A BAYESIAN APPROACH TO RIDGE FITTING IN THE ω-k DIAGRAM OF THE SOLAR FIVE-MINUTE OSCILLATIONS

Cherilynn A. Morrow Timothy M. Brown
University of Colorado / High Altitude Observatory, NCAR[*]
P.O. Box 3000
Boulder, CO 80307
U.S.A.

ABSTRACT. The acoustic oscillation modes of the Sun cluster along ridges of power in the ω-k plane. Fitting curves to these ridges provides input for methods that reveal information about the Sun's interior. This curve-fitting task is difficult due to noise in the data, close spacing between ridges at low k, and heuristic approaches to the fitting problem. The procedure we are investigating employs a simple but powerful rule from Bayesian decision theory in an effort to minimize the impact of such problems. This Bayesian approach allows one to make systematic use of prior physical and phenomenological information to assign a prior probability that a candidate curve gives the best fit to a ridge. Bayes' rule then permits one to update this probability using the new ridge power data. The maximally probable candidate curve given both new and prior information is chosen as the best fit.

1. INTRODUCTION

An ω-k diagram consists of a power distribution in the frequency (ω), horizontal wavenumber (k) plane. In such a power spectrum, solar acoustic oscillation modes (i.e. p modes) are represented by regions of high power clustered along ridges. Each ridge corresponds to oscillator modes with a particular radial wave number (n) and is labeled accordingly: P1 for n=1, P2 for n=2, etc. Efforts to fit such ridges involve using the power data to construct an analytical function that best reflects how the mode frequencies vary with horizontal wavenumber. This function provides input to both direct and inversion methods that aim to infer the physical properties of the Sun's interior. Attempts to learn about internal rotation, sound speed, and composition all depend on accurate representations of the ω-k ridges.

The ridge-fitting task is complicated by additional sources of power that confuse the mode peak locations. Ideal mode peaks are disguised by noise inherent in the nature of the signal, noise from background sources, and mixing and/or beating with nearby modes. Other complications arise from the close ridge-spacing at low k values, lack of observational resolution in time and/or space, and heuristic approaches to the fitting problem that may give inconsistent results. The Bayesian approach being investigated is intended to avoid or reduce the impact of such difficulties by providing a general framework for fitting patterns in the noisy and complex data of solar oscillations power spectra.

2. THE BAYESIAN APPROACH

Bayes' rule for manipulating probabilities is easily derived in two steps. Let P(A·O|H) represent the probability that the inferences symbolized by A and O are both true given that

---

[*] The National Center for Atmospheric Research is sponsored by the National Science Foundation.

485

the collection of assumptions represented by H are true. Following Cox (1961), one can write:

$$P(A \cdot O \mid H) = [P(A \mid H)] \, [P(O \mid H \cdot A)] \tag{1a}$$

$$P(O \cdot A \mid H) = [P(O \mid H)] \, [P(A \mid H \cdot O)] \tag{1b}$$

Since the conjunctive inferences A·O and O·A are identical, their probabilities may be equated to produce Bayes' rule:

$$P(A \mid H \cdot O) = \frac{[P(A \mid H)] \, [P(O \mid H \cdot A)]}{P(O \mid H)} \tag{2}$$

A = an inference whose probability depends on prior and new information

H = a set of hypotheses expressing prior information

O = new observational information

$P(A \mid H \cdot O)$ = posteriori probability of A (after consideration of O)

$P(A \mid H)$ = prior probability of A (before consideration of O)

$P(O \mid H \cdot A)$ = conditional probability of O (conditioned on truth of A)

$P(O \mid H)$ = normalization probability (must be nonzero)

Bayes' rule updates the prior probability of A to the posterior probability of A by means of the conditional probability which accounts for how new information in O either supports or denies prior beliefs about A. If O has no bearing on A, then P(O|H·A)=P(O|H) and one is left with the prior degree of belief in A. The decision process using Bayes' rule involves considering a domain of candidate inferences in search of the inference A that maximizes the posteriori probability, P(A|H·O). When found, this maximally probable inference is taken as the best answer to the decision-making problem.

