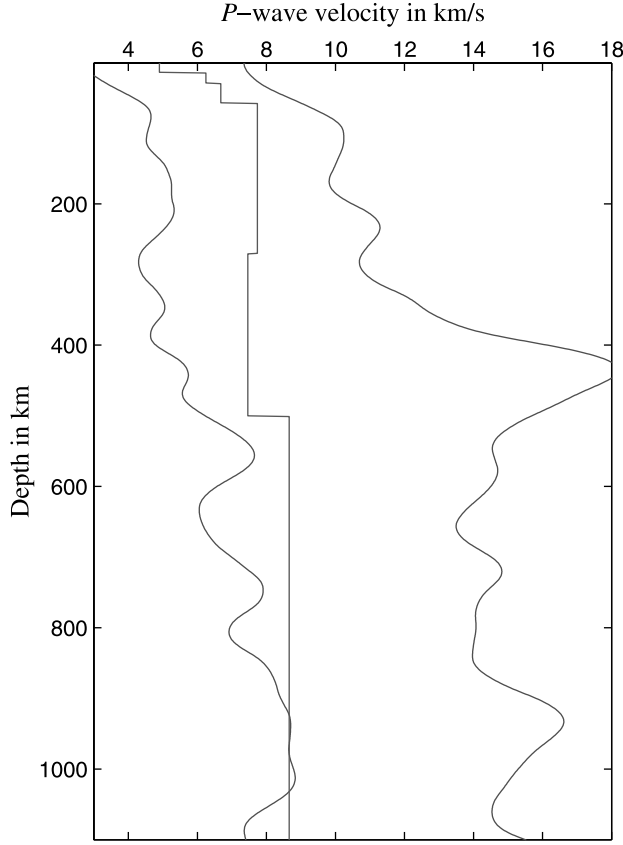


Figure 7. The lunar P wave velocity structure shown using 99% credible intervals. The broad velocity ranges evince the uncertainties involved. The structure in the middle shows the Nakamura velocity model [Nakamura *et al.*, 1982] so as to highlight consistency with earlier models.

velocity models shown here have been displayed using marginal posterior probability distributions, one-dimensional (1-D) marginals, actually. What this means is that if we were to choose the most probable velocity at a given depth, this would correspond to the most likely velocity at this particular depth only, while disregarding the information on all other parameters.

[41] Given that $\sigma(\mathbf{m})$ is a rather complicated entity, it is more convenient and probably also sufficient for general orientation regarding the uncertainty about \mathbf{m} simply to describe regions of given probability under $\sigma(\mathbf{m})$. From this point of view the identification of intervals containing 50, 90, or 99% of the probability under $\sigma(\mathbf{m})$ might actually suffice to give a good idea of the general quantitative messages implicit in $\sigma(\mathbf{m})$, leading us to the calculation of credible intervals. The credible interval is defined as the shortest possible interval containing a given probability, and in Figure 7 we have depicted credible intervals containing 99% of the marginal probability. At first glance these credible intervals appear to be quite broad, which essentially indicates that the absolute value of the velocity in a given layer is not very well determined. As a consequence, the interpretation of the marginals shown in Figure 6 is a subtle issue where the use of qualities such as prudence and restraint are to be advocated. For example, if it happens that at a certain depth or within a certain depth range there might be a large probability for a somewhat high

velocity, this is, because of the aforementioned ill-determined nature of the absolute velocities, by no means to be interpreted as the results stating this particular fact.

[42] This conundrum inherent in nonlinear problems is further illustrated with the following simple example. Figure 8a shows the nonlinear function investigated for the purpose (a third-degree polynomial). Also included is the true model, m_{true} , giving rise to noise-free data d_{true} . Let us imagine that we have performed some sort of physical measurement which results in an observed datum, d_{obs} , which is also shown. Let us also assume, mainly for simplicity, that noise concerning this data point is normally distributed with uncertainty σ . In this particular example, d_{obs} turns out to be displaced one σ from d_{true} (experimental details are included in the caption of Figure 8). Now, the situation is this: we have measured a datum, d_{obs} , and given this observation we would like to infer information about the model m (here one model parameter) by computing its posterior probability density (the prior probability density is assumed constant). What has been said so far is exactly what we have been describing hitherto in dealing with inverse problems. Next, the posterior probability distribution for a set of model parameter values is easily calculated, and Figure 8b shows this probability distribution. Here m_{true} has also been depicted, and it is obvious that the true model is situated far from the model of maximum posterior probability.

[43] This example more than anything else reveals the intricacies hidden in nonlinear problems and explains why the most probable model is not necessarily the one of utmost interest, since it might be far displaced, as indeed it is in the example outlined here (by more than 100%), from the true model, and this in spite of the fact that data

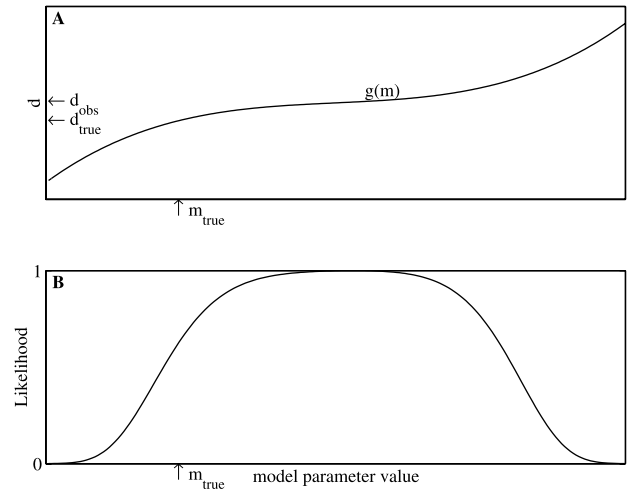


Figure 8. A simple illustration of a nonlinear problem: (a) the nonlinear function and (b) the posterior probability distribution for the model parameter m . Here the true model is displaced by more than 100% from the model with the greatest posterior probability, ultimately showing that care has to be exercised in interpreting the outcome of nonlinear problems. Details of the experiment are as follows: $g(m) = (\alpha m)^3 + \beta m$, where $\alpha = 0.7$ and $\beta = 700$; d_{obs} is assumed to be Gaussian distributed with uncertainty $\sigma = 10^5$ and whose actual value is given by $d_{\text{true}} + \sigma$. The posterior probability distribution is calculated from $L(m) = \exp(-\|d_{\text{obs}} - g(m)\|^2 / 2\sigma^2)$.

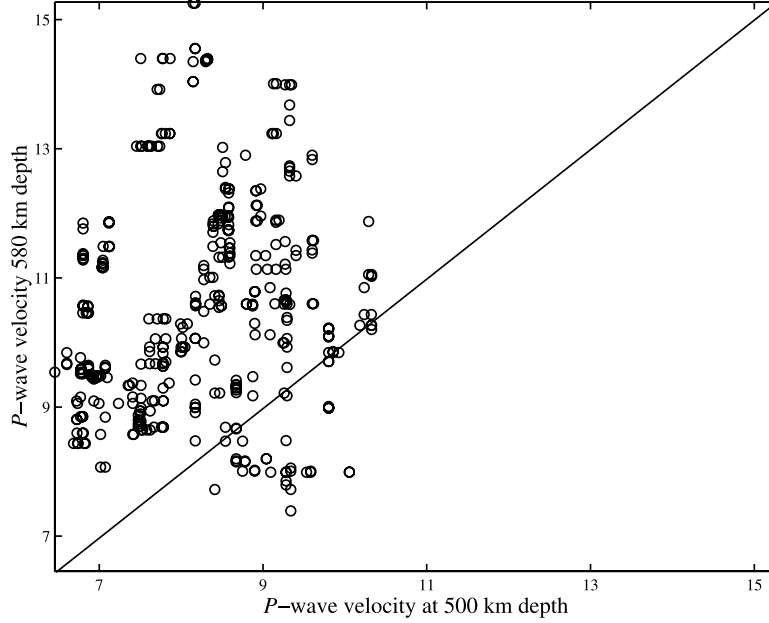


Figure 9. Two-dimensional marginal P wave velocity distribution depicting the correlation between sampled velocity parameters at 500 and 580 km depth. The distribution is clearly seen to be translated toward higher velocities at a depth of 580 km, indicating the existence of a velocity jump. Eighty-eight percent of the probability distribution lies above the equality line.

uncertainty has been modeled using a Gaussian distribution. With this example in mind it is also clear why, in the case of nonlinear problems, the model with the highest posterior probability does not necessarily correspond to the right solution and thus why interpretations using the most probable model can be erroneous.

[44] Returning to the credible intervals, we further superposed the Nakamura model on top of our credible intervals to highlight consistency. Although the Nakamura model at certain ranges lies just barely within ours, it is nonetheless contained within the 99% credible region, which shows that there is no inconsistency between our results and his model.

[45] On the other hand, an item which turns out to be well determined is the existence of a velocity discontinuity. This is easily evidenced using two-dimensional marginals. As an example, we plotted in Figure 9 the marginal probability distribution among two model parameters, the P wave velocity at 500 km and at 580 km depth, respectively, which depicts the correlation that exists between the two model parameters. From this it is apparent that by far the largest probability is displaced toward the model parameter at 580 km depth, clearly indicating that higher velocities at this depth are favored in comparison to those above. This translation of the probability distribution then is evidence of a velocity discontinuity. Whether the purported discontinuity is sharp or gradual is more difficult to assess with this method. It should be noted that this inference was made without any reference as to absolute velocity values.

[46] Of course, we could also display 3-D marginals, depicting the probability distribution among three model parameters at a time. This is exemplified in Figure 10, where we have shown examples of 3-D marginals depicting the sampled hypocentral coordinates of all deep moon-

quakes, giving a good representation of their distribution in space.

