

Inferring Particle Distribution in a Proton Accelerator Experiment

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

Abstract. A beam of protons is produced by a linear charged particle accelerator, then focused through the use of successive quadrupoles. The initial state of the beam is unknown, in terms of particle position and momentum. Wire scans provide the only available data on the current state of the beam as it passes through and beyond the focusing region; the goal is to infer the initial state from these position histograms. This setup is that of an inverse problem, in which a computer simulator is used to link an initial state configuration to observable values (wire scans), and then inference is performed for the distribution of the initial state. Our Bayesian approach allows estimation of uncertainty in our initial distributions and beam predictions.

Keywords: computer simulator, inverse problem, exponentially-dampened cosine correlation

1 Introduction

Particle accelerators are used in a variety of experiments in physics. For an accelerator to be useful, it is important to understand exactly what the accelerator is producing. First, the particle beam emitted from the accelerator must be focused, so that it can be directed to the region of interest. The focusing process depends on the initial state of the beam. Second, information about the emitted particles may be critical in future calculations of the experiment. Thus, the statistical problem of interest is that of inferring the initial distribution (position and momentum) of the particles when they are first emitted from the accelerator.

The challenge of the problem arises because it is difficult to directly measure information about the particles. What can be observed are one-dimensional histograms of particle frequencies at various points along the path of the beam. Measurements are taken as the beam passes through a series of focusing quadrupole magnets. A computer simulator can be used to link an initial distribution state to future spatial location distributions. We are thus faced with a classic inverse problem, in that we are trying to learn about the unobservable initial state from highly transformed and simplified data,

*Department of Applied Mathematics and Statistics at the University of California, Santa Cruz, CA, <http://www.ams.ucsc.edu/~herbie/>

†Department of Applied Mathematics and Statistics at the University of California, Santa Cruz, CA, <http://www.ams.ucsc.edu/~bruno/>

‡Department of Applied Mathematics and Statistics at the University of California, Santa Cruz, CA, <http://www.ams.ucsc.edu/~zhouwn/>

§Division of Statistical Sciences, Los Alamos National Laboratory, Los Alamos, NM, <http://www.stat.lanl.gov/people/dhigdon.shtml>

with computer code providing the link (see for example, [Yeh 1986](#)). Proposed initial states can be run through the simulator, the predicted results computed, and then the initial proposal can be modified in an attempt to better match the computed results and the observed data. This process is iterated until convergence.

We take a Bayesian approach as it allows better accounting of uncertainty, particularly in the context of computer experiments and inverse problems (c. f. [Kennedy and O'Hagan 2001](#), where in addition to finding the calibration parameters, they also attempt to model the computer simulator). In many inverse problems, the problem is underspecified, in that many initial states will be able to produce similar fits for the data. Thus it is helpful for the statistician to produce a range of highly plausible initial states, which can be done naturally through the Bayesian paradigm by reporting posterior distributions or intervals.

In the next section we describe the physical experiment, along with the data that are collected. The following section discusses our statistical model for this problem, which accounts for some interesting features in the data. We then present some results, and conclude with some comments and future directions.

2 Physical Setup and Data

The LEDA accelerator is a linear accelerator that produces a beam of protons. The exact composition of the beam is not known, so the goal of this experiment is to infer the composition of the beam. The states of the particles in the beam are determined by their cross-sectional distributions (in the x and y directions) of position and momentum, denoted (x, p_x, y, p_y) . The x and y dimensions are treated as independent. However the position and momentum are expected to be correlated within a dimension, as discussed below.

With the initial distributions of position and momentum, the future paths of the particles can be predicted reasonably accurately with physical models (employing, for example, the Vlasov equation and the Poisson equation) for the particle movement as well as accounting for external forces on the particles (the focusing magnets that will be described shortly) and the inter-particle Coulomb field ([Dragt et al. 1988](#)). We are working with computer code (MLI 5.0) supplied by Los Alamos that simulates the particle paths via numerical solutions of the differential equations generated by the physical system ([Qiang et al. 2000](#)). A typical high fidelity run on a SunBlade 1000 workstation takes about six minutes; a lower fidelity run (with only 8,000 particles instead of 100,000) runs in about two minutes.

In order for this beam to be useful, it must be further focused, which is done with a series of magnets called *quadrupoles*. Magnets are used in sets of pairs, with the first one focusing the beam in the y direction, but de-focusing in the x direction. The second pair focuses x but de-focuses y . Through iterative focusing and de-focusing, the beam is gradually sharpened in both directions. [Figure 1](#) shows a simulation (via the computer code) of the 5-th, 15-th, . . . , 95-th percentiles of the positions of particles in the beam

