Inter-annual variation in large-scale movement of Atlantic bluefin tuna (*Thunnus thynnus*) determined from pop-up satellite archival tags

Supplementary Material

John R. Sibert[1], Molly E. Lutcavage[2], Anders Nielsen[1], Richard W. Brill[3], and Steven G. Wilson[4]

April 15, 2006

[1]Joint Institute of Marine and Atmospheric Research, University of Hawai'i at Manoa, Honolulu, HI 96822, U.S.A.; sibert@hawaii.edu

[2]Department of Zoology, University of New Hampshire, Durham, NH, USA 03824

[3]Virginia Institute of Marine Science, Gloucester Point, VA, USA 23062

[4]Hubbs-SeaWorld Research Institute, 2595 Ingraham Street, San Diego, CA 92109, USA

1 Model description

The state-space extended Kalman filter model applied to the analysis of data from archival tags has been refined and extended since 2001 when Sibert *et al.* (2003) was submitted for publication. The model and extensions used in this paper are completely described here.

1.1 The basic model

The basic model for an observed track is a state–space model, where the state equation describes the movements of a fish in an axis–parallel plane. A biased random walk model is assumed:

$$\alpha_i = \alpha_{i-1} + c_i + \eta_i, \quad i = 1, \ldots, T \tag{1}$$

Here α_i is a two dimensional vector containing the coordinates at time t_i, c_i is the drift (or bias) vector describing the deterministic part of the movement, and η_i is the noise vector a describing the random part of the movement. The deterministic part of the movement is assumed to be proportional to time:

$$c_i = \begin{pmatrix} u\Delta t_i \\ v\Delta t_i \end{pmatrix} \tag{2}$$

The random part is assumed to be serially uncorrelated and follow a two dimensional Gaussian distribution with mean vector 0 and covariance matrix Q_i, where

$$Q_i = \begin{pmatrix} 2D\Delta t_i & 0 \\ 0 & 2D\Delta t_i \end{pmatrix}. \tag{3}$$

The measurement equation of the state–space model is a non–linear mapping of the coordinates on the axis–parallel plane to the sphere. The original coordinates are in Nautical miles and the coordinates on the sphere are in degrees of longitude and latitude. The measurement equation describing the actual measured position y_i is:

$$y_i = z(\alpha_i) + d_i + \varepsilon_i, \quad i = 1, \ldots, T \tag{4}$$

where z is the coordinate change function given by:

$$z(\alpha_i) = \begin{pmatrix} \frac{\alpha_{i,1}}{60\cos(\alpha_{i,2}\pi/180/60)} \\ \frac{\alpha_{i,2}}{60} \end{pmatrix} \tag{5}$$

d_i is the observation bias:

$$d_i = \begin{pmatrix} b_x \\ b_y \end{pmatrix} \tag{6}$$

and ε_i is the measurement error which is assumed to follow a Gaussian distribution with mean vector 0 and covariance matrix H_i, where

$$H_i = \begin{pmatrix} \sigma_x^2 & 0 \\ 0 & \sigma_{y_i}^2 \end{pmatrix} \tag{7}$$

The variance of the latitude measurements is closely related to the equinox. Measurements close to the equinox have large latitude errors. The following variance structure is assumed:

$$\sigma_{y_i}^2 = \sigma_{y0}^2 \Big/ \Big(\cos\left(2\pi(J_i + (-1)^{s_i} b_0)/365.25\right) + a_0 \Big) \tag{8}$$

where J_i is the number of days since last solstice prior to all observations, s_i is the season number since the beginning of the track (one for the first 182.625 days, then two for the next 182.625, then three and so on). a_0, b_0 and σ_{y0}^2 are model parameters.

1.2 The extended Kalman filter

As the model described in the previous section is non–linear, an approximation is needed to apply the Kalman filter. The extended Kalman filter simply uses a first order Taylor approximation around the optimal estimator $a_{i|i-1}$:

$$z(\alpha_i) \approx z(a_{i|i-1}) + \widehat{Z}_i \cdot (\alpha_i - a_{i|i-1}) \tag{9}$$

Here $\widehat{Z}_i$ is the first derivative (or the Jacobi matrix) of the function z, which is calculated as:

$$z'(\alpha_i) = \frac{\partial z(\alpha_i)}{\partial \alpha_i} = \begin{pmatrix} \frac{1}{60\cos(\alpha_{i,2}\cdot\pi/180/60)} & \frac{\alpha_{i,1}\pi\sin(\alpha_{i,2}\cdot\pi/180/60)}{180(60\cos(\alpha_{i,2}\cdot\pi/180/60))^2} \\ 0 & \frac{1}{60} \end{pmatrix} \tag{10}$$