3. THE BAYESIAN APPROACH APPLIED TO RIDGE-FITTING

In the context of fitting one ridge at a time in an ω-k power spectrum, let A,O & H be defined as follows:

A = a candidate curve is best fit to a ridge of power

O = solar oscillations data has been observed to give ridge power data

H = pertinent and reasonable assumptions that have been derived from physical theory, phenomenology, exploratory analysis of the current data, previous experience with similar data, and knowledge of the fitting method.

Given the above definitions of A,O & H, the specific meaning of the probability terms of the Bayes' formula can be examined. For convenience of computation and of reference to terms, the equation is considered in logarithmic form:

$$\ln P(A \mid H \cdot O) = \ln P(A \mid H) + \ln P(O \mid H \cdot A) - \ln P(O \mid H) \tag{3}$$

P(A|H·O) maps the probability that a candidate curve is best fit to the ridge given all assumptions and the ridge power data. The object is to find the candidate curve which maximizes this probability. To do this, one must choose the form and parameterization of the fitting function. A reasonable choice is a curve parameterized by a set of amplitudes $\{a_i\}$ of basis cubic splines $\{B_i\}$:

$$\omega(k) = a_1 B_1(k) + a_2 B_2(k) + \cdots + a_{nknot+2} B_{nknot+2}(k) \tag{4}$$

An optimization code (provided by R. Schnabel at the University of Colorado) then seeks, through computation of Bayes' formula, the $\{a_i\}$ corresponding to the maximally probable

candidate cubic spline. For each candidate (A) one need only compute the first two terms on the right hand side of Eq.3. The third, lnP(O|H), is independent of the choice of the candidate fit in A and thus has no influence on an optimizing scheme in its quest to maximize the value of Bayes' formula.

lnP(A|H) depends on the probability that a candidate curve is the best fit given only the assumptions in H. P(A|H) is called the prior probability because it reflects a rational degree of belief in the candidate curve being best prior to considering any of the new observational information in O. This term allows one to limit the domain of candidate inferences through use of prior knowledge and experience. A reasonable assumption is that a Duvall law (Duvall 1982) derived from previous data provides initial fits to the ridges whose ω values are very likely to be within a defined distance, dω (k), of the best fit at each k. Thus, the prior probability term we use is a box structure with half-width dω(k) at each k, and with smooth edges to avoid confusing the optimization routine. The prior probability term is most easily understood if one imagines that the amplitude of the boxes are all zero near the initial curve and go discontinuously to $-m$ at a distance d$\omega(k)$ from it. In this case, if *nout* of the nkpts points of a candidate spline are located outside of the boxes, then :

$$\ln P(A \mid H) = \frac{(\text{nkpts} - nout) \times (0) + (nout) \times (-m)}{\text{nkpts}} \tag{5}$$

This expression corresponds to a prior probability that is proportional to exp{-mf } where $f = nout$ /nkpts $=$ the fraction of points outside the box widths. The magnitude of m can be set depending on how sensitive one wants to be to a given number of points outside the region of prior expectation. If $f = 0$, then lnP(A | H) $= 0$ and this signals that the candidate curve completely satisfies prior beliefs about the location of the best fit and leaves the refinement to the conditional probability term which incorporates the new power data information.

