Inference for a Proton Accelerator Using Convolution Models

Herbert K. H. Lee, Bruno Sansó, Weining Zhou, and David M. Higdon

Proton beams present difficulties in analysis because of the limited data that can be collected. The study of such beams must depend on complex computer simulators that incorporate detailed physical equations. The statistical problem of interest is to infer the initial state of the beam from the limited data collected as the beam passes through a series of focusing magnets. We are thus faced with a classic inverse problem where the computer simulator links the initial state to the observables. We propose a new model for the initial distribution that is derived from the discretized process convolution approach. This model provides a computationally tractable method for this highly challenging problem. Taking a Bayesian perspective allows better estimation of the uncertainty and propagation of this uncertainty.

KEY WORDS: Bayesian statistics; Computer simulator; Density estimation; Inverse problem.

1. INTRODUCTION

The particle accelerator is a major tool in modern experimental physics. For it to be an effective tool, scientists must be able to understand the characteristics of the particle beam and optimally focus the beam. The combination of physical laws and fast computers allows an accurate description of the beam if the initial conditions are known. But these conditions cannot be measured directly, and thus inferring these conditions is the primary statistical goal. In our work we look at a particular proton accelerator, but the methodology developed herein should easily generalize to other accelerators (as well as other applications) and can play an important role in the design of future accelerators.

This problem is difficult because of the limited information that can be collected about the particle beam. What is measured during an experiment are one-dimensional histograms of the relative frequencies of the particles at certain points along the path of the beam. From this information, scientists need to learn about the initial distribution of the particles in both their x- and y-dimension phase spaces, that is, their position and momentum in each dimension. A key computational tool is a detailed computer simulator that predicts the evolution of the beam given a particular set of initial conditions. Thus we have a classic inverse problem, where we are attempting to learn about the unobservable initial state from highly transformed and simplified data with computer code providing the link (see, e.g., Yeh 1986). Inference involves an iterative procedure whereby initial states are proposed and run through the simulator, the predicted results are computed and compared with the observed data, and then the initial proposal is modified in an attempt to improve the match between the computed results and the data.

We model the initial distributions nonparametrically through a convolution of a discrete gamma process with a Gaussian kernel. This approach turns out to be equivalent to using a mixture of Gaussians for density estimation, except with the mixture components fixed to be identical distributions except with different centers that are restricted to lie on the lattice points of a regular grid. This discrete approximation is a compromise that provides a high degree of flexibility yet also can be fit in practice with the available data. In many inverse problems, it is important to avoid models with more degrees of freedom than can be fit with the scarce data available. Our approach provides sufficient flexibility without going too far.

This work builds on an earlier analysis of this dataset (Lee, Sansó, Zhou, and Higdon 2006), wherein it was assumed that the initial joint density of position and momentum was bivariate normal for each of the horizontal and vertical dimensions. Although this assumption is not uncommon in physics (e.g., with the RMS beam envelope equations (Sacherer 1971)), it is also a very strong assumption. It tends to be good at characterizing the main features of the beam but is unable to model finer nuances. Here we eliminate the restrictive parametric assumption, allowing more detailed inferences to be drawn.

We work within the Bayesian paradigm, because it provides a framework for better accounting for uncertainty, particularly in the context of computer experiments and inverse problems (e.g., Kennedy and O’Hagan 2001). As with many inverse problems, a range of initial states can produce similar fits for the data, so that the problem here is underspecified (see, e.g., Oliver, Cunha, and Reynolds 1997; Lee, Higdon, Bi, Ferreira, and West 2002). The Bayesian approach naturally allows us to find a range of highly plausible initial states that can be reported through posterior distributions or intervals.

The next section provides details on the accelerator experiment, including the physical setup and the data collected. This is followed by an introduction to our statistical model, some results, and finally some comments and future directions.

2. PARTICLE ACCELERATOR EXPERIMENTS

The low-energy demonstration accelerator (LEDA) is a 6.7-MeV proton accelerator developed to study high-current beams. It is a linear accelerator with 52 sets of magnets to focus the beam over a span of 11 meters. One primary area of investigation for physicists is the halo effect, the tendency of some particles to spread away from the central area of the beam. These particles can end up leaving the beam and escaping into the surroundings, creating a radioactivity hazard with repeated use of the accelerator. A better understanding of the halo-forming...
process would allow countermeasures to be taken; in particular, the focusing magnets could be adjusted to reduce the halo. Such focusing techniques require good understanding of initial state of the protons as they leave the accelerator so that the magnets can be positioned properly. Thus the statistical inference for the initial distribution is a key component of the entire process. Our work here is confined to beam inference, with study of the halo effect itself a separate research project for our collaborating physicists.