The optimal estimator $a_{i|i-1}$ is inserted into the matrix in (10) to obtain $\widehat{Z}_i$. The updated Kalman filter equations are written as:

$$a_{i|i-1} = a_{i-1} + c_i \tag{11}$$
$$P_{i|i-1} = P_{i-1} + Q_i \tag{12}$$
$$F_i = \widehat{Z}_i P_{i|i-1} \widehat{Z}_i' + H_i \tag{13}$$
$$w_i = y_i - z(a_{i|i-1}) - d_i \tag{14}$$
$$a_i = a_{i|i-1} + P_{i|i-1} \widehat{Z}_i' F_i^{-1} w_i \tag{15}$$
$$P_i = P_{i|i-1} - P_{i|i-1} \widehat{Z}_i' F_i^{-1} \widehat{Z}_i P_{i|i-1} \tag{16}$$

The filter is started by calculating $a_0 = z^{-1}(y_0)$ and assuming this position to be known without error ($P_0 = 0_{2\times 2}$).

1.3 Pop–off model

If at any point during the observed track the tag came off the fish, the movement pattern would be expected to change instantaneously. The measurement equation (4) is unchanged, as it only describes how the true position of the tag is translated into the observed measurements. The movement pattern is described by the state equation (1) modified to accommodate variation in c_i and η_i along the track.

Consider a given pop–off time τ. If the tag came off the fish after τ days at liberty, the parameters describing the movement of the tag (u, v and D) would be expected to be different while the tag was on the fish, than while the tag was simply floating by itself. Denote the parameters before pop–off as u_1, v_1 and D_1, and after pop–off as u_2, v_2 and D_2.

Observations are only available at discrete time points, so the movement between two observations may be partly before the pop–off and partly after. As a consequence the terms c_i and η_i should be calculated as a weighted average. For instance the first coordinate of the drift term c_i should be $\widetilde{u}_i \Delta t_i$, where

$$\widetilde{u}_i = \mathsf{wave}(u_1, u_2, t_{i-1}, t_i, \tau) \stackrel{\mathsf{def.}}{=} \begin{cases} u_1 & , \text{ if } \tau \geq t_i \\ \frac{\tau - t_{i-1}}{\Delta t_i} u_1 + \frac{t_i - \tau}{\Delta t_i} u_2 & , \text{ if } t_{i-1} < \tau < t_i \\ u_2 & , \text{ if } \tau \leq t_{i-1} \end{cases} \tag{17}$$

Following this example the second coordinate of c_i should be $\widetilde{v_i}\Delta t_i$, and the diagonal elements of the covariance matrix Q_i of η_i should be $2\widetilde{D_i}\Delta t_i$, where

$$\widetilde{v_i} = \mathsf{wave}(v_1, v_2, t_{i-1}, t_i, \tau) \quad \& \quad \widetilde{D_i} = \mathsf{wave}(D_1, D_2, t_{i-1}, t_i, \tau). \tag{18}$$

The basic model is clearly a sub–model of the pop–off model, as it correspond to the case where $u_1 = u_2$, $v_1 = v_2$ and $D_1 = D_2$. Compared to the basic model, the pop-off model has four additional parameters (τ and the three extra movement parameters).

In the actual implementation of the pop–off model a smooth (differentiable) version of the wave function was used. This modification allows for automatic differentiation of the model and the approximation error is negligible.

A smooth (differentiable) approximation of the wave function defined in (17) is found in the following way. The wave function can also be defined as:

$$\widetilde{u_i} = \mathsf{wave}(u_1, u_2, t_{i-1}, t_i, \tau) = \frac{1}{\Delta t_i} \int_{t_{i-1}}^{t_i} 1_{\{t<\tau\}} u_1 + 1_{\{t\geq\tau\}} u_2 \; dt \tag{19}$$

This view on the weighted average function inspired the approximation. The smooth weighted average function swave is defined as:

$$\widetilde{u_i} \approx \mathsf{swave}(u_1, u_2, t_{i-1}, t_i, \tau) \stackrel{\mathsf{def.}}{=} \frac{1}{\Delta t_i} \int_{t_{i-1}}^{t_i} (u_2 - u_1) \left(\frac{\mathrm{atan}(S(t-\tau))}{\pi} + \frac{1}{2} \right) + u_1 \; dt \tag{20}$$

$$= \frac{u_2 - u_1}{\pi \Delta t_i S} \Bigg(S(t_i - \tau)\mathrm{atan}(S(t_i - \tau)) - \frac{1}{2}\log(S^2(t_i - \tau)^2 + 1) - \tag{21}$$

$$S(t_i - \tau)\mathrm{atan}(S(t_i - \tau)) + \frac{1}{2}\log(S^2(t_i - \tau)^2 + 1) \Bigg) + \frac{u_1 + u_2}{2}$$

Here S is the scaling constant. Large values of S gives a close approximation of the weighted average function wave. In the implementation a value of $S = 100$ was used, which gives a very close approximation.