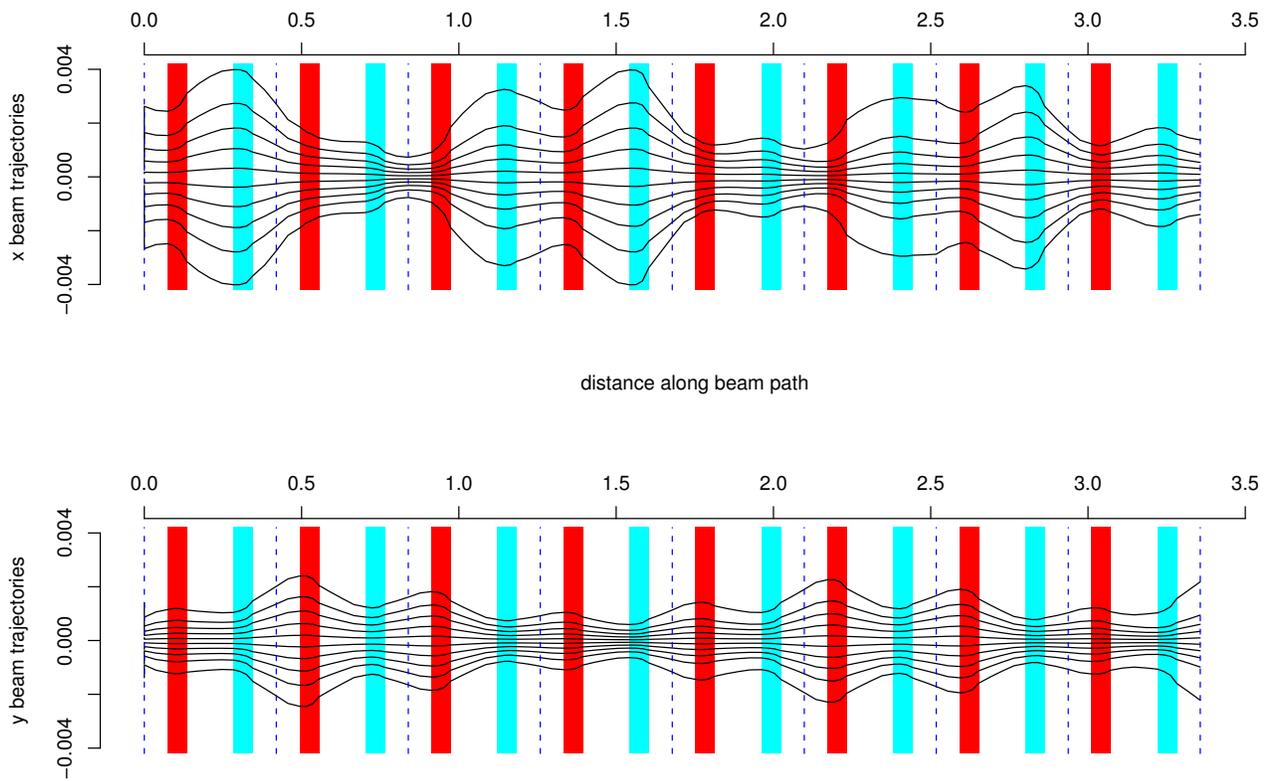


Figure 1: Simulation of particle beamlines passing through a series of quadrupole magnets. The upper panel corresponds to the progression of the particles in the x dimension; the lower panel is the y dimension. Magnets are denoted by shaded areas, with alternating magnets focusing and de-focusing in each direction. The wires are denoted by dashed lines.

as they pass through eight pairs of magnets. The y focusing magnets are shown in red (dark gray in grayscale), x focusing in blue (light gray). Notice how the beam widens in the x dimension as it passes through the red magnets and narrows passing through the blue ones, with the reverse effect for the beam in the y dimension.

Note that this focusing affects the beam. In particular, it induces a negative correlation between the position and momentum—i.e., it causes particles further away from the center to be pushed back toward the center, with the force being a function of the distance. This is a consequence of Maxwell’s equations for electromagnetic fields. In contrast, de-focusing will induce a positive correlation.

The dashed vertical lines in Figure 1 represent wirescans. As the beam crosses these lines, electrical current is produced and a histogram (here with 256 bins, which makes it appear sort of like a curve) of particle positions can be created. It is these nine wirescans in each dimension that are the observable data. Figure 2 shows some additional views. The top row are the first, third, fifth, seventh, and ninth (final) wirescans in the x direction, with the bottom row showing the y wirescans. For illustration, particle clouds are also shown in Figure 2. The second row shows the x -position versus x -momentum distribution at these five wirescan locations, and the fourth row are the corresponding distributions for the y position and momentum. The middle row shows the joint x and y positions. Notice that these are independent.

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, where the trajectories are shown as 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 shows the highly peaked histogram for y position. After passing through two sets of magnets, the second column of Figure 2 shows the wirescans and particle clouds at the third wirescan (third dashed line from the left in each panel of Figure 1). Now the x positions are relatively tight while the y distribution is wider. Note also the non-linear behavior of the particle clouds, showing intriguing relationships between position and momentum in each direction.