6. Results and Discussion

[47] In interpreting the results, we have to be aware of the limitations which have been imposed in part on us by way of the nature of the data and in part by us in our choice of our particular model of the Moon and the inversion. Chief limitations imposed by data are, of course, due to the fact that only four stations were operative, covering only a small area of the front side of the Moon with two stations placed close to each other, raising the question whether they really represent independent degrees of freedom to determine the unknowns. And of course the complex lunar signal characteristics have right from the beginning been a major obstacle, limiting modeling to essentially first arrivals, although attempts at modeling the low-frequency part of the signals have also been undertaken [Loudin, 1979; Khan and Mosegaard, 2001].

[48] In terms of our model of the Moon the necessary assumption of spherical symmetry has to be cited as a limiting factor, and the obtained velocity structure is thus to be viewed as an average over the front side of the Moon, covered by the four seismic stations, with near-surface lateral heterogeneity modeled using local corrections. Although these local corrections are a simple way of dealing with lateral heterogeneity, by being a vertical average over the 1 km surficial low-velocity layer which has been stripped off, they nonetheless contain a clue to the differences among the four stations. However, whether this difference is due to actual differences in velocity, thereby reflecting different geological settings, or simply due to differences in topography among the four stations cannot be assessed by the local corrections. The importance of the local corrections,

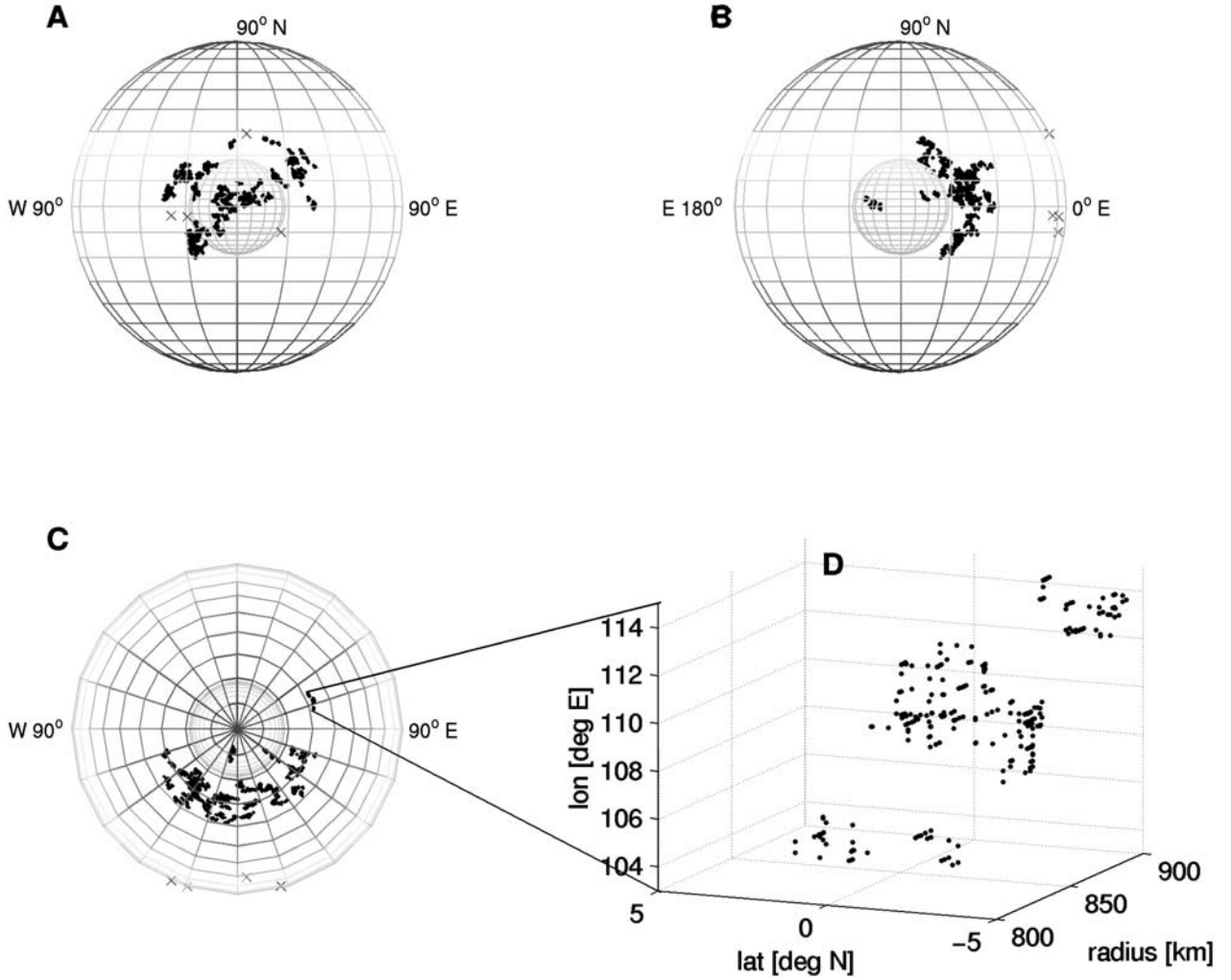


Figure 10. Sampled deep moonquake hypocenter coordinates showing the spatial distribution of quakes in the lunar interior. Figures 10a–10c display all sampled hypocenter coordinates for each deep moonquake source region from a number of different viewpoints. Since the individual source regions contain a cluster of many thousand samples, which are individually not distinguishable, Figure 10d depicts the distribution of sampled hypocenter coordinates for the lone farside moonquake, A_{33} , for enhancement. Figures 10a–10c also contain a central sphere with a radius of 500 km, which was included so as to enhance illustration of the spatial distribution of the moonquakes and is not meant to signify a lunar core. In Figure 10c we are viewing down on the lunar north pole, and the only farside moonquake observed to date, A_{33} , lying just over the eastern limb, is clearly visible. Crosses denote seismic stations. See color version of this figure at back of this issue.

however, becomes obvious when one considers the question as to whether or not stations 12 and 14 can be regarded as being independent sources of information. It is clear that with the large number of parameters that have to be determined the arrival times from these two stations should carry full weight. However, if the arrival time is delayed at one station relative to the other and this happens to be due to local structural differences beneath the two stations, such differences directly affect large systematic errors in structural parameters thus determined, as noted by *Nakamura* [1983]. The significance of the local corrections is thus obvious, since structural differences if extant are dealt with by these local parameters. Figure 11 depicts the sampled local corrections in terms of velocities, showing differences to be present beneath each of the four stations.

[49] From the point of view of the inversion there exists the usual trade-off between resolution and uncertainty. In this study it has to be emphasized that we have chosen the highest possible resolution given the constraint we had to satisfy to accommodate ray theory. The uncertainties are therefore to be viewed in the light of this. Had we, for example, chosen to increase the minimum layer thickness, this would have reduced the uncertainties on the velocities. This is partly the reason why the *Nakamura* model has smaller uncertainties. Accordingly, we have chosen the best possible resolution given ray theory and thus the most conservative estimate of model parameter uncertainties. As regards the MCMC algorithm, it has to be said that in the end we are able to pick only a limited number of samples from the posterior distribution. If the problem is

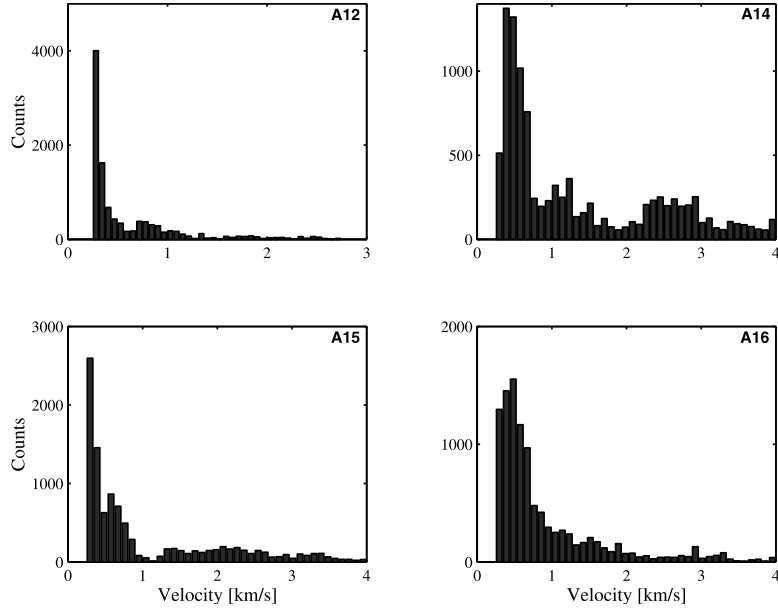


Figure 11. The P wave velocity structure of the uppermost 1 km of the lunar crust, comprising the very low velocity layer beneath each of the four Apollo landing sites. Mean values are $v_{12} = 0.6 \text{ km s}^{-1}$, $v_{14} = 1.3 \text{ km s}^{-1}$, $v_{15} = 1.0 \text{ km s}^{-1}$, and $v_{16} = 0.9 \text{ km s}^{-1}$.

too complex, one might tend to bypass solutions in the model space. This would result in an underestimation of the width of the uncertainty, because of the inadequate coverage of the model space. Stationarity in the likelihood function and model parameters, while sampling, is a necessary condition, but not a sufficient one to ensure that we have actually sampled the posterior distribution in a satisfactory manner. Despite these shortcomings in mathematical details, this method is still a much better tool than deterministic ones (like linearized methods), in that the deterministic methods tend to sample a model which lies close to its start, introducing a bias, since the uncertainty on the obtained model essentially depends on where one starts off.