1.4 Maximum likelihood principle

The negative log–likelihood function for the Kalman filter is given by (Harvey, 1990):

$$\ell(\theta) = -\log L(Y|\theta) = T\log(2\pi) + 0.5\sum_{i=1}^{T}\log(|F_i|) + 0.5\sum_{i=1}^{T} w_i' F_i^{-1} w_i. \quad (22)$$

The estimated values of the model parameters are found by minimizing the negative log–likelihood function as a function of the model parameters.

$$\widehat{\theta} = \underset{\theta\in\Theta}{\operatorname{argmin}}\ \ell(\theta) \quad (23)$$

where Θ is the domain of the model parameters expressing upper and lower bounds.

Let M_1 denote the pop–off model and M_2 denote the basic model. The model parameters of model M_1 are $\theta_1 = (u_1, u_2, v_1, v_2, D_1, D_2, \tau, b_x, b_y, \sigma_x, \sigma_y, b_0)$, and the model parameters of model M_2 are $\theta_2 = (u, v, D, b_x, b_y, \sigma_x, \sigma_y, b_0)$. The standard likelihood–ratio test can be used for any fixed τ to compare the two models M_1 and M_2, where M_2 is a sub–model of M_1 The test is calculated as:

$$R_{M_1\to M_2} = \frac{L_{M_2}(Y|\theta_2)}{L_{M_1}(Y|\theta_1)} \quad (24)$$

$$\text{P-value}_{M_1\to M_2} = P\left(\chi_4^2 \geq -2\log R_{M_1\to M_2}\right) \quad (25)$$

The four degrees of freedom is calculated as the difference between the number of parameters in the two models.

The P-value computed by equation 25 is only exact if the value of τ is fixed. Our method searches for the value of τ which maximizes the likelihood ratio, and thus the P-value is not an accurate measure of the desired probability. Nevertheless, the likelihood ratio is a sensitive measure of deviance and a valid criterion for estimating τ. Nielsen (2004) presents a computer intensive method for estimating a more exact P-value.

1.5 Locating the pop–off position τ

The most probable pop–off position is the maximum likelihood estimate of the model parameter τ in the pop–off model. $\widehat{\tau}$ denotes the maximum likelihood estimate of τ. $\widehat{\tau}$ describes the optimal position to divide the movement

model of the track into two, or in other words the 'cut–position' resulting in the highest likelihood.

It proved difficult to do a standard minimization of the negative log likelihood ℓ corresponding to the pop–off model, as random fluctuations of the track can result in many local minima. To ensure that a global minimum is found, the following 'brute force' algorithm was used.

Initial scan: For a number (20 or so) of different *fixed* values of τ covering the entire track, the negative log likelihood corresponding to the pop–off model is minimized by estimating the remaining parameters u_1, u_2, v_1, v_2, D_1, D_2, b_x, b_y, σ_x, σ_y and b_0.

Final minimization: Starting from the pop–off position τ where the initial scan returned the smallest negative log likelihood value, a final minimization of the negative log likelihood corresponding to the entire pop–off model is done. In this minimization the parameters u_1, u_2, v_1, v_2, D_1, D_2, b_x, b_y, σ_x, σ_y, b_0 and τ are included.

If enough fixed values of τ is included in the initial search, this algorithm will return the global minimum.

1.6 The most probable track

The Kalman filter and the maximum likelihood principle supply estimates of the model parameters and the predicted track. A point on the predicted track at any given time point is calculated using all observations available at that time, that is $a_i = E(\alpha_i|y_1, \ldots, y_i)$. Intuitively, it seems reasonable that better estimates can be produced, once the entire track is known. The most probable track is calculated using all observations, and after the parameters has been estimated. A point on the most probable track is $a_{i|T} = E(\alpha_i|y_1, \ldots, y_T)$.