The initial state of the beam is determined by the distribution of the particles that compose the beam, in particular the cross-sectional distributions (in the $x$ and $y$ directions) of position and momentum, denoted $(x, p_x, y, p_y)$. Whereas the $x$ and $y$ dimensions can be treated as independent, the position and momentum are expected to be correlated within a dimension, as discussed later.

If the initial distributions of position and momentum are known, then the future paths of the particles can be predicted with reasonable accuracy using physical models of particle movement (with, e.g., the Vlasov and Poisson equations) as well as modeling external forces on the particles and the interparticle Coulomb field (Dragt et al. 1988). These models are solved numerically with computer code. The scientists at Los Alamos have supplied us with the program ML1 5.0, which simulates the particle paths in such a manner (Qiang, Ryne, Habib, and Decyk 2000). A typical high-fidelity run on a SunBlade 1000 workstation takes about 6 minutes. A lower-fidelity run (with only 8,000 particles instead of 100,000) runs in about 2 minutes and tends to produce similar results. Thus the simulator is reasonably fast but is clearly the limiting factor in iterative methods.

After the beam emerges from the accelerator, it is focused with a series of magnets called quadrupoles. Magnets are used in sets of pairs, with the first pair focusing the beam in the $x$ direction, but defocusing it in the $y$ direction. The second pair focuses $x$ but defocuses $y$. Through iterative focusing and defocusing, the beam is gradually controlled in both directions. Figure 1 shows a simulation (from computer code) of the 5th, 15th, ..., 95th percentiles of the positions of particles in the beam as they pass through eight pairs of magnets. The magnets that focus in the $y$ direction are shown in red (dark gray in grayscale), and the $x$ focusing magnets are shown in blue (light gray). In the figure the beam can be seen widening in the $x$ dimension as it passes through the red magnets and narrowing as it passes through the blue ones, with the reverse effect in the $y$ dimension. By passing the beam through a large number of quadrupoles (on the order of 50), the beam can be slowly focused in both directions. This can be difficult to see in Figure 1, because each individual quadrupole only makes a small incremental improvement in the combined two-dimensional focusing problem.

This focusing affects the beam in that it induces a negative correlation between the position and momentum; that is, particles further away from the center are pushed back toward the center, with the force being a function of the distance. This action is a consequence of Maxwell’s equations for electromagnetic fields. In contrast, defocusing induces a positive correlation. The net result of repeated focusing and defocusing can produce interesting nonlinear effects, posing challenges in studying the halo effect.

Figure 1. Simulation of particle beamlines as they are focused by a series of quadrupole magnets. (a) Progression of the particles in the $x$ dimension; (b) progression of the particles in the $y$ dimension. Magnets are denoted by shaded areas, with alternating magnets focusing and defocusing in each direction. The wires are shown with dashed lines.

It is important to note that although the simulator works by tracking the progress of individual particles as they move through the apparatus, this particle-level detail can be done only through simulation. In a real experiment, particles cannot be observed individually, and we cannot jointly observe the position and momentum information. Instead, the data available are histograms of particle positions at fixed locations along the beam. The dashed vertical lines in Figure 1 represent wirescans. Physical wires are set up, as the beam crosses these wires, electrical current is produced, and a histogram (here with 256 bins, which is sufficient resolution to make it appear like a curve) of particle positions can be created. These wirescans in each dimension provide the only observable data. Figure 2 provides some additional illustration. The top row are the first, third, fifth, seventh, and ninth (final) wirescans in the $x$ direction, and the bottom row shows the corresponding $y$ wirescans. The figure also shows the simulated particle clouds. The second row shows the $x$-position versus $x$-momentum distribution at these five wirescan locations, and the fourth row shows the corresponding distributions for the $y$ position and momentum. The middle row shows the joint $x$ and $y$ positions. Note that in this middle row, the positions are independent across dimensions, whereas within dimensions, the position and momentum are dependent.