[50] Another issue concerning this study which merits a comment is the very limited data set, giving in part rise to the error bounds on the results. Earlier studies assumed all arrival time readings equally uncertain. This has the unfortunate consequence of conferring equal weight to all data points. Good as well as bad arrival time readings are thus equally probable, which might lead to inconsistencies. Now, the point of view adopted here of assigning to a given arrival time an a priori uncertainty and the subsequent process of looking for and discarding outliers, resulting in a revision of the posterior uncertainty, should protect the

results from these inconsistencies. On the other hand, it must be said that the procedure of eliminating certain data points on the grounds that they deviate from the assigned arrival time reading \pm the uncertainty is a somewhat crude approach, since the given data point is strictly an event, which, however, owing to the complex nature of the seismogram, is extremely difficult to pick.

[51] In spite of this, inversion of the artificial impacts led to the crustal velocity structure (to a depth of roughly 100 km), inversion of meteoroid impacts in combination with the shallow moonquakes provided information on the lunar upper mantle (depth range roughly 100 to 500 km), and finally, inversion of the deep moonquakes provided information about the middle mantle velocity structure (depth range 500–1100 km). About 5×10^6 models were sampled in all, from which 5×10^4 were used for analysis in this study. The results are shown in Figure 6. Figure 12 shows the marginal distributions of sampled hypocenter coordinates for deep moonquake source A_7 , depicting another way of presenting the full posterior distribution (Figure 10). The hypocentral coordinates for all deep and shallow moonquakes and the epicentral coordinates for the meteoroid impacts are compiled in Tables 2, 3, and 4, respectively.

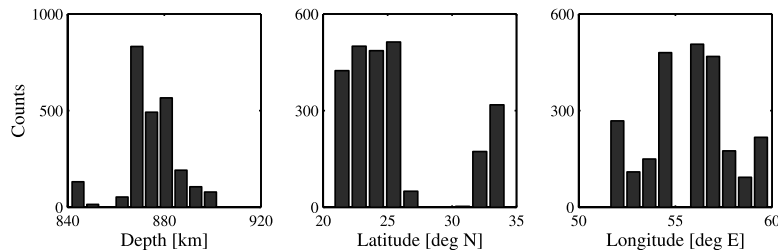


Figure 12. Marginal distributions depicting the sampled source depths and epicentral coordinates in terms of latitude and longitude for deep moonquake source A_7 .

Table 2. Shallow Moonquake Hypocenter Estimates^a

Event		Hypocenter Coordinates		
Year	Day	Depth, km	Latitude, °N	Longitude, °E
1971	107	74.5 ± 21.2	52.5 ± 2.8	26.9 ± 3.3
1971	140	47.5 ± 46.4	37.4 ± 2.1	-19.4 ± 3.7
1971	192	220.8 ± 48.8	49.4 ± 2.5	-27.2 ± 8.7
1972	002	172.5 ± 54.7	56.9 ± 4.2	111.4 ± 5.5
1973	072	48.9 ± 34.9	-81.9 ± 2.9	-130.7 ± 3.4
1973	171	114.8 ± 22.7	-5.4 ± 2.7	-68.1 ± 2.8
1974	192	132.6 ± 40.7	26.1 ± 4.9	85.1 ± 2.3
1975	003	74.8 ± 16.2	27.3 ± 3.0	-102.2 ± 3.5
1975	012	101.9 ± 25.6	63.1 ± 5.5	46.9 ± 3.5
1975	044	129.9 ± 70.3	-16.4 ± 2.3	-24.6 ± 3.1
1975	314	73.2 ± 18.7	-10.6 ± 2.7	58.9 ± 3.2
1976	004	88.0 ± 38.8	44.5 ± 3.2	30.5 ± 3.7
1976	066	74.4 ± 26.9	45.6 ± 3.9	-26.3 ± 4.8
1976	068	119.8 ± 23.9	-18.2 ± 2.3	-10.8 ± 1.6

^a All estimates are mean values, and the values indicated by ± signify one standard deviation.

[52] A detailed discussion of how to interpret the data has been advanced, and there it was made rather obvious that, because of the large credible region encompassed by the models, the absolute velocities were ill-determined. However, velocity discontinuities were found by employing 2-D marginals. In Figure 9 we depicted the correlation between two model parameters at 500 and 580 km depth, respectively, and it was clearly seen that the probability distribution was displaced toward higher velocities at 580 km depth, indicating a transition in this general region. This estimate can be made quantitatively, since the actual probability for there being such a discontinuity can easily be calculated using (5), which states that all we have to do is to count the number of models purporting whatever feature we might be interested in. In the above case the probability is 88%, which speaks for itself. Figure 13 shows the outcome of a similar analysis conducted for the *S* wave models which gave a 99% probability for there being a velocity increase in the same region. While our results (Figure 6) appear to suggest constant velocity zones for the upper mantle, Nakamura *et al.* [1976] inferred a decrease of shear wave velocity starting at roughly 300 km depth associated with a lower *Q* for shear waves, using the decay of shear wave amplitude with distance and the relative arrival times of *P*

Table 3. Meteoroid Impact Epicenter Estimates^a

Event		Epicenter Coordinates	
Year	Day	Latitude, °N	Longitude, °E
1971	143	-6.5 ± 4.4	-18.3 ± 3.2
1971	163	27.8 ± 2.5	-36.7 ± 4.4
1971	293	32.1 ± 2.8	-36.4 ± 2.1
1972	134	-0.2 ± 2.1	-14.6 ± 1.3
1972	199	15.2 ± 12.6	133.6 ± 2.5
1972	213	29.0 ± 1.5	-6.3 ± 1.5
1972	324	65.4 ± 3.3	-21.5 ± 6.2
1974	325	-6.3 ± 2.1	22.8 ± 0.9
1974	349	-10.2 ± 7.7	-12.4 ± 2.1
1975	102	5.4 ± 1.9	33.8 ± 1.7
1975	124	-57.4 ± 11.2	-117.4 ± 3.7
1976	025	-2.5 ± 5.0	-71.4 ± 1.7
1976	319	4.1 ± 6.7	-86.9 ± 5.5
1977	107	-37.9 ± 5.9	-60.3 ± 4.6

^a All estimates are mean values. The values indicated by ± signify one standard deviation.

Table 4. Deep Moonquake Hypocenter Estimates^a

Hypocenter Coordinates			
Source	Depth, km	Latitude, °N	Longitude, °E
A ₁	921 ± 6	-19.9 ± 1.3	-40.9 ± 1.5
A ₅	702 ± 10	17.4 ± 1.5	-39.1 ± 3.4
A ₆	847 ± 8	36.2 ± 2.9	54.2 ± 1.4
A ₇	876 ± 6	22.9 ± 1.4	55.6 ± 2.7
A ₈	942 ± 14	-33.7 ± 2.3	-37.3 ± 2.3
A ₉	992 ± 8	-7.9 ± 1.1	-14.4 ± 3.9
A ₁₁	822 ± 11	6.6 ± 1.2	6.8 ± 2.0
A ₁₄	928 ± 13	-25.2 ± 0.9	-37.6 ± 1.2
A ₁₅	820 ± 24	-0.2 ± 2.5	-3.8 ± 3.4
A ₁₆	1161 ± 29	7.1 ± 2.1	3.7 ± 1.5
A ₁₇	820 ± 37	26.7 ± 1.4	-21.5 ± 1.0
A ₁₈	918 ± 7	22.3 ± 1.6	32.1 ± 1.2
A ₁₉	799 ± 4	17.0 ± 3.8	30.8 ± 3.3
A ₂₀	960 ± 4	25.1 ± 1.3	-34.0 ± 1.6
A ₂₄	985 ± 22	-37.8 ± 1.8	-42.7 ± 2.1
A ₂₅	959 ± 12	37.0 ± 1.0	67.7 ± 2.4
A ₂₇	1056 ± 13	20.9 ± 2.6	21.1 ± 2.0
A ₂₈	1048 ± 8	8.3 ± 1.8	28.3 ± 2.1
A ₃₀	915 ± 17	13.1 ± 1.5	-33.2 ± 1.6
A ₃₃	902 ± 11	3.7 ± 2.1	112.3 ± 2.8
A ₃₄	993 ± 13	6.2 ± 2.4	-7.6 ± 1.3
A ₃₆	1016 ± 9	64.4 ± 2.5	-12.5 ± 1.5
A ₃₉	933 ± 6	-16.9 ± 4.5	-14.6 ± 1.8
A ₄₀	905 ± 23	-0.3 ± 1.4	-10.6 ± 2.9
A ₄₁	801 ± 4	19.7 ± 1.2	-34.6 ± 0.9
A ₄₂	915 ± 9	27.9 ± 3.1	-51.1 ± 3.0
A ₄₄	941 ± 8	61.2 ± 2.4	52.9 ± 0.9
A ₅₀	836 ± 16	11.4 ± 1.3	-52.9 ± 1.7
A ₅₃	932 ± 12	-27.9 ± 3.6	-31.0 ± 2.1
A ₅₄	968 ± 18	8.4 ± 1.8	-64.0 ± 4.7
A ₆₁	801 ± 5	21.8 ± 1.1	43.4 ± 1.1
A ₇₁	945 ± 26	14.3 ± 1.0	-12.4 ± 2.0
A ₈₄	879 ± 18	-13.6 ± 2.8	-31.0 ± 1.6
A ₈₅	821 ± 7	35.9 ± 2.7	68.1 ± 1.7
A ₉₇	948 ± 10	4.5 ± 1.8	12.4 ± 3.6

^a All estimates are mean values and the values indicated by ± signify one standard deviation. The source numbering follows that of Nakamura *et al.*, [1982].

and *S* waves. This decrease in v_s is easily investigated using 2-D marginals. The result is depicted in Figure 14, which clearly shows the probability distribution displaced toward higher velocities at 280 km than at 320 km depth with a value of 95%, although the involved velocity decreases are minor. No significant velocity changes were found for the *P* wave models.