The conditional probability term, P(O|H·A), reflects the degree to which the power data (O) around a candidate curve is consistent with the assertion that the candidate curve is best fit to the ridge (A). In order to compute this probability, one must define in H what the power data should look like around a best ridge fit. The two important aspects in making this definition are 1) the average cross-sectional shape (in ω) of a ridge and 2) the statistical behavior of the power data about that average shape. Experience suggests that a Lorentzian distribution superposed on a background level offers a good description of an average ridge cross section. Further, if the solar oscillations are produced by random driving of a damped harmonic oscillator, then the power spectrum of the resulting signal may be considered to have the statistical characteristics of a noise spectrum. Following Groth (1975) the probability density of power for one pixel of a noise spectrum is exp{ $- \Pi_{obs}$ /Π_{exp} } where Π_{obs} is the observed power in that pixel and Π_{exp} is the pixel's expected power ($\Pi_{exp} =$ the Lorentzian for pixels in a symmetric interval about any candidate curve). So the exponential provides an expression (based on reasonable assumptions) of the probability of observing Π_{obs} at (k,ω) given that Π_{exp} is what would be expected if the position of the candidate curve were best. This is precisely P(O|H·A) for one pixel. Since each data pixel is considered independent:

$$\ln P(O \mid H \cdot A) = -\sum_{k=1}^{\text{nkpts}} \sum_{\omega=\omega 1}^{\omega 2} \frac{\Pi_{obs}}{\Pi_{exp}} \left(\frac{1}{n\omega pts(k)} \right) \left(\frac{1}{\text{nkpts}} \right) \tag{6}$$

nkpts $=$ number of k points in candidate curve

$\omega 1, \omega 2 = \omega$ values at endpoints of interval about candidate curve at any k

n$\omega pts(k) =$ number of ω points in the summation interval at k

$\Pi_{exp} =$ expected power about candidate curve presumed best fit (Lorentzian)

$\Pi_{obs} =$ observed power at (k,ω)

The above expression is essentially a cross correlation, at each k, of the power data around a candidate curve with the negative inverse of the Lorentzian profile. Its value should be largest when the power data around a candidate curve most conforms to the expected shape of the power around a best fit to a ridge. Moreover, it properly accounts for the expected statistical behavior of power near the ridge and thus avoids unnecessary and possibly serious errors in the determination of a maximum cross correlation.

4. PRELIMINARY RESULTS

We applied the Bayesian scheme outlined above to an ω-k diagram provided by D. Haber at the University of Colorado. The velocity data (an 11 hour time string obtained in 1981 by F. Hill using a 512x128 diode array with the Vacuum Tower Telescope at the Sacramento Peak Observatory) was interpolated onto a spherical grid and collapsed in the short (N-S) spatial dimension so that the power spectrum primarily represents sectorial modes of oscillation.

The figure illustrates the initial fits (dashed) which are derived from a Duvall law and the fits resulting from the Bayesian approach (solid). There are both promising and disturbing aspects of these fits. At low and intermediate k values the Bayesian procedure brings the final fit into the central concentration of the ridge power. The results for P1 through P3 are particularly striking in this regard. Where the initial fits appear best -- low k, high n -- the Bayesian spline fitting process does not significantly affect the position of the curve. At high k values, however, it is disturbing that the final fits appear to stay attached to the path of the initial fit even though the initial position is away from the power. This problem is not yet fully understood, but it is possibly related to the optimizing scheme. Local maxima in the neighborhood of the global maximum of the Bayes' expression could fool the optimizer into settling for a lesser maximum. Use of a more sophisticated optimization procedure that is capable of locating a global extremum in a sea of local extrema may offer a solution. Efforts are currently being made to implement such an optimization scheme.

5. DISCUSSION

This Bayesian approach is promising and very general in its application to fitting problems. The user must:

* cast the problem into the Bayesian formalism by defining A,O & H and the meaning of the probability terms

* choose and parameterize the appropriate type of fitting function

* define some region of these parameters within which it is physically or phenomenologically reasonable to find the best answer

*specify the prior and conditional probability terms of the Bayesian expression in some plausible and effective way that depends on the physical and statistical nature of the data being fit.

*provide an optimization scheme for seeking the maximum of the Bayes' expression

The primary advantages of such an approach are the capability of significantly limiting the domain of possible best fits through use of prior knowledge, the capability of incorporating the actual physical and statistical nature of the data into the fitting process and the systematic method for combining the two capabilities via Bayes' rule.