The plots in these two figures all fit together. For example, at the start of the beam (the left side) in Figure 1 is the first wirescan, with the trajectories somewhat spread apart for $x$ and relatively tight for $y$. The upper left plot of Figure 2 shows the resulting spread-out $x$ wirescan, and the lower left plot shows the highly peaked histogram for $y$ position. The second column of Figure 2 shows the wirescans and particle clouds at the third
Figure 2. Simulation of wirescans and particle clouds at the locations of wirescans 1, 3, 5, 7, and 9 for the same experiment shown in Figure 1. The top row is the \( x \) wirescans, the second row is \( x \)-momentum versus position, the third row is \( y \)-position versus \( x \)-position, the fourth row is \( y \)-momentum versus position, and the bottom row is the \( y \) wirescans.

wirescan (third dashed line from the left in each panel of Fig. 1) after passing through two pairs of magnets. Now the \( x \) positions are relatively tight, whereas the \( y \) distribution is wider. Also note the nonlinear behavior of the particle clouds, showing intriguing relationships between position and momentum in each direction.

We want to emphasize that in practice we can only observe wirescans, and that none of the true trajectory or particle cloud information is available. In fact, the Heisenberg uncertainty principle declares that it is not possible to measure both the position and momentum of a particle. Thus all of our statistical inference must be based only on the wirescan position histograms (the top and bottom rows of Fig. 2). Although the first wirescan is highly informative about the initial position distribution, by itself it contains no information about the initial momentum distribution. Multiple additional wirescans are needed to draw inferences on momentum due to the indirect and nonlinear information available from the experiment.

The physicists involved tried to use maximum entropy methods (Mottershead 1996) to draw inferences about the initial configuration, but found that they were unable to reconstruct the noisy data of this experiment. Thus they brought the problem to us, and we proceeded with a fully Bayesian statistical model, as described in the next section.

3. STATISTICAL MODEL

Because the wirescans provide the only observable data, we must build our likelihood from the difference between the observed wirescans and those predicted by the computer simulator for a particular initial distribution. Thinking nonparametrically, our unknown parameter is the four-dimensional initial distribution of position and momentum in each of the \( x \) and \( y \) dimensions, that is, the distribution of \((x, p_x, y, p_y)\). For a given value of this distribution, we can simulate particles from this distribution, run the particles through the simulator, and obtain predicted wirescans \( \hat{w} \). We then compare these predictions with the observed data \( w \). We use a Gaussian error structure for the difference,

\[
L = \left( \frac{\tau^2}{2\pi} \right)^{n_{scan}n_{bin}/2} \exp \left\{ -\frac{\tau^2}{2} \left( (w_x - \hat{w}_x)^T (w_x - \hat{w}_x) + (w_y - \hat{w}_y)^T (w_y - \hat{w}_y) \right) \right\}, \tag{1}
\]

where \( w_x \) is the wirescan data for the \( x \) dimension and \( w_y \) is that for the \( y \) dimension, \( \tau^2 \) is a precision parameter, \( n_{scan} \) is the number of wirescans in each dimension, and \( n_{bin} \) is the number of bins per wirescan. We use an independent error structure, similar to an \( L^2 \) distance metric, because we have found that
it represents a reasonable trade-off between model details and computational efficiency. In some cases, there can be strong dependencies in the errors (Lee et al. 2006). In other cases, the errors show much less dependence and can be effectively modeled by an independence structure. Here we have found that assuming independent errors within a scan produces results comparable to those from more complex models, so we keep the analysis more tractable by using an independent error structure.

3.1 Nonparametric Density Estimation From a Convolution Perspective

Because we are essentially interested in density estimation for the initial distribution, it would seem that we have various approaches from which to choose. Although we make no claim to have tried all of them, we have tried a number of standard approaches and found them unproductive for this application. The key difficulty is that we want to do density estimation for unobservable data. The initial distribution of states for the particles is inherently unobservable, so we must base our inference on the histograms of some of the dimensions of the transformed particle cloud. Although we want sufficient flexibility in our model to accommodate a wide variety of possible initial distributions, we also cannot allow too much freedom or else estimating the parameters becomes impractical. In essence, the amount of model flexibility provided by modern parametric and nonparametric density estimation methods results in an underspecified problem. We also must take into account that each evaluation of (1) requires several minutes. Our approach balances flexibility with the ability to fit the model in a practical time period using the data available in this problem.

To achieve a maximally flexible model, we base our approach on the convolution representation of a Gaussian process, because a Gaussian process is itself a highly flexible nonparametric model for spatially related data in multiple dimensions. A Gaussian process is a random process $Z(s)$ such that for any finite subset of indices $s_1, \ldots, s_n \in S$, $Z = (Z(s_1), \ldots, Z(s_n))^T$ has a multivariate normal distribution. It is common to assume a constant or polynomial form for the mean and to assume that the covariance structure is stationary and isotropic, so that the correlation between any two points depends only on their distance. A further common assumption is that this correlation is a simple parametric function of distance (Cressie 1993; Wackernagel 1998; Stein 1999).