[53] Turning to a different method of analysis of the posterior distribution leads us to our investigation of the much debated lunar crustal thickness using Bayesian hypothesis testing. Our analysis resulted in a Bayes factor of 4.2, which undeniably leads to the conclusion that hypothesis 1 is more plausible than hypothesis 2, thereby favoring the depth to the lunar Moho in the range 35–45 km. As a further aid in arguing for a shallow lunar crust we have displayed the models, satisfying the two hypotheses, where Figure 15 depicts the posterior velocity models satisfying \mathcal{H}_1 and \mathcal{H}_2 , respectively. Of prime interest in Figure 15a is the obviously discernible discontinuity at $\sim 38 \pm 3$ km depth, where the results indicate an increase in velocity from 6.8 ± 0.7 to 7.8 ± 0.6 km s⁻¹. This is slightly thinner than the value of 45 ± 5 km inferred so far for the crustal thickness [Khan *et al.*, 2000], although it should be stressed that these two values do not mutually disagree, since the former has been obtained from a subset of all

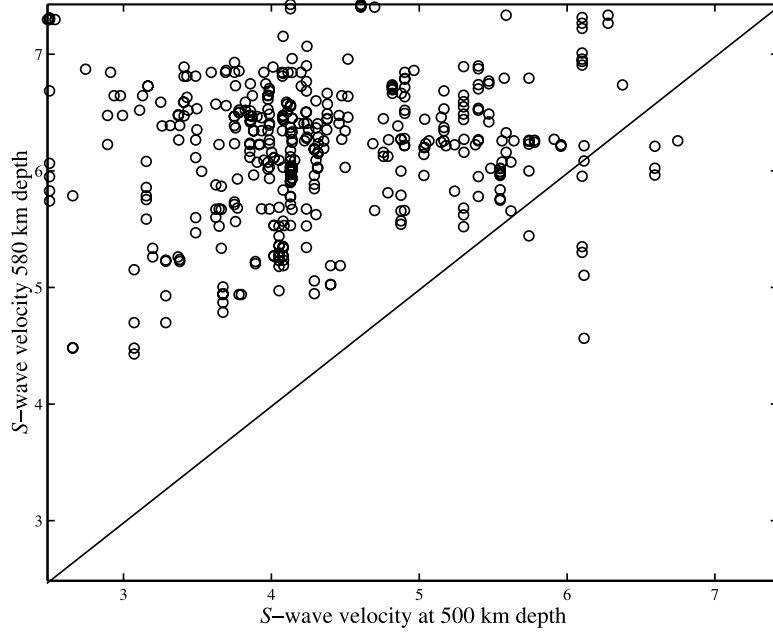


Figure 13. Two-dimensional marginal S wave velocity distribution depicting the correlation between sampled velocity parameters at 500 and 580 km depth. The distribution is clearly seen to be translated toward higher velocities at a depth of 580 km, indicating the existence of a velocity jump. Ninety-nine percent of the probability distribution is situated above the equality line.

sampled models, namely, those models that additionally concur with the constraints imposed by \mathcal{H}_1 . The models which were found to have discontinuities in the depth range 50–70 km are displayed in Figure 15b. One might be tempted to label these models unphysical, but what is actually apparent here is the nonuniqueness inherent in

the data we are dealing with, which leads to a multitude of possible models with clearly quite different implications for the lunar crustal velocity structure. The interpretation of Figure 15b is such that it shows that in order for the MCMC algorithm to generate models incorporating velocity discontinuities at depths around 60 km it seems at the same

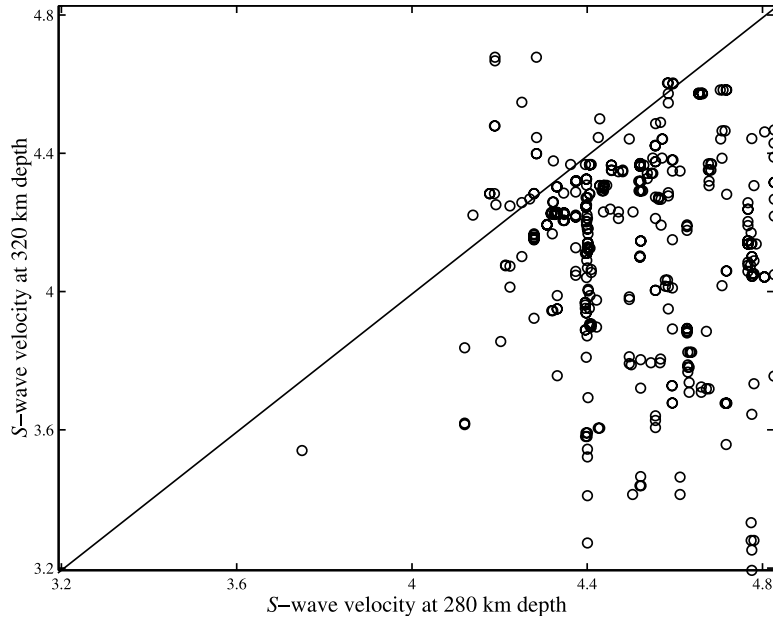


Figure 14. Two-dimensional marginal S wave velocity distribution depicting the correlation between sampled velocity parameters at 280 and 320 km depth. The distribution is clearly seen to be translated toward higher velocities at a depth of 280 km, indicating the existence of a velocity decrease in this region. Ninety-five percent of the probability distribution lies below the equality line.

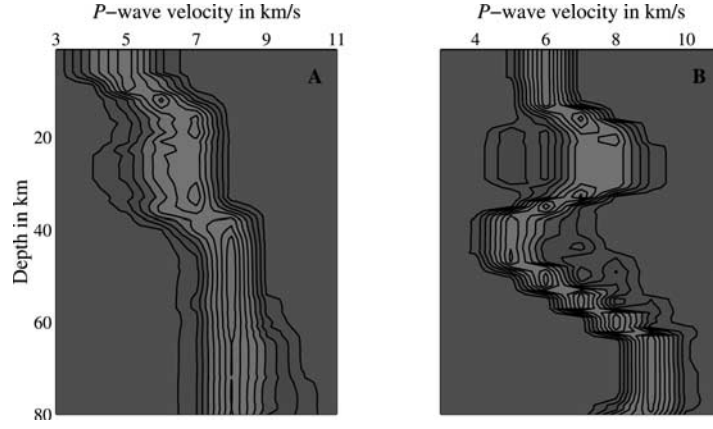
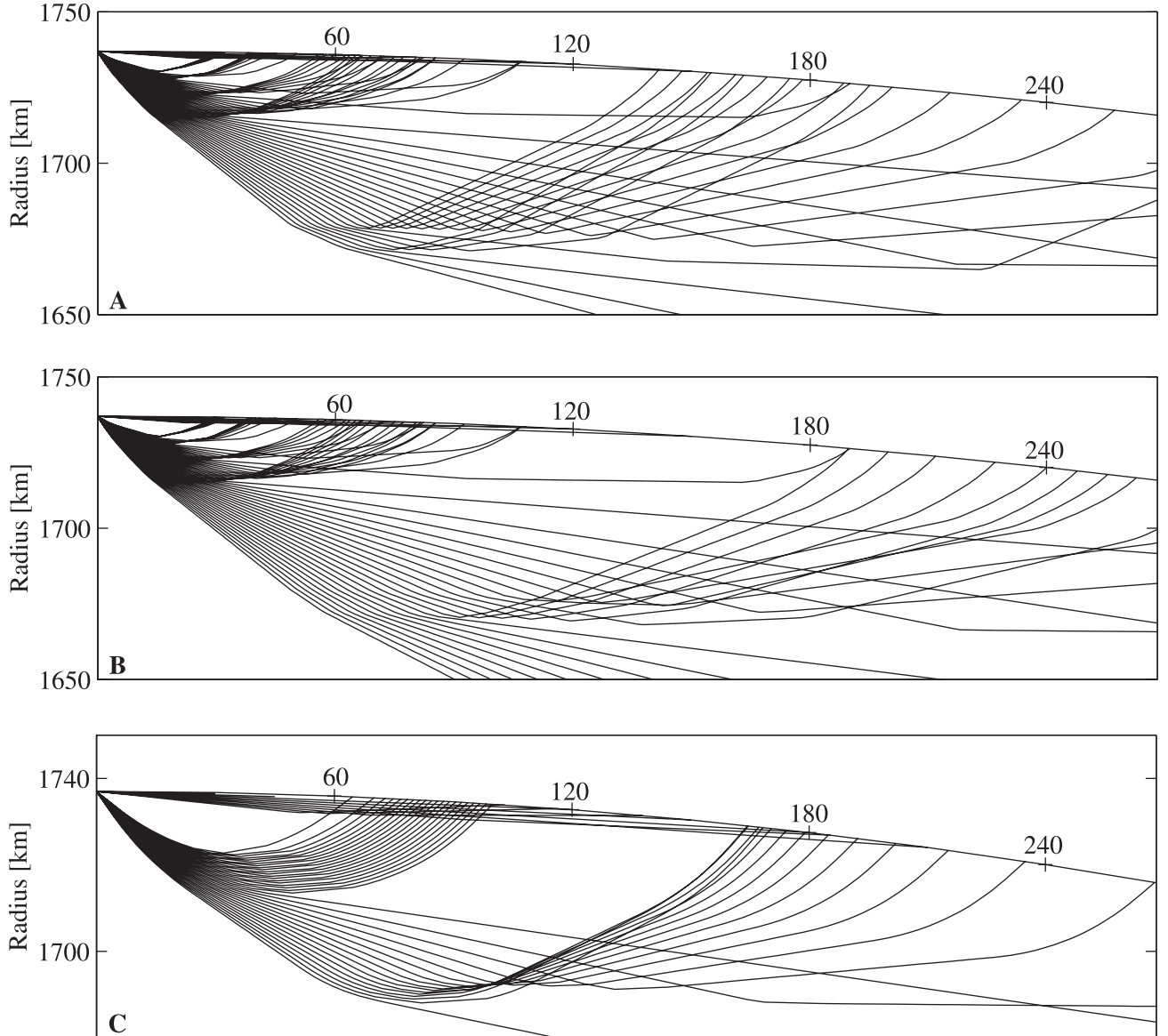


Figure 15. (a) Marginal posterior P wave velocity distributions depicting the lunar crustal velocity structure obtained from inversion of artificial impacts. The figure has been constructed from a total of 2709 models having satisfied \mathcal{H}_1 . The contour lines define eight equal-sized probability density intervals for the distributions. (b) Marginal posterior P wave velocity distributions for those models that satisfied \mathcal{H}_2 . The figure has been constructed from a total of 905 models. See color version of this figure at back of this issue.