In practice, we will only be able to observe wirescans, and not any of the trajectories or particle clouds, which we have been able to plot here by running the computer code on simulated data. Furthermore, the Heisenberg uncertainty principle declares that it is not possible to measure both the position and momentum of a particle. Thus we must make do with just the series of wirescan position histograms (the top and bottom rows of Figure 2) and attempt to work backwards by combining this information with the particle simulator to infer the initial distribution statistically.

3 Statistical Model

We start our analysis by performing a simulation study using a high fidelity simulated beam of 100,000 particles as a proxy for a real particle accelerator beam. We use the simulator with a much lower fidelity beam of 8,000 particles to explore the known initial distribution of the high fidelity beam. We configure the simulator to have nine wires

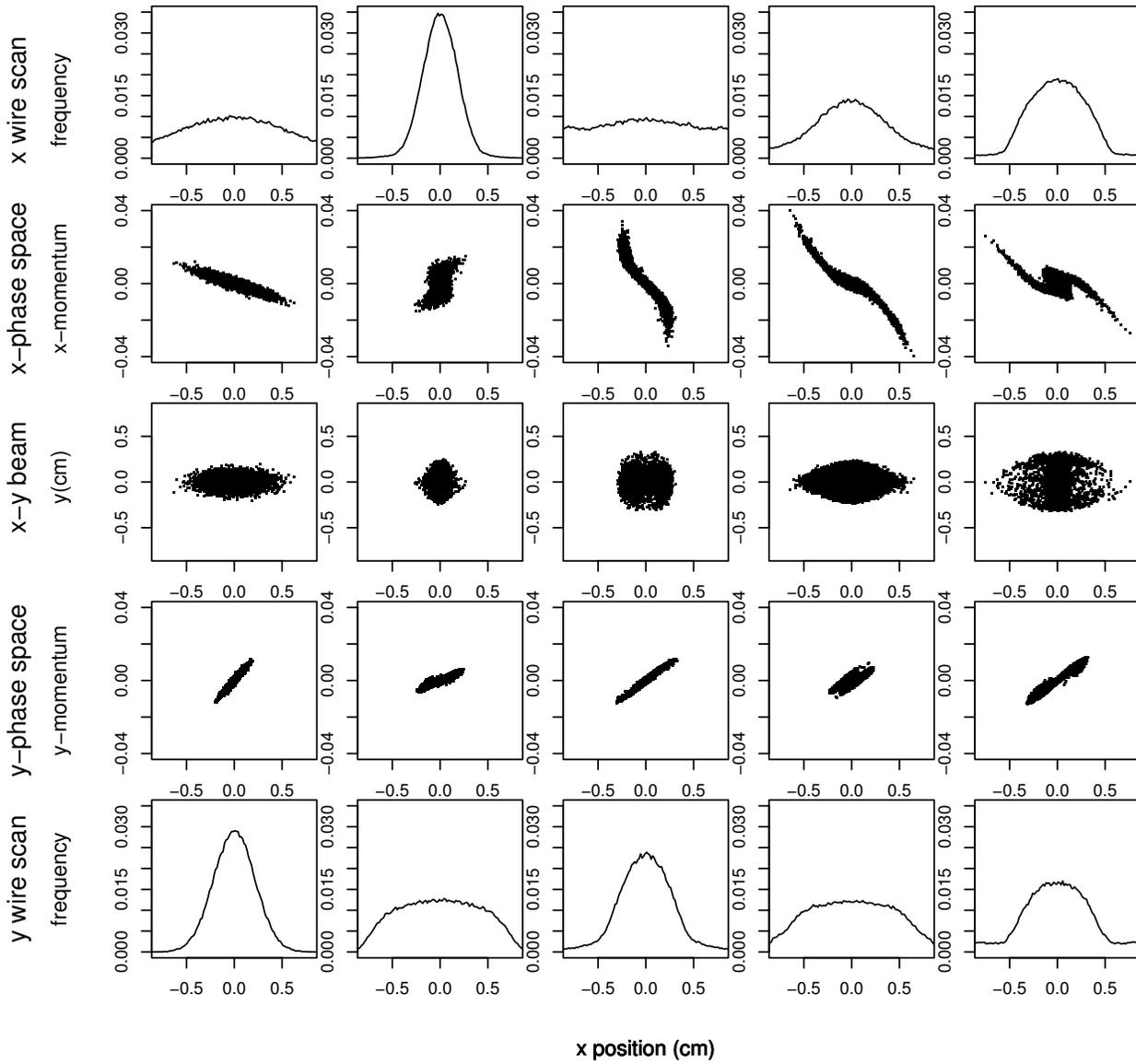


Figure 2: Simulation of wirescans and particle clouds at the locations of wirescans 1, 3, 5, 7, and 9 for the same experiment as in Figure 1. Top row is the x wirescans, second row is x -momentum vs. position, third row is y -position vs. x -position, fourth row is y -momentum vs. position, and bottom row is the y wirescans.

with 256 bins each. 8,000 particles were chosen since such number provides reasonable approximations to the high fidelity beam, yet low enough computational times to make MCMC feasible. We use a family of Gaussian distributions to parameterize the initial distribution for the cloud of particles. The likelihood is based on a modified squared-error loss, in that we use a correlated Gaussian structure for the discrepancies between the high fidelity wirescans and the low fidelity ones. We expect errors in nearby bins to be dependent as explained below.