In the ω-k ridge-fitting problem, prescribing an initial Duvall ridge gives a much needed head start at low k values where the ridge spacing is very close. The structure of the prior probability term is based on this initial fit because the Duvall law gives a good idea of the ridge's position and the amount and location of the power data that is relevant to that ridge. The structure of the conditional probability term, which works through Bayes' rule to update the

prior probability, is based on physical knowledge, the results of exploratory statistical analysis, and previous experience — all of which are a part of the set of hypotheses in H. A noisy, crowded ridge with a Lorentzian-enveloped cross-section superposed on a significant background can be accommodated and fit. In principle, better hypotheses (in H) about the nature of the data lead to better fitting results. High quality data (in O) would also improve the quality of the outcome, but even very noisy data can be handled if appropriate assumptions are made and incorporated into the Bayesian probability terms. The Bayesian approach to decision-making offers a good systematic framework for arriving at the best answer to a complex fitting problem such as that encountered in solar oscillations power spectra; it remains, as always, for the user to employ the information that will ensure the success of the process.
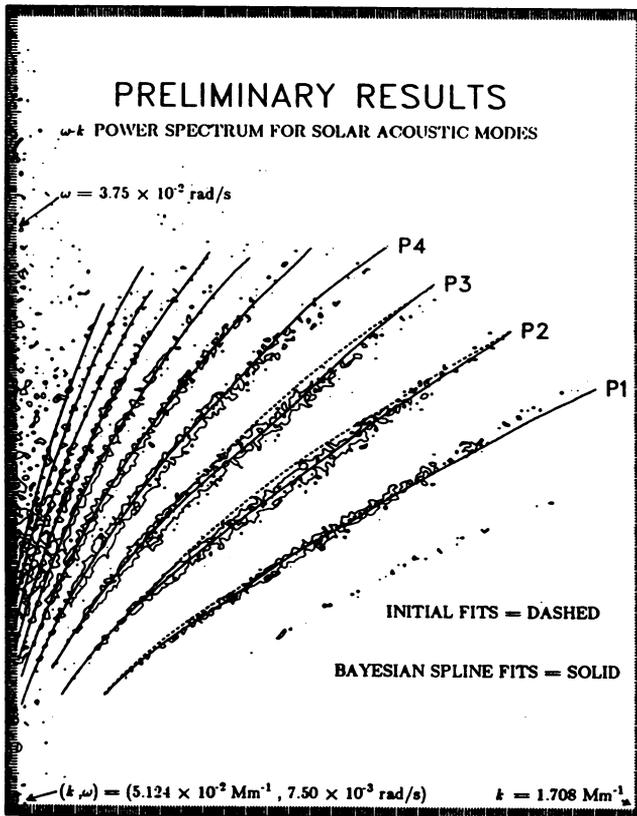

PRELIMINARY RESULTS

ω-k POWER SPECTRUM FOR SOLAR ACOUSTIC MODES

$\omega = 3.75 \times 10^{-2}$ rad/s

P4
P3
P2
P1

INITIAL FITS = DASHED

BAYESIAN SPLINE FITS = SOLID

$(k, \omega) = (5.124 \times 10^{-2}$ Mm^{-1}, 7.50×10^{-3} rad/s$)$ $k = 1.708$ Mm^{-1}

Figure: *Dashed lines are initial fits to the ω-k ridges derived from a Duvall law; solid lines are the results of the Bayesian approach to spline fitting. These preliminary results are promising but need improvement. Please see text for further discussion.*

REFERENCES

Cox, R.T., 1961, *The Algebra of Probable Inference*, Johns Hopkins Press, Baltimore, p. 16.
Duvall, T. L. Jr., 1982, *Nature* **300**, 242.
Groth, E. J., 1975, *Ap. J. Suppl.* **29**, 285.