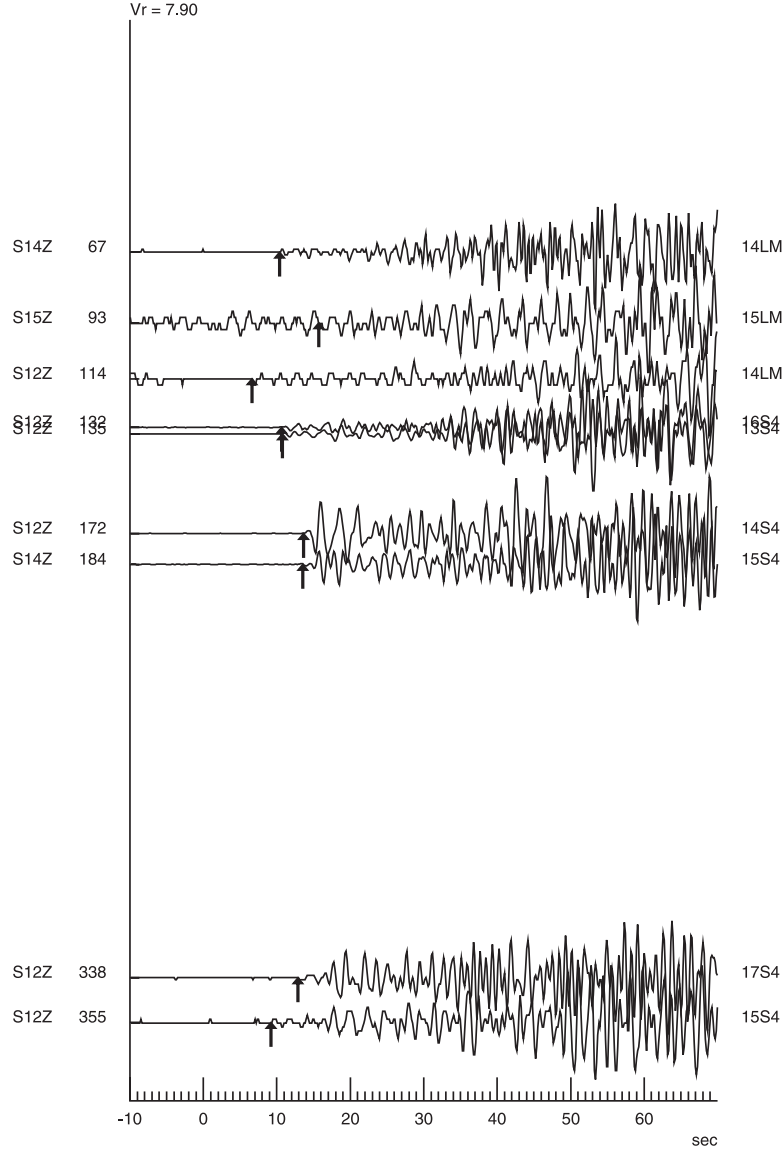


Figure 17. Composite plot of seismograms recorded from all artificial impacts on vertical components as a function of epicentral distance. Notation “S14Z 67” on the left of the first trace means station 14 vertical component recorded at an epicentral distance of 67 km, whereas “14LM” on the right of the trace signifies the mission and the impactor, in this case the lunar module of mission 14. Note the large-amplitude arrival at a distance of 172 km. Arrows indicate the position of first *P* wave arrivals as read by J. Gagnepain-Beyneix (manuscript in preparation, 2002). Seismograms are plotted to the same scale.

time to have a preference for generating substantial velocity decreases in the lower crust. Now, the sampling of such geologically unsound models could easily have been restricted. However, this would have required the introduction of prior information into the inverse problem, which implies constraining the numerical system, thereby reducing its nonuniqueness and not exposing the true model variability, which, as noted, was one of the primary motivations

for conducting our reanalysis of the Apollo lunar seismic data. Moreover, what actually constitutes “geologically unsound” features in a model is often debatable given its subjectivity.

[54] Of relevance in assessing Bayesian hypothesis testing as a tool for inferring conclusions is to inquire whether the conclusions reached are contingent upon the hypotheses; that is, is the Bayes factor dependent upon the specific

Figure 16. (opposite) Seismic ray paths in the Moon for a surface source. Figures 16a and 16b correspond to the Toksöz model with velocity discontinuities of 2 km/s and 1 km/s placed at 60 km depth. Figure 16c shows ray paths using the velocity model obtained here, with a velocity discontinuity of 1 km/s at a depth of 42 km. Where the density of rays is high, focusing of energy has occurred, leading to large amplitudes. Numbers on the lunar surface denote epicentral distance in kilometers.

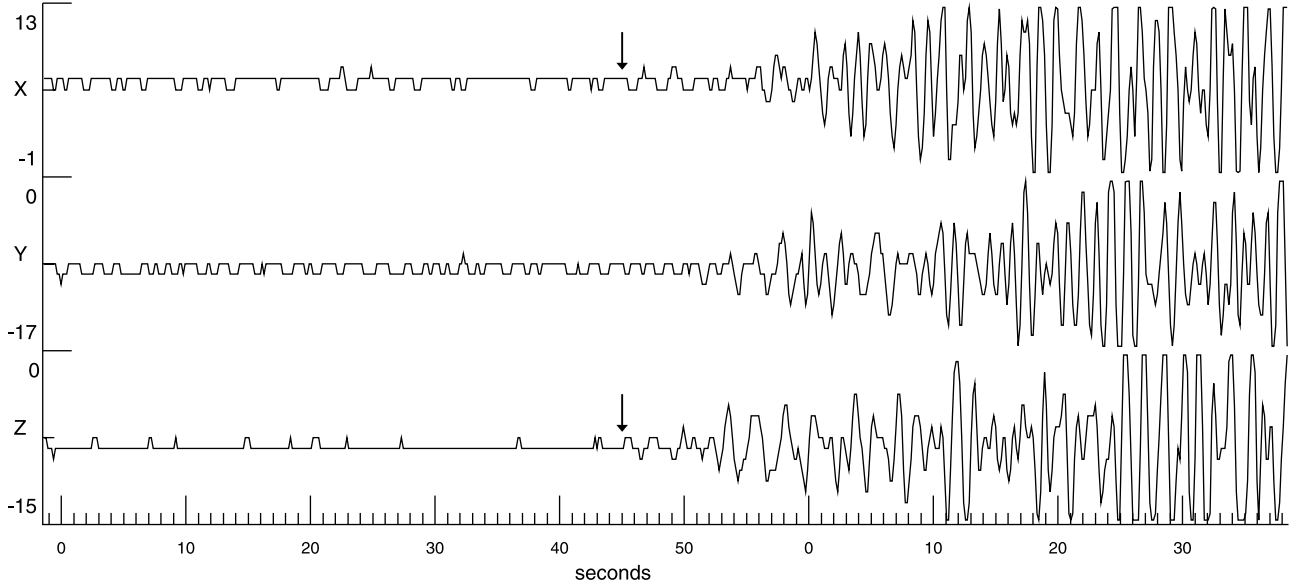


Figure 18. The SIVB 15 impact, 29 July 1971, as recorded at ALSEP 12 (LP) at a distance of 356 km. Traces start at 20 h 58 min 50.5 s. Ground motion is given in μm . Here X and Y represent the horizontal components of ground motion, and Z corresponds to vertical motion. The two peaks immediately after the arrow are clearly visible on the Z component and were identified as noise by Toksöz *et al.* [1972]. Ahead of these two pulses the signal itself contains very little noise, and the present authors are therefore more prone to picking the first arrival at the arrow, rather than 5 s later. While the Y component does not reveal anything, the X component contains a signal (indicated by the arrow) that corresponds to the one on the vertical component.

constraints employed in describing the physical features of interest? This is easily verified by simply altering the constraints. As an example, let us look for discontinuities in our models specified according to the following definitions:

1. Velocity gradients should reach $0.3 \text{ km s}^{-1} \text{ km}^{-1}$.
2. The velocity beneath the crust should reach a value of 8.0 km s^{-1} .

[55] We also extended the depth range in which we seek the crust-mantle boundary of hypothesis 1, from 35–45 km to 30–50 km, so as to have the same size as the range considered in hypothesis 2, but intuitively searching a greater depth range should just result in an augmented number of models satisfying \mathcal{H}_1 . The Bayes factor for this analysis is $\mathcal{B}_{12} = 7.5$, leading to a further increase in the plausibility of \mathcal{H}_1 relative to \mathcal{H}_2 . This should confirm the reliability of hypothesis testing.