We first validated our model using the high fidelity simulations. Next we considered data obtained from actual readings of four wirescans in a particle accelerator.

3.1 Probability model for the initial particle clouds

We model x position and momentum as bivariate normal and y position and momentum as bivariate normal independently of x . Thus,

$$x, p_x, y, p_y \sim N_4 \left(\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{bmatrix} \sigma_x^2 & \rho_x \sigma_x \sigma_{p_x} & 0 & 0 \\ \rho_x \sigma_x \sigma_{p_x} & \sigma_{p_x}^2 & 0 & 0 \\ 0 & 0 & \sigma_y^2 & \rho_y \sigma_y \sigma_{p_y} \\ 0 & 0 & \rho_y \sigma_y \sigma_{p_y} & \sigma_{p_y}^2 \end{bmatrix} \right),$$

and so the initial configuration is described by parameters: σ_x , σ_{p_x} , σ_y , σ_{p_y} , ρ_x and ρ_y . Our prior specifications are fairly vague, in fact, we assume that

$$\frac{1}{\sigma_x^2}, \frac{1}{\sigma_y^2} \sim \text{Exp} \left(\frac{1}{100} \right); \quad \frac{1}{\sigma_{p_x}^2}, \frac{1}{\sigma_{p_y}^2} \sim \text{Exp} \left(\frac{1}{1000} \right); \quad \rho_x \sim U(0, 1) \quad \text{and} \quad \rho_y \sim U(-1, 0).$$

We note that ρ_x and ρ_y must have opposite signs because of physical constraints as discussed previously.

3.2 Correlation structure of the wirescans

Inverse problems typically require that the likelihood function be fully specified, since there is insufficient information in the data to estimate both the initial configuration and additional parameters in the likelihood (see, for example, [Oliver et al. 1997](#); [Lee et al. 2002](#)). In order to specify our likelihood, we used the errors between the readings at the nine wirescans produced by the high fidelity and the low fidelity simulations, using the same initial configuration. The errors in bin counts are not independent, as nearby bins are correlated. Heuristically, the scientists expect ‘‘blobs’’ of mass in the initial distribution (e.g., ellipsoid contours). Such a structure leads to a relatively smooth wirescan, and so if one bin is too large, we also expect the nearby bins to be too large. As the frequencies must total to one, we also expect negative correlation at larger distances. We obtained empirical correlograms and observed a distinctive sinusoidal decay. Thus we model the error correlation between bins with an exponentially-dampened cosine function. For each wirescan j , the correlation function is defined by parameters λ_j and

ω_j . Let d be the distance between two bins, then

$$\rho_j(d) = e^{-3d/\lambda_j} \cos(\omega_j d). \quad (1)$$

This defines a valid positive definite covariance, as shown in [Yaglom \(1986\)](#). [Figure 3](#) shows the fits obtained for some of the wirescans using ρ_j for both x and y dimensions.

The likelihood is obtained by assuming that the errors of the wirescans for each dimension are conditionally normal. Let Σ_{x_j} and Σ_{y_j} denote the covariances corresponding to the j -th wirescans in the x and y dimensions respectively. These are obtained using the correlation in (1). Denote the corresponding errors as e_{x_j} and e_{y_j} . Then the likelihood is proportional to

$$\prod_{j=1}^9 (2\pi\tau^2)^{-2*256/2} \exp \left\{ -\frac{1}{2\tau^2} (e'_{x_j} \Sigma_{x_j}^{-1} e_{x_j} + e'_{y_j} \Sigma_{y_j}^{-1} e_{y_j}) \right\}.$$

We assume that τ^2 , Σ_{x_j} and Σ_{y_j} , $j = 1, \dots, 9$ are known. The Σ s were determined from simulation experiments; τ^2 requires additional attention. First we note that in many inverse problems, such as this one, there is not enough information to fully estimate all of the parameters and here τ^2 is a problematic one (see, for example, [Oliver et al. 1997](#); [Lee et al. 2002](#)). Ideally it will be chosen based on considerations about the expected size of discrepancies in the wirescans, relying on knowledge from subject area experts in a real application. In practice, the choice of τ^2 represents a trade-off between (i) the closeness of the fitted wirescans to the observed ones and (ii) the convergence properties of the MCMC sampler. Smaller values of τ^2 will produce better matches between the fitted and observed wirescans. The problem is that if τ^2 is too small, it can be nearly impossible to get the MCMC sampler to fully explore the posterior space, and the chain will typically fail to find a good fit in a reasonable amount of time (i.e., not getting anywhere near the posterior mode in two weeks). Without starting MCMC very close to the posterior mode (which would not be known ahead of time in practice), the chain will fail to converge properly when τ^2 is too small. Thus τ^2 must be chosen to be large enough for the chain to effectively explore the parameter space, yet not too large that the discrepancies in the fits are unbearable. Our choice here is a practical compromise.