An alternative method of obtaining a stationary Gaussian process is by convolving white noise with a smoothing kernel (Thiébaux and Pedder 1987; Barry and Ver Hoef 1996; Higdon 2002). Let $x(s)$ be a white noise process (or Wiener process “derivative”; e.g., Priestley 1981) and let $k(\cdot; \phi)$ be a kernel, possibly depending on a low-dimensional parameter $\phi$. Convolving $x$ with this kernel produces a Gaussian process,

$$Z(s) = \int_S k(u-s; \phi)x(u)\,du$$

where $W$ is a Wiener process. The process $Z(s)$ is such that for $s, s' \in S$ and $d = s - s'$,

$$c(d) = \text{cov}(Z(s), Z(s'))$$

that is, $c(d)$ is the convolution of the kernel with itself (Kern 2000).

In practice, a discretized version of (2) is used as an approximation. Defining the background process on a grid $s_1, \ldots, s_m$ and suitably normalizing the kernel gives

$$Z(s) \approx \sum_{i=1}^{m} k(s_i - s; \phi)x(s_i),$$

where $x(\cdot)$ is white noise. This approximation provides a computationally efficient mechanism for using Gaussian processes in practice when the dimension of $S$ is small, particularly for larger datasets or inverse problems, because of the reduction in effective dimension of the parameterization (Higdon, Lee, and Holloman 2003).

Here we seek to take advantage of this lower-dimensional approximate representation of a highly flexible process. We need to modify this model because the proton distribution must be nonnegative and must be a proper density that integrates to 1. Our approach here shares some similarities with methods of Ickstadt and Wolpert (1999). We replace the white noise process $x(\cdot)$ with a normalized gamma process,

$$x(s_i) \overset{iid}{\sim} \Gamma(\alpha, \beta),$$

$$x(s_i) = \frac{x(s_i)}{\sum_{i=1}^{m} x(s_i)}.$$  

The joint set of points $x$ thus has a Dirichlet distribution, but we avoid calling $x(\cdot)$ a “Dirichlet process,” because that terminology already has another well-established meaning. We also note that additional flexibility can be gained by allowing the $\alpha$ and $\beta$ parameters to vary by location. In the simpler iid case, which we use here, the value of $\beta$ is irrelevant because of the normalization, so we do not consider it any further. In either case, the discrete convolution gives

$$Z(s) = \sum_{i=1}^{m} k(s_i - s; \phi)x(s_i).$$

The resulting process $Z(\cdot)$ thus approximates a nonparametric density estimate, where the actual parameters being fit are $\{x(s_1), \ldots, x(s_m)\}$. By fixing $\alpha$ and $\phi$, we obtain a method that retains much flexibility but also is of sufficiently low effective dimension that we can actually fit the model for our proton accelerator application. In choosing a Gaussian kernel, we obtain a special case of several other methods. This approach can be seen as a radial basis network (e.g., Bishop 1995, chap. 5) with basis functions restricted to a grid or, equivalently, as kernel density estimation with kernels of identical width and restricted to a grid. The advantage of viewing our approach in the context of convolutions is that it gives us a theoretical framework to see that not too much of the original nonparametric flexibility is lost when compared with standard density estimation techniques (Barry and Ver Hoef 1996; Higdon 2002).
One of the most flexible nonparametric density estimation approaches involves assuming that a density \( f(\psi) \) can be represented using the mixture \( \int k(\psi|\theta) dG(\theta) \), where \( k \) is a kernel dependent on parameters \( \theta \) and \( G \) is a random probability measure (see, e.g., Müller and Quintana 2004 and the references therein). A mixture model for densities with compact support in \( \mathbb{R}^2 \), was developed by Kottas and Sansó (2007). This would be theoretically appropriate for our present application, but the computations involved in the estimation are very substantial (see Neal 2000). To appreciate the difference between the nonparametric mixture model approach and the one we present in this article, consider a discretization of the mixture given by \( \sum_i f(\psi|\theta_i)g(\theta_i). \) Here we have that the number, location, and weights of the kernel parameters are random. As for most inverse problems, the information provided by the data is not sufficient to inform all these degrees of freedom. Thus in our model we fix the \( \theta_i \)'s and assume that the weights are random.