[56] In view of the discrepancy between the crustal thickness estimate presented in this study and the ones obtained earlier, with values ranging from 65 km [Toksöz *et al.*, 1972] over 55 km [Toksöz *et al.*, 1974] to 58 km [Nakamura *et al.*, 1982] for the region below the Apollo 12 and 14 landing sites, a few comments are merited. (For the sake of completeness, it should be noted that another recent reanalysis of the Apollo lunar seismic data set, although using a different inversion technique, has also resulted in a thinner estimate for the crustal thickness [Chenet *et al.*, 2002].) In the study of Toksöz *et al.* [1972, 1974], amplitude data and synthetic seismograms were employed in addition to the travel times from the man-made impacts.

[57] Now, amplitude data do provide additional information on velocity discontinuities, since phenomena such as geometric focusing, defocusing, and reflection of seismic rays are controlled by velocity gradients. These effects are illustrated in Figures 16a, 16b, and 16c for three different velocity models. Figures 16a and 16b depict ray paths in the crust using the Toksöz model, with velocity discontinuities of 2 and 1 km/s situated at 60 km depth, respectively. It is clear that in order to focus rays at the appropriate epicentral distance with the Toksöz model so as to obtain the observed large first P wave arrival at a distance of 172 km (see Figure 17), a discontinuity of 2 km/s is needed, which is probably difficult to realize. Figure 16c, on the other hand, shows that it is indeed possible to fit the same data by a model corresponding to the one presented in this study, including a 1 km/s discontinuity at a depth of 42 km.

[58] The positive velocity discontinuity will, of course, produce a multiplicity in the travel time curve corresponding to the different arrivals, like the refracted first arrival, the direct arrival, and, if present, any reflected arrivals. Therefore the presence of any later P wave arrivals will provide additional evidence for velocity gradients or discontinuities inside the Moon. In the studies of Toksöz *et al.* [1972, 1974], later arrivals as well as surface reflected phases were reportedly identified in the seismograms and used to further evince the existence of a high-velocity discontinuity. However, the notorious coda seen on lunar seismograms, produced by intense scattering, obscures secondary and all later arriving phases, thereby severely hampering their identification. This

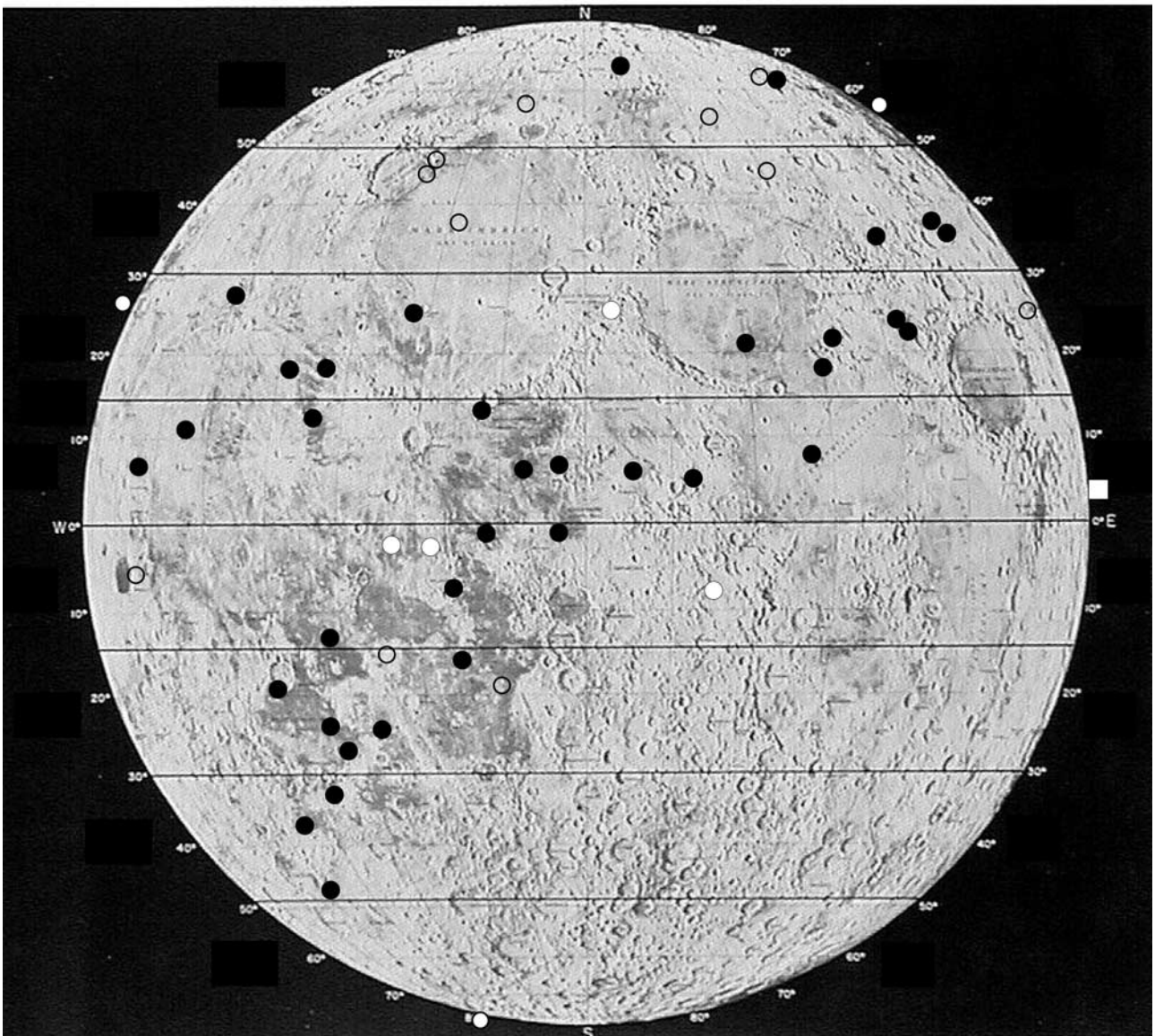


Figure 19. Map of the lunar nearside showing the four Apollo stations as well as deep and shallow moonquake epicenters taken from Tables 2 and 4, respectively. The white dots on the face of the Moon denote the four stations. Black dots indicate deep moonquakes, and open circles denote shallow moonquakes. On the rim of the Moon are additionally located three white dots indicating the location of the shallow moonquakes on the lunar farside. The white square on the right near the lunar equator shows the location of the lone farside deep moonquake.

does not necessarily mean that none of the purported secondary arrivals are real. Some of them may truly represent reflected and converted phases from some interface, but others may quite possibly be due to side scattering (Y. Nakamura, personal communication, 2001).

[59] The calculation of theoretical seismograms defines a more definitive approach, since these can be directly compared to their observed equivalents. The computation of synthetic seismograms is mostly dependent on the velocity structure but, of course, also depends on the seismic source and the response of the seismometer. For details of how seismograms were computed the reader is referred to *Toksöz et al.* [1972]. Figure 6 of that manuscript, in particular, shows a comparison of observed and synthetic seismograms, and it is apparent even from a first look that discrepancies are

present, especially when considering the impact recorded at an epicentral distance of 356 km. Concerning this particular seismogram, the authors state that (p. 497): "...the first two peaks of the observed seismograms are noise pulses and can be clearly identified as such in unfiltered seismograms." However, an independent review of this event, by the present authors (see Figure 18) does not qualify the two pulses as noise, but rather as the start of the signal, which, of course, leaves the synthetics and thereby the velocity model used in deriving these questionable.

[60] Having said this, one should be aware that the act of using uncertain data, in the sense of those discussed here when solving inverse problems, should be cautioned, since the probable inconsistency with which, as an example, secondary later arrivals have been read can lead to biasing

of the final result, as it is succinctly stated by *Tarantola* [1987]:

....With such kinds of data, it is clear that the subjectivity of the scientist plays a major role. It is indeed the case, whichever inverse method is used, that results obtained by different scientists (for instance for the location of a hypocenter) from such data sets are different. Objectivity can only be attained if the data redundancy is great enough that differences in data interpretation among different observers do not significantly alter the models obtained.

[61] In having performed a relocalization of all events, except the artificial impacts, we will briefly discuss the spatial distribution of the natural seismic events. The current depth distribution (50–220 km depth) of shallow moonquakes confirms earlier evidence suggesting the upper mantle to be tectonically active, implying that this region of the Moon is the only zone in the lunar interior where tectonic stresses thought to arise from thermal contraction and expansion are high enough to cause mechanical failure [*Nakamura et al.*, 1979]. When the epicenters are drawn on the lunar surface, some shallow moonquakes appear to occur, now as then, on the edges of impact basins, which had led to the suggestion that they were correlated [*Nakamura et al.*, 1979] (see Figure 19). However, because of the limited number of detected events, the statistical significance of this correlation must be considered inconclusive.

[62] Concerning the deep moonquakes, it was observed in earlier studies [*Lammlein et al.*, 1974; *Lammlein*, 1977; *Nakamura et al.*, 1982] that their source regions clustered within seismic belts. Two deep belts were identified by *Lammlein et al.* [1974], one belt extending ENE to WSW and another belt extending south. In addition, *Nakamura et al.* [1982] discerned two more belts in the NW quadrant trending W and WNW. The present distribution seems more to define a main SW to NE trending belt extending from source A_{24} at 37.8°S and 42.7°W to A_{85} at 35.9°N and 68.1°E , comprising 60% of the events investigated here and lying on an arc of a great circle (see Figure 19). This belt is ~ 250 km in width and 3800 km in length as measured on the lunar surface. The linear trends inferred earlier in the NW quadrant are less conspicuous with the present distribution. The only farside moonquake observed to date, situated at 3.7°N and 112.3°E , most clearly presents an isolated, well-located source of moonquake activity and does not appear to be part of the main NE to SW trending belt as previously noted [*Lammlein et al.*, 1974]. The curious lack of moonquake activity in the SE quadrant was also observed earlier and was ascribed as being due to the presence of a localized high attenuation zone or alternatively may simply be due to the absence of deep moonquakes beneath this mostly highland region [*Nakamura et al.*, 1982].