3.3 MCMC

We use a Metropolis-Hastings algorithm (see for example, [Gelman 1997](#)) to explore the distributions of the six parameters that define the distribution of the initial configuration. We sample the parameters in two blocks, one for the variances and correlation of the x dimension and another for those of the y dimension. To produce proposals for the variances we use random walks on the log scale. Proposals for the correlation parameters are obtained with constrained random walks. The x dimension correlation is constrained to the interval $(0, 1)$, while the y dimension correlation is constrained to $(-1, 0)$, since the physics of the problem requires that the correlations have opposite signs (as discussed in [Section 2](#)).

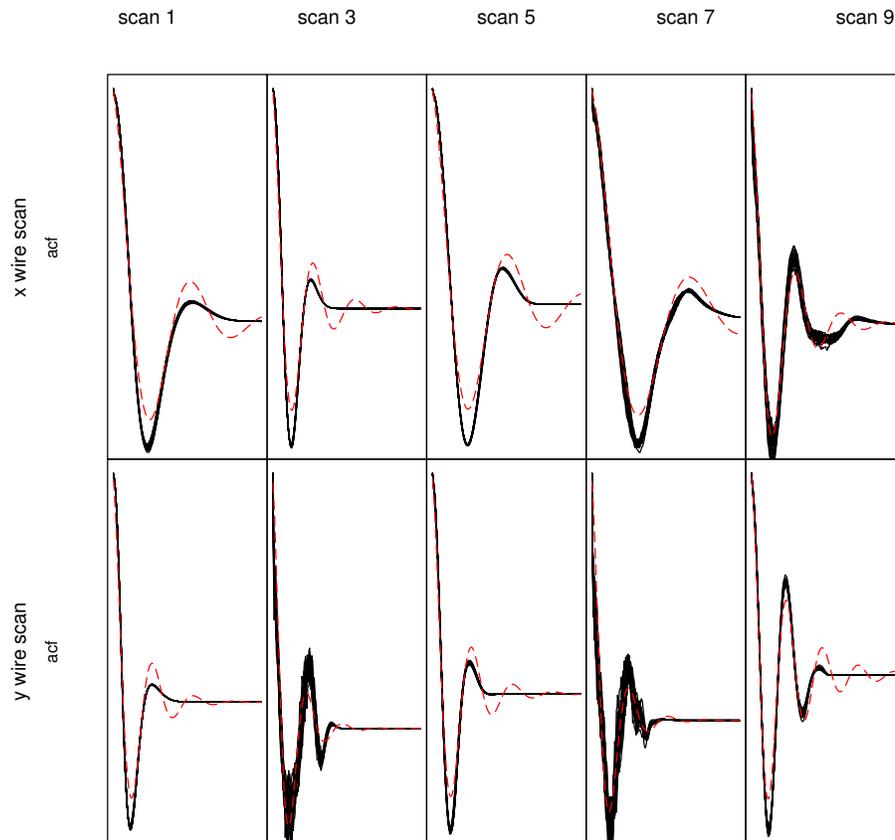


Figure 3: Autocorrelation function plots for the differences between wire scans of high and low fidelity simulated beams. Only the odd numbered wire scans are shown. The dotted lines correspond to the empirical autocorrelations. The continuous line corresponds to the least squares fit using the correlation function defined in (1).

3.4 Simulation Results

Figure 4 shows our results on the simulated data. Each column represents a wire. There were actually nine wirescans, but only four are shown to improve the visibility of the graphs; the other five are similar in character. The top two rows are for the x dimension, the bottom two rows for the y dimension. The first and third rows are the estimated posterior distributions of the particle cloud (x or y position and momentum) as the beam passes through the quadrupole magnets. The second and fourth rows show the true wirescans (circles), the estimated posterior mean scans (solid line), and posterior interval estimates (dashed lines). Only 32 bins are pictured so that the results can be seen visually (using all of them produces a smear of black ink). This number was chosen to match the number of bins used in the real data example of the next section. The posterior mean is generally close to the truth, and the credible intervals provide a measure of our uncertainty.

Figure 5 shows the estimated posterior distribution for the six parameters of our inverse problem (the variance of position and momentum for each of x and y , and the correlation between the position and momentum for each of x and y), along with the true values (the large black dots). We are pleased that the estimated posterior has most of its mass near the truth for all six parameters.