Thus our prior model for \( x \) position and momentum is a two-dimensional process, \( Z_x \), of the form (4), and our model for the \( y \) position and momentum is another such process, \( Z_y \), independent of the \( Z_x \) process. We use a \( 32 \times 32 \) grid for each background \( s \) process. This was the smallest grid that provided sufficient flexibility for a good fit; larger grids did not much improve fits but took significantly longer to run. Based on conversations with scientists at Los Alamos and on experience with simulated examples, we set \( \alpha = 1 \). The kernel \( k \) is a bivariate normal with mean 0 and covariance matrices for \( x \) and \( y \),

\[
\Sigma_x = \begin{bmatrix} 3.9 \times 10^{-4} & 0 \\ 0 & 4.8 \times 10^{-7} \end{bmatrix}
\]

and

\[
\Sigma_y = \begin{bmatrix} 3.4 \times 10^{-4} & 0 \\ 0 & 4.3 \times 10^{-7} \end{bmatrix}.
\]

Figure 3 shows the grid of our background \( s \) process with the contours of one of the kernels, to give an idea of the smoothing effects of the kernels.

This prior is then combined with the likelihood from (1) to obtain the posterior. The value of the precision, \( \tau^2 \), in the likelihood is set a priori to \( 1/(1.2 \times 10^{-5}) \). Ideally we would fit this parameter from the data. However, as with many inverse problems (see, e.g., Oliver et al. 1997; Lee et al. 2002), the data do not provide sufficient information to directly estimate this parameter. We have tried putting a prior on \( \tau^2 \) and estimating its posterior, only to find that the posterior looks very similar to the prior for a wide range of possible priors. Thus \( \tau^2 \) must be fixed, and we set it to \( 1/(1.2 \times 10^{-5}) \). This choice represents a combination of knowledge from subject area experts and practical computing considerations, in that its value affects convergence properties of the MCMC sampler. Increasing \( \tau^2 \) forces the model to produce predictions that more closely match the observed wirescans. But large values of \( \tau^2 \) also cause the MCMC sampler to move very slowly through the parameter space. Because each iteration takes several minutes, it can be impossible to obtain a good run in a reasonable amount of time (such as 2 weeks). Thus the choice of \( \tau^2 \) is a practical compromise between expected scientific accuracy and computability.

3.2 Model Fitting

Because the likelihood cannot be written in closed form, but must be evaluated numerically by running the particle simulator, we use a Metropolis–Hastings algorithm (see, e.g., Gamerman 1997) to explore the posterior distribution of the background processes and hence the two position–momentum space distributions.

The standardized latent values \( x(s) \) are fit using Metropolis–Hastings by doing a random walk on the log-transformed samples. For each iteration, evaluating the likelihood requires running the simulator, which requires a sample of particles. New values for the latent process \( x^*(s) \) are proposed, and, conditioned on these new values, we draw a sample of size 8,000 from 1,024 lattice cell locations with the probabilities specified by the distribution induced by the convolution of the \( x^* \) process for each of the \( x \) and \( y \) dimensions independently, as per (3). Thus we obtain a sample of size 8,000 of \( (x, p_x, y, p_y) \) from the product of convolutions. Next, we run the proposed initial configuration of particles through the simulator to obtain a fitted wirescan. Now the likelihood (1) can be evaluated, and the new likelihood compared to that of the previous iteration. The proposed values are accepted according to the standard Metropolis–Hasting probabilities.

Because it would be rather time-consuming and computationally complex to update each latent value \( x(s) \) for each lattice cell individually for both the \( x \) and \( y \) dimensions, we collect latent values \( x(s) \) into several blocks according to their positions in space and update each group separately for both dimensions.

4. SIMULATED EXAMPLE

To check the validity of our proposed approach, we examine its behavior on a simulated example, so that we can check our results against a known truth. Because we have been told by scientists at Los Alamos to expect some clumping in the phase space distributions (position vs. momentum), we created...
the true distribution to be a mixture of two normals for each of the $x$ and $y$ phase spaces. These distributions are shown in the bottom row of Figure 4. We generated a cloud of particles from this joint distribution and ran it through the simulator to produce wirescan data, and then used only that wirescan data to try to infer the starting distribution. The top row of Figure 4 shows our posterior mean fitted distributions. We are pleased that our nonparametric methodology finds a good match when the truth is known.