[63] As concerns the depth distribution of deep moonquakes, we have found that 16 events, corresponding to almost 50%, fall in the narrow range from 850 to 950 km.

7. Conclusion

[64] In this paper we have presented a detailed analysis of a general inverse problem as is constituted by the inversion of the Apollo lunar seismic data. This was done by using a

MCMC sampling algorithm, that is, by performing a random walk in a multidimensional model space that combines prior information with information from measurements and from the theoretical relationship between data and model parameters. Input to the algorithm were random models generated according to the prior distribution and the likelihood function. As output we assimilated random realizations of the posterior distribution, which contains all the information about the parameterized physical system derivable from available sources.

[65] We furthermore gave examples of how to investigate the models comprising the complex posterior distribution in a statistical fashion, using credible intervals, 2-D marginals, and Bayesian hypothesis testing. Common to all of them is the notion of conjuring up questions relating to any particular property of our interest, like the depth of a particular deep moonquake, for instance. All models exhibit an example of this property, and it may turn out that all of them provide the same value for it, in which case we would say that the depth of this moonquake is well-constrained by data. On the other hand, it might also happen that all models end up giving different answers to that question, thus rendering this feature ill-determined. Whatever the particular method, the point to note here is that within the probabilistic formulation of the general inverse problem, doing statistics is an all important asset which provides us with a clear probabilistic answer.

[66] It was also shown that because of the large credible intervals containing 99% of the probability distribution, the absolute velocities were not very well-determined. Two-dimensional marginals, on the other hand, could be used to investigate the correlation between any two parameters, shedding light on the presence of velocity changes, discontinuities, and the like. Moreover, Bayesian hypothesis testing was shown to be a very informative tool in providing evidence for the relative plausibility between any two hypotheses concerning features of interest. The advantage of using this form of hypothesis testing to examine a set of models as to certain features was made obvious. Instead of having to a priori constrain the system in certain ways from the outset, which means having to run the algorithm again with different sets of prior information, we chose to invoke as few prior constraints as possible so as to facilitate the comparison between any two hypotheses in order to gauge which particular feature is most probable given data and prior information. We specifically employed Bayesian hypothesis testing to distinguish between a thin and a thick lunar crust, which resulted in a Bayes factor B_{ij} of 4.2, clearly favoring a thinner crust with a thickness around 38 km.

8. Future Work

[67] Finally, we would like to touch on the matter alluded to earlier regarding the sampling of P and S wave velocities. Let us consider the following two relations governing the wave velocities as a function of the material parameters:

$$v_p = \sqrt{\frac{\kappa + 4/3\mu}{\rho}}, \quad (9)$$

$$v_s = \sqrt{\frac{\mu}{\rho}}. \quad (10)$$

From these two equations it is immediately apparent that the parameters v_p and v_s considered in this study are not independent when considered in terms of the elastic constants describing the medium. This raises the interesting problem of inverting for these parameters instead of the wave velocities using the same arrival time data set, since the inherent dependency that is here brought out would probably result in narrower distributions.

[68] **Acknowledgments.** We are grateful to Yosio Nakamura for a constructive review of this paper. An anonymous reviewer is also thanked for comments. Stimulating and highly informative discussions with Albert Tarantola concerning this study and inverse problems in general are much appreciated by the first author. Finally, we would like to extend our gratitude to Jeannine Gagnepain-Beyneix for making available Figures 17 and 18.

References

- Barnett, V., and T. Lewis, *Outliers in Statistical Data*, John Wiley, New York, 1984.
- Bernardo, J., and A. Smith, *Bayesian Theory*, John Wiley, 586 pp., New York, 1994.
- Chenet, H., J. Gagnepain-Beyneix, and P. Lognonne, A new geophysical view of the Moon, *Lunar Planet. Sci.*, XXXIII, abstract 1684, 2002.
- Duennebier, F., and G. Sutton, Thermal moonquakes, *J. Geophys. Res.*, 79, 4351, 1974.
- Goins, N., A. Dainty, and M. Toksöz, Seismic energy release of the Moon, *J. Geophys. Res.*, 86, 378, 1981a.
- Goins, N., A. Dainty, and M. Toksöz, Lunar seismology: The internal structure of the Moon, *J. Geophys. Res.*, 86, 5061, 1981b.
- Good, I., The interface between statistics and philosophy of science, *Stat. Sci.*, 3, 386, 1988.
- Hampel, F., P. Rousseeuw, E. Ronchetti, and W. Stabel, *Robust Statistics: The Approach Based on Influence Functions*, John Wiley, New York, 1986.
- Khan, A., K. Mosegaard, and K. Rasmussen, A new seismic velocity model for the Moon from a Monte Carlo inversion of the Apollo lunar seismic data, *Geophys. Res. Lett.*, 27, 1591, 2000.
- Khan, A., and K. Mosegaard, New information on the deep lunar interior from an inversion of lunar free oscillation periods, *Geophys. Res. Lett.*, 28, 1791, 2001.
- Kovach, R., and J. Watkins, Apollo 17 seismic profiling—Probing the lunar crust, *Science*, 180, 1063, 1973.
- Koyama, J., and Y. Nakamura, Focal mechanism of deep moonquakes, *Proc. Lunar Planet. Sci. Conf. 11th*, 1855, 1980.
- Lammlein, D., Lunar seismicity and tectonics, *Phys. Earth Planet. Inter.*, 14, 224, 1977.
- Lammlein, D., G. Latham, J. Dorman, Y. Nakamura, and M. Ewing, Lunar seismicity, structure and tectonics, *Rev. Geophys.*, 12, 1, 1974.
- Latham, G., M. Ewing, F. Press, and G. Sutton, Apollo Passive Seismic Experiment, *Science*, 165, 241, 1969.
- Latham, G., et al., Seismic data from man-made impacts on the Moon, *Science*, 170, 620, 1970.
- Latham, G., M. Ewing, F. Press, G. Sutton, J. Dorman, Y. Nakamura, M. Toksöz, D. Lammlein, and F. Duennebier, Passive Seismic Experiment, in *Apollo 16 Preliminary Science Report, NASA Spec. Publ., NASA SP-315*, sec. 9, 29 pp., 1972.
- Loudin, M., Structural-compositional models of the lunar interior, M.S. thesis, Pa. State Univ., Univ. Park, 1979.
- Mosegaard, K., Resolution analysis of general inverse problems through inverse Monte Carlo sampling, *Inverse Problems*, 14, 405, 1998.
- Mosegaard, K., and A. Tarantola, Monte Carlo sampling of solutions to inverse problems, *J. Geophys. Res.*, 100, 12,431, 1995.
- Nakamura, Y., HFT events: Shallow moonquakes?, *Phys. Earth Planet. Inter.*, 14, 217, 1977.
- Nakamura, Y., A₁ moonquakes: Source distribution and mechanism, *Proc. Lunar Planet. Sci. Conf. 9th*, 3589, 1978.
- Nakamura, Y., Shallow moonquakes: How they compare with earthquakes, *Proc. Lunar Sci. Conf. 11th*, 1847, 1980.
- Nakamura, Y., Seismic velocity structure of the lunar mantle, *J. Geophys. Res.*, 88, 677, 1983.
- Nakamura, Y., D. Lammlein, G. Latham, M. Ewing, J. Dorman, F. Press, and M. Toksöz, New seismic data on the state of the deep lunar interior, *Science*, 181, 49, 1973.
- Nakamura, Y., J. Dorman, F. Duennebier, M. Ewing, D. Lammlein, and G. Latham, High-frequency lunar teleseismic events, *Proc. Lunar Sci. Conf. 5th*, 2883, 1974a.
- Nakamura, Y., G. Latham, D. Lammlein, M. Ewing, F. Duennebier, and J. Dorman, Deep lunar interior inferred from recent seismic data, *Geophys. Res. Lett.*, 1, 137, 1974b.
- Nakamura, Y., F. Duennebier, G. Latham, and H. Dorman, Structure of the lunar mantle, *J. Geophys. Res.*, 81, 4818, 1976.
- Nakamura, Y., G. Latham, H. Dorman, A. Ibrahim, J. Koyama, and P. Horvath, Shallow moonquakes: Depth, distribution and implications as to the present state of the lunar interior, *Proc. Lunar Planet. Sci. Conf. 10th*, 2299, 1979.
- Nakamura, Y., G. Latham, and J. Dorman, How we processed Apollo lunar seismic data, *Phys. Earth Planet. Inter.*, 21, 218, 1980.
- Nakamura, Y., G. Latham, J. Dorman, and J. Harris, Passive Seismic Experiment long-period event catalog, Final Version, 1969 day 202–1977 day 273, *Galveston Geophys. Lab. Contrib.* 491, 314 pp., Univ. of Tex. at Austin, 1981.
- Nakamura, Y., G. Latham, and J. Dorman, Apollo Lunar Seismic Experiment—Final summary, *J. Geophys. Res.*, 87, A117, 1982.
- Sellers, P., Seismic evidence for a low-velocity lunar core, *J. Geophys. Res.*, 97, 11,663, 1992.
- Tarantola, A., *Inverse Problem Theory*, 613 pp., Elsevier Sci., New York, 1987.
- Tarantola, A., and B. Valette, Inverse problems: Quest for information, *J. Geophys.*, 50, 159, 1982.
- Toksöz, M., Geophysical data and the interior of the Moon, *Annu. Rev. Earth Planet. Sci.*, 2, 151, 1974.
- Toksöz, M., et al., Velocity structure and properties of the lunar crust, *Moon*, 4, 490, 1972.
- Toksöz, M., A. Dainty, S. Solomon, and K. Anderson, Structure of the Moon, *Rev. Geophys.*, 12, 539, 1974.