We note that the fit is not perfect in some of the plots. This is partly a result of the choice of τ^2 , which represents a trade-off between goodness-of-fit and computational efficiency in convergence of the MCMC. Here we used a value of $\tau^2 = 0.0005$, which leaves the likelihood somewhat “loose” in the sense that the wirescans match most of the time but not all, as shown in Figure 4. The reason we do this is that each MCMC iteration requires a run of the simulator and that takes about two minutes. So a 3000-iteration MCMC run takes about five days on a SunBlade 1000 workstation. With this value of τ^2 , this number of iterations is sufficient for convergence of the chain from typical arbitrary starting values. When we try to use smaller values of τ^2 to decrease the error in the fitted values, we find that convergence of the chain decreases significantly. An order of magnitude decrease in τ^2 appears to increase the time to convergence by at least an order of magnitude, which makes the computational requirements impractical with the current technology. As computers get faster, we expect that our approach will be able to achieve better fidelity. For the present, we must accept some error in our fits in order to be able to produce results in a reasonable amount of time. In our goal of inferring the initial distribution, Figure 5 shows that we are doing a reasonable job in that regard.

4 Application to Real Data

We now apply our methodology to a real dataset provided by scientists at Los Alamos National Laboratory (Allen et al. 2002; Allen and Pattengale 2002). In this setting, only four wirescans are available, analogous to the first four scans in the simulated example above. From these four scans we attempt to infer the unknown initial distribution of the beam. The data seem a bit more complex than can be perfectly matched under our

paired bivariate normal model, but the model captures the key features in the data. We note that for the real dataset, the correlation parameter for x is positive and y is negative, the reverse of our simulated example (the physical constraint is merely that the signs must be opposite), so we modify the relevant priors and proposal distributions accordingly.

Figure 6 shows our results on this dataset. As in Figure 4, each column is a wire; the top two rows are for x , the bottom two for y , with estimated particle cloud posteriors and fitted scans with 95% credible intervals for each. As with the simulated data, there is room for improvement in some of the fits, but the computational requirements make it difficult to achieve significant improvements. Our methodology provides a good starting point for solving this inverse problem. The bivariate normal assumption is useful for its intuitive simplicity yet it provides enough flexibility to do a reasonable job of modeling the true process.

5 Conclusions

The Bayesian approach is helpful in this problem as it gives a natural measure of uncertainty. Such an uncertainty estimate would be nearly impossible to obtain from a classical analysis, yet is valuable in understanding the functioning of the particle accelerator. As with many inverse problems, multiple initial conditions can be consistent with the observed data, and the Bayesian approach also provides a natural mechanism for either exploring this multimodal surface, or for restricting the problem through the imposition of structure in the prior based on substantive information (as we do here).

Extensions of the current model can be considered in several directions. One is the acceleration of computations and the other is exploring more complex distributions for the initial configuration. A typical MCMC run would take several days because of the time spent running the simulator at each iteration. To make the MCMC faster, we can consider a multiresolution approach where a very low fidelity simulator (faster but less accurate) is coupled with a high fidelity one (slower but more reliable); we expect that a multiresolution approach will also help improve the time necessary for convergence of the MCMC sampler. An alternative approach is that of replacing the current simulator with a simplified version that uses linear or non-linear approximations between subsequent positions of the accelerator, such as in [Craig et al. \(1996\)](#) or [O’Hagan et al. \(1999\)](#). Improvements to the initial configuration may be obtained by using a more flexible family of distributions, such as mixtures of normals or Gaussian processes.

References

- Allen, C. K., Chan, K. C. D., Colestock, P. L., Crandall, K. R., Garnett, R. W., Gilpatrick, J. D., Lysenko, W., Qiang, J., Schneider, J. D., Schulze, M. E., Sheffield, R. L., Smith, H. V., and Wangler, T. P. (2002). “Beam-halo measurements in high-current proton beams.” *Physical Review Letters*, 89(214802). 257

- Allen, C. K. and Pattengale, N. D. (2002). “Theory and technique of beam envelope simulation: Simulation of bunched particle beams with ellipsoidal symmetry and linear space charge forces.” Technical Report LA-UR-02-4979, Los Alamos National Laboratory. 257
- Craig, P. S., Goldstein, M., Seheult, A. H., and Smith, J. A. (1996). “Bayes Linear Strategies for History Matching of Hydrocarbon Reservoirs.” In Bernardo, J. M., Berger, J. O., Dawid, A. P., and Smith, A. F. M. (eds.), *Bayesian Statistics 5*, 69–95. Oxford: Clarendon Press. 258
- Dragt, A., Neri, F., Rangarjan, G., Douglas, D., Healy, L., and Ryne, R. (1988). “Lie Algebraic Treatment of Linear and Nonlinear Beam Dynamics.” *Annual Review of Nuclear and Particle Science*, 38: 455–496. 250
- Gamerman, D. (1997). *Markov Chain Monte Carlo*. London, UK: Chapman and Hall. 255
- Kennedy, M. C. and O’Hagan, A. (2001). “Bayesian Calibration of Computer Models.” *Journal of the Royal Statistical Society, Series B*, 63: 425–464. 250
- Lee, H., Higdon, D., Bi, Z., Ferreira, M., and West, M. (2002). “Markov Random Field Models for High-Dimensional Parameters in Simulations of Fluid Flow in Porous Media.” *Technometrics*, 44: 230–241. 254, 255
- O’Hagan, A., Kennedy, M. C., and Oakley, J. E. (1999). “Uncertainty Analysis and other Inference Tools for Complex Computer Codes.” In Bernardo, J. M., Berger, J. O., Dawid, A. P., and Smith, A. F. M. (eds.), *Bayesian Statistics 6*, 503–524. Oxford University Press. 258
- Oliver, D. S., Cunha, L. B., and Reynolds, A. C. (1997). “Markov Chain Monte Carlo Methods for Conditioning a Permeability Field to Pressure Data.” *Mathematical Geology*, 29(1): 61–91. 254, 255
- Qiang, J., Ryne, R. D., Habib, S., and Decyk, V. (2000). “An Object-Oriented Parallel Particle-In-Cell Code for Beam Dynamics Simulation in Linear Accelerators.” *Journal of Computational Physics*, 163: 434–451. 250
- Yaglom, A. (1986). *Correlation Theory of Stationary and Related Random Functions I: Basic Results*. Springer Series in Statistics. New York: Springer-Verlag. 255
- Yeh, W. W. (1986). “Review of Parameter Identification in Groundwater Hydrology: the Inverse Problem.” *Water Resources Research*, 22: 95–108. 250