Figure 5 shows additional results about the fit for this example. In that figure, each column represents a wire. The top two rows are for the $x$ dimension, and the bottom two rows show the $y$ dimension. The first and third rows are realizations from the estimated posterior distributions of the particle cloud ($x$ or $y$ position and momentum) as the beam passes through and is reshaped by the quadrupole magnets. We show realizations because the simulator requires a set of particles as its input rather than a distribution, so we can only show how the particle cloud is propagated by using realizations. The second and fourth rows show the true wirescan data (circles), the estimated posterior mean scans (solid line), and posterior interval estimates (dashed lines). The posterior mean is quite close to the truth, and the credible intervals provide a measure of our uncertainty. The intervals are relatively narrow, and we conjecture that this is because of a combination of two factors. First, the ground truth is a fairly “simple” distribution that can be fit relatively cleanly by our model, although we note that it did take some effort in model development to get our nonparametric approach to consistently fit such an example, which is actually relatively smooth. A poorly tuned convolution-based approach can lead to fits with a larger number of well-separated thin peaks, which would fail to capture the truth in this case. Second, there is some nonidentifiability in the model, meaning that different input distributions can lead to very similar wirescans, so that uncertainty in the distributions may appear to be diminished in fitted wirescan plots. Figure 6 displays three realizations from the posterior to help show the variability in the fit. These realizations all provide reasonably good fits to the wirescan data and so give a sense of both our posterior uncertainty and the nonidentifiability in this problem.

5. APPLICATION TO REAL DATA

Returning to the motivating application, we now present our results on data from a real run of a proton accelerator. Scientists at Los Alamos National Laboratory have provided us with four pairs of wirescan data (one member of each pair for the $x$ coordinates, one for the $y$ coordinates) (Allen et al. 2002; Allen and Pattengale 2002). From these scans, we are to infer the initial distribution of the beam in terms of its $x$ and $y$ position and momentum distributions.

Figure 7 shows our results on this dataset. As in Figure 5, each column is a wire, the top two rows are for $x$ and the bottom two rows are for $y$, with estimated particle cloud posterior distributions and fitted scans with 95% credible intervals for each. The fitted wirescans are reasonably close to the observed values, indicating a good fit. Figure 8 shows the resulting posterior mean for the initial distributions of position and momentum for each of the $x$ and $y$ dimensions. In both cases there is a primary mode representing most of the mass, along with several auxiliary modes. We have been told by collaborators at Los Alamos that separate “clumps” of mass are to be expected in such plots. As a further check on our results, we did a number of restarts
of the MCMC process with different initial solutions. Those chains that converged tended to converge to this same solution, although some would get stuck in local modes and some would fail to converge in a reasonable amount of time (e.g., 2 weeks). Given a reasonable starting location, an MCMC run takes about 2 weeks on a SunBlade 1000 workstation, with nearly all of the computational effort going into running the simulator to evaluate the likelihood function.

The clumps in the posterior that are further from the center show that the beam is in need of additional focusing. Using this information, along with additional simulations, the physicists can attempt to tune the focusing process by changing the positions of the quadrupole magnets to improve the behavior of the beam. We note that although an earlier analysis (Lee et al. 2006) using a multivariate normal approximation did well at quantifying the primary mode of the density, that parametric assumption did not allow discovery of the smaller secondary modes in the initial density.

We also note that fitting by MCMC provides a full posterior for the initial distributions, so that uncertainty can be propagated in planning for future analysis of the beam. Figure 9 shows three realizations from the posterior, which gives some idea of the variability in the posterior and the uncertainty in our results.

6. CONCLUSIONS

Inverse problems are often challenging, because no data are directly observed on the process for which inference is to be drawn. In such cases, one must strike a balance between using an oversimplified yet fully identifiable model, and using a fully flexible but underidentified model. For our proton accelerator application, we have found that traditional density estimation techniques provide more flexibility than our limited data can allow us to fit. Instead, we have presented a process convolution perspective to describe our new approach that retains much of its original nonparametric flexibility yet is still a model that can be fit from our available data.

Because the likelihood is not in closed form, uncertainty estimates are typically difficult to obtain for inverse problems. Taking the Bayesian approach provides a straightforward method of estimating our uncertainty. Such estimates are helpful for the scientists in establishing design conditions of future experiments.

[Received December 2005. Revised March 2007.]
REFERENCES


Figure 7. Results for the real dataset: The columns are the positions of the four wirescans. The top row shows the estimated posterior particle cloud distributions for $x$ dimension versus momentum. The second row shows the data ($\circ$), posterior mean (---), and posterior 95\% interval estimates (---). The third and fourth rows are the analogous plots for the $y$ dimension.

Figure 8. Estimated input phase spaces for the real dataset.
Figure 9. Posterior realizations of phase spaces.