A. Khan and K. Mosegaard, Department of Geophysics, Niels Bohr Institute for Astronomy, Physics and Geophysics, University of Copenhagen, Juliane Maries Vej 30, 2100 Copenhagen O, Denmark. (amir@gfy.ku.dk; klaus@gfy.ku.dk)

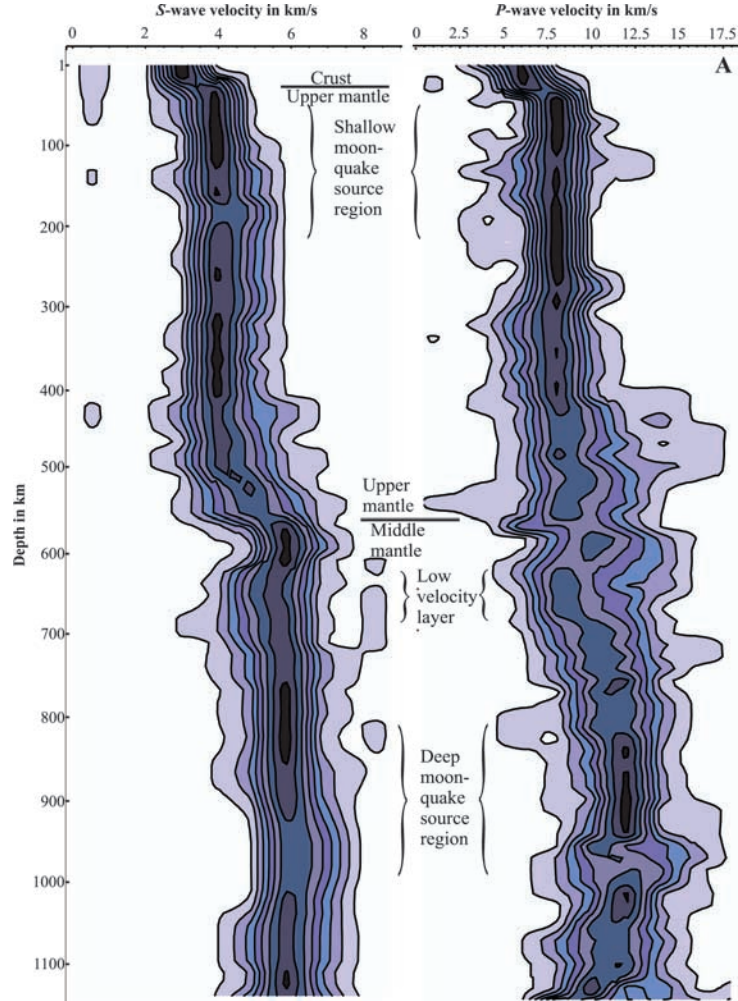


Figure 6. The marginal posterior velocity distributions depicting the velocity structure of the Moon. A total of 50,000 models have been used in constructing the two results. For each kilometer a histogram reflecting the marginal probability distribution of sampled velocities has been set up. By lining up these marginals, the velocity as a function of depth is envisioned as contours directly relating their probability of occurrence. The contour lines define nine equal-sized probability density intervals for the distributions. The uncertainties on the results are in part due to the large uncertainty in arrival time readings. It should be kept in mind that the velocity models are depicted using marginal probability distributions, and as such a model incorporating velocities of maximum probability does not necessarily correspond to the most likely model.

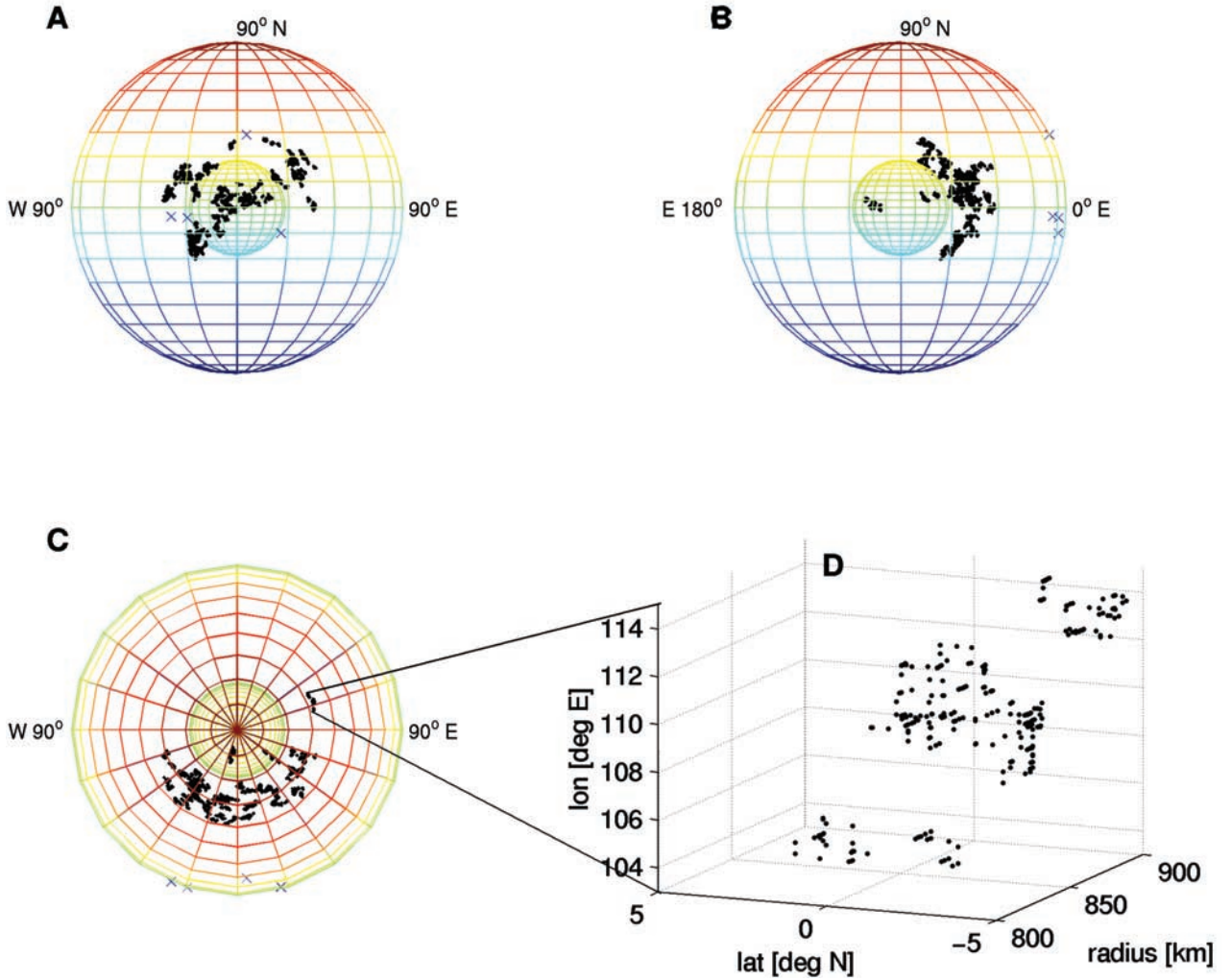


Figure 10. Sampled deep moonquake hypocenter coordinates showing the spatial distribution of quakes in the lunar interior. Figures 10a–10c display all sampled hypocenter coordinates for each deep moonquake source region from a number of different viewpoints. Since the individual source regions contain a cluster of many thousand samples, which are individually not distinguishable, Figure 10d depicts the distribution of sampled hypocenter coordinates for the lone farside moonquake, A_{33} , for enhancement. Figures 10a–10c also contain a central sphere with a radius of 500 km, which was included so as to enhance illustration of the spatial distribution of the moonquakes and is not meant to signify a lunar core. In Figure 10c we are viewing down on the lunar north pole, and the only farside moonquake observed to date, A_{33} , lying just over the eastern limb, is clearly visible. Crosses denote seismic stations.

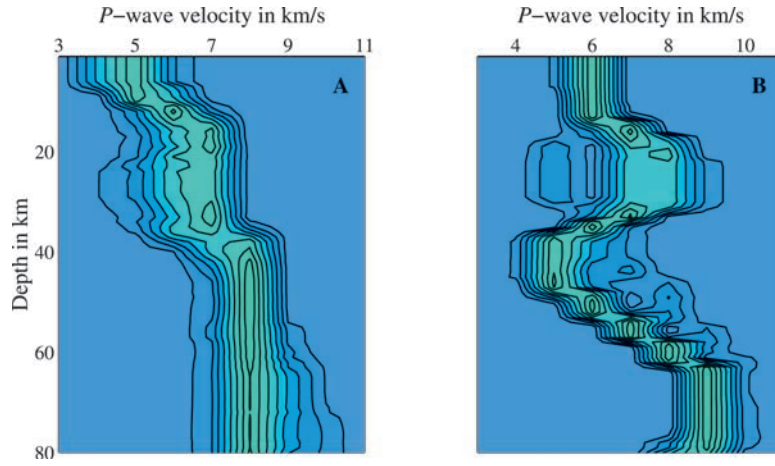


Figure 15. (a) Marginal posterior P wave velocity distributions depicting the lunar crustal velocity structure obtained from inversion of artificial impacts. The figure has been constructed from a total of 2709 models having satisfied \mathcal{H}_1 . The contour lines define eight equal-sized probability density intervals for the distributions. (b) Marginal posterior P wave velocity distributions for those models that satisfied \mathcal{H}_2 . The figure has been constructed from a total of 905 models.