Acknowledgments

Herbert Lee’s work was partially supported by grant DMS 0233710 from the National Science Foundation. Bruno Sansó’s work was partially supported by grant 74328-001-03 3D from Los Alamos National Laboratory. Weining Zhou’s work was partially supported by grant 74328-001-03 3D from Los Alamos National Laboratory. The authors would like to thank the editors,

two anonymous referees, Charles Nakhleh, and Vidya Kumar for their helpful comments and suggestions.

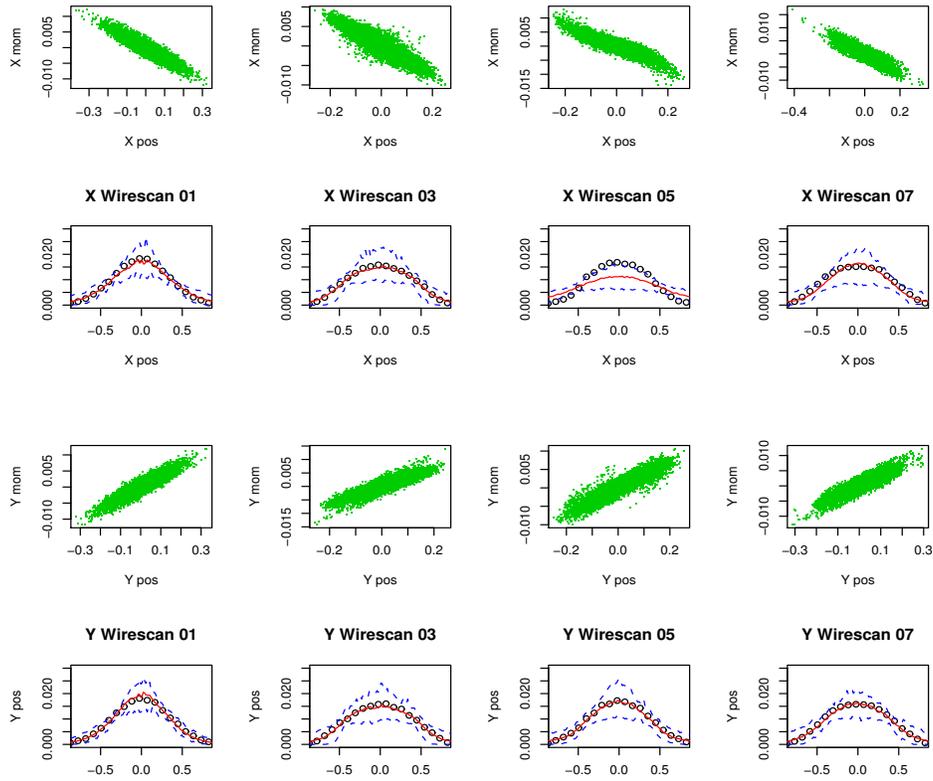


Figure 4: Results for the simulated dataset: the columns are the positions of the four wire-scans. The top row shows the estimated posterior particle cloud distributions for x position vs. momentum. The second row shows the data (circle), posterior mean (solid line), and posterior 95% interval estimates (dashed lines). The third and fourth rows are the analogous plots for the y dimension.

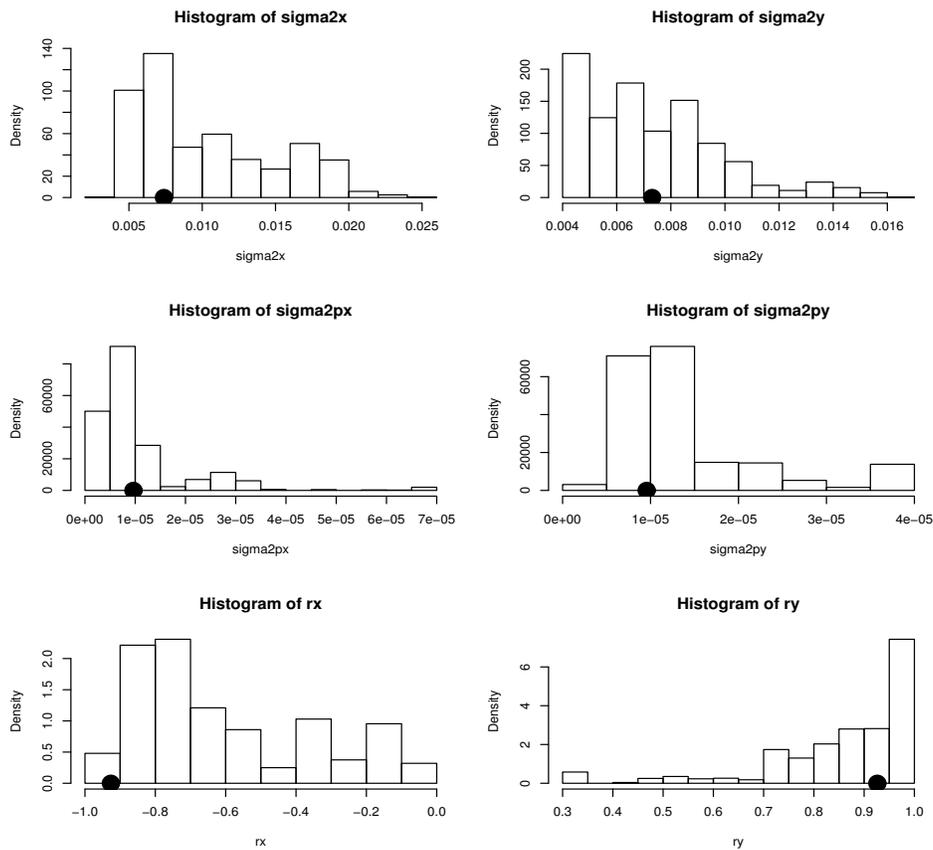


Figure 5: Posterior distribution estimates for the parameters for the simulated data set. The truth is shown as the large black dot.

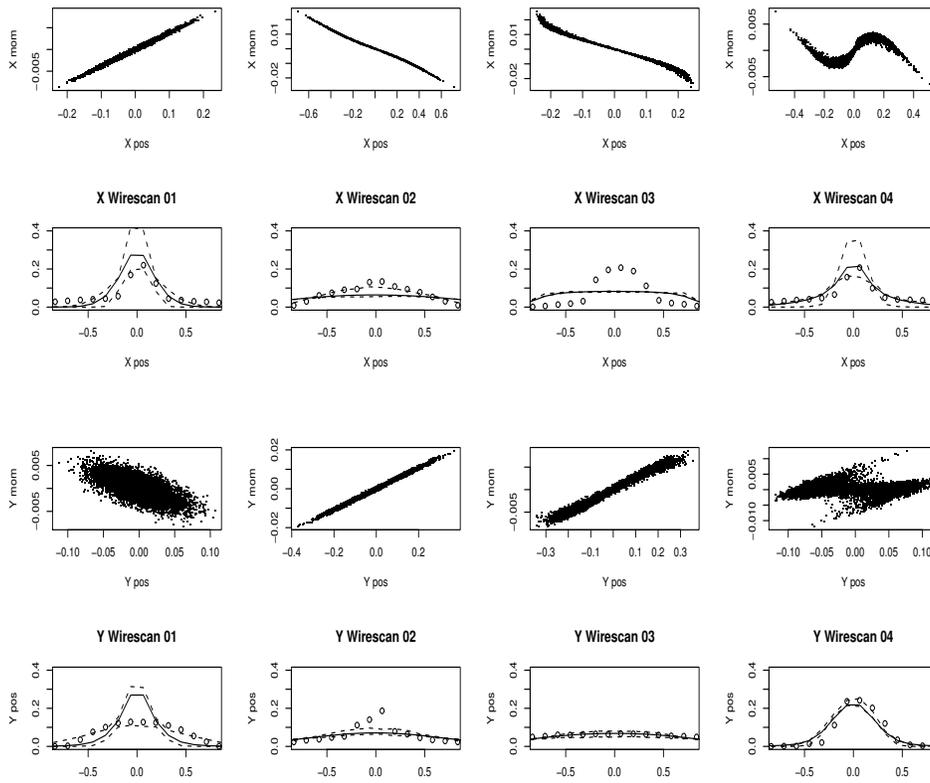


Figure 6: Results for the real dataset: the columns are the positions of the four wirescans. The top row shows the estimated posterior particle cloud distributions for x dimension vs. momentum. The second row shows the data (circle), posterior mean (solid line), and posterior 95% interval estimates (dashed lines). The third and fourth rows are the analogous plots for the y dimension.