The actual computation of the most probable track is done in a single backwards updating sweep of the predicted track. The last point of the most probable track is identical to the last point of the predicted track, as all observations were available to the predicted track at the final point. The last point of the most probable track is known $a_{T|T} = a_T$ and the prediction error

of the last point $P_{T|T} = P_T$ The following equations are used to compute the previous points of the most probable track.

$$a_{i|T} = a_i + P_i^\star \cdot (a_{i+1|T} - a_i - c_{i-1}) \tag{26}$$

$$P_{i|T} = P_i + P_i^\star \cdot (P_{i+1|T} - P_{i+1|i}) P_i^{\star\prime} \ , \quad \text{where} \quad P_i^\star = P_i P_{i+1|i}^{-1} \tag{27}$$

This technique is known as smoothing in textbooks on the Kalman filter, as the resulting track most often is smoother than the predicted track.

1.7 Last point fixed

If the last position is known without error, or with an error term which is negligible compared to the other observations, it can be used as a fixed point. This corresponds to setting the measurement error and the measurement bias to zero for the last observation ($H_T = 0_{2\times 2}$ and $d_T = 0_{2\times 1}$). This results in a predicted (and most probable) track ending up in the fixed pop–off location. As the last point is now assumed known without error, the number of observations T in the likelihood function (22) should now be replaced by $T - 1$.

1.8 Multi–segment model

The pop-off model (1.3) is easily extended to a general notion of change in apparent behavior. The change in behavior from one movement pattern, characterized by distinct values of the movement parameters (u, v, D), to another is signaled by some event. In the pop-off model, the event is the instant when the time at liberty exceeds the pop-off time, τ. Movement from one geographic region to another is an alternative way to specify the behavior change event. The multi-segment model defines a number of geographic regions consisting of closed polygons defined by a series of longitude and latitude pairs. Let R_j be a $2 \times n_j$ matrix containing n_j pairs of arbitrarily specified geographic coordinates defining a region j. In a manner analogous to equations 17 and 18, the terms of c_i and η_i are defined according to ther current estimate of the tag position a_i (equation 15), by:

$$a_i \in R_j \begin{cases} \widetilde{u_i} = u_j \\ \widetilde{v_i} = v_j \\ \widetilde{D_i} = D_j \end{cases} \tag{28}$$

where (by an abuse of notation) $a_i \in R_j$ means that the point a_i lies within the polygon defined by R_j. This parameterization of regional variation in movement is similar to that used by Sibert *et al.* (1999) in large scale diffusion models.

1.9 Multi–track model

The multi-track model is a simple extension of the general model in which the likelihood is maximized over all tracks with a common set of estimated parameters. That is the likelihood function $\ell(\theta)$ in equation 22 is replaced by

$$\ell(\theta) = -\log L(Y|\theta) = \sum_{k=1}^{m} \left\{ T_k \log(2\pi) + 0.5 \sum_{i=1}^{T_k} \log(|F_{ki}|) + 0.5 \sum_{i=1}^{T_k} w'_{ki} F_{ki}^{-1} w_{ki} \right\}. \tag{29}$$

where m is the number of tracks.

2 Markov Chain Monte Carlo Simulation

The state-space Kalman filter model parameters are normally estimated by quasi-Newton numerical minimization of the extended Kalman filter likelihood function. The minimization algorithm occasionally failed to converge to a solution for multi-track models, possibly due to multiple local minima. Further, the introduction of spatially defined multi-segment tracks causes problems of non-differentiability in the likelihood function. Markov chain Monte Carlo (MCMC) optimization was used to estimate approximate probability distributions of the model parameters for the multi-track and two-region KF variants. Although MCMC is often associated with Bayesian analysis, the method is general and has wider applications in optimization (Geyer, 1996; Gelman, *et al.* 2004, chapter 11). The state space Kalman filter model is implemented in ADModel Builder (Otter Research Ltd.) which uses the Metropolis-Hastings algorithm for MCMC simulation.

We ran 500,000 simulations for each model and sampled the parameter values after every 100 simulations. Inspection of the sampled Markov chains indicates whether the algorithm has found parameter values that produce a relatively stationary likelihood. The mode of the sampled Markov chains were interpreted as point estimates of model parameters.

In the following, two figures are shown for every set of combined tracks. The movement parameters (u, v, D) for the offshore areas are shown in the thick gray (light blue) lines. The Markov chains are shown by plotting the parameter value against the sample number and by plots of the empirical likelihood histograms for each estimated parameter. In both cases, the ranges of the parameter values are scaled to the extremes of last 4,000 samples of the Markov chains.

The MCMC converged to stationary values within the first 10,000 simulations (100 samples) for all three sets of tracks from combining the 1999, 2000 and 2002 deployments separately. However some of the combinations of tracks which combine the 2000 required longer simulation to converge to a solution. For example, the Markov chains did not reach stationary values until more than 50,000 simulations had been conducted when the 2000 deployments were combined with the 1999 deployments, Figures 7 and 8, suggest that the parameters are not well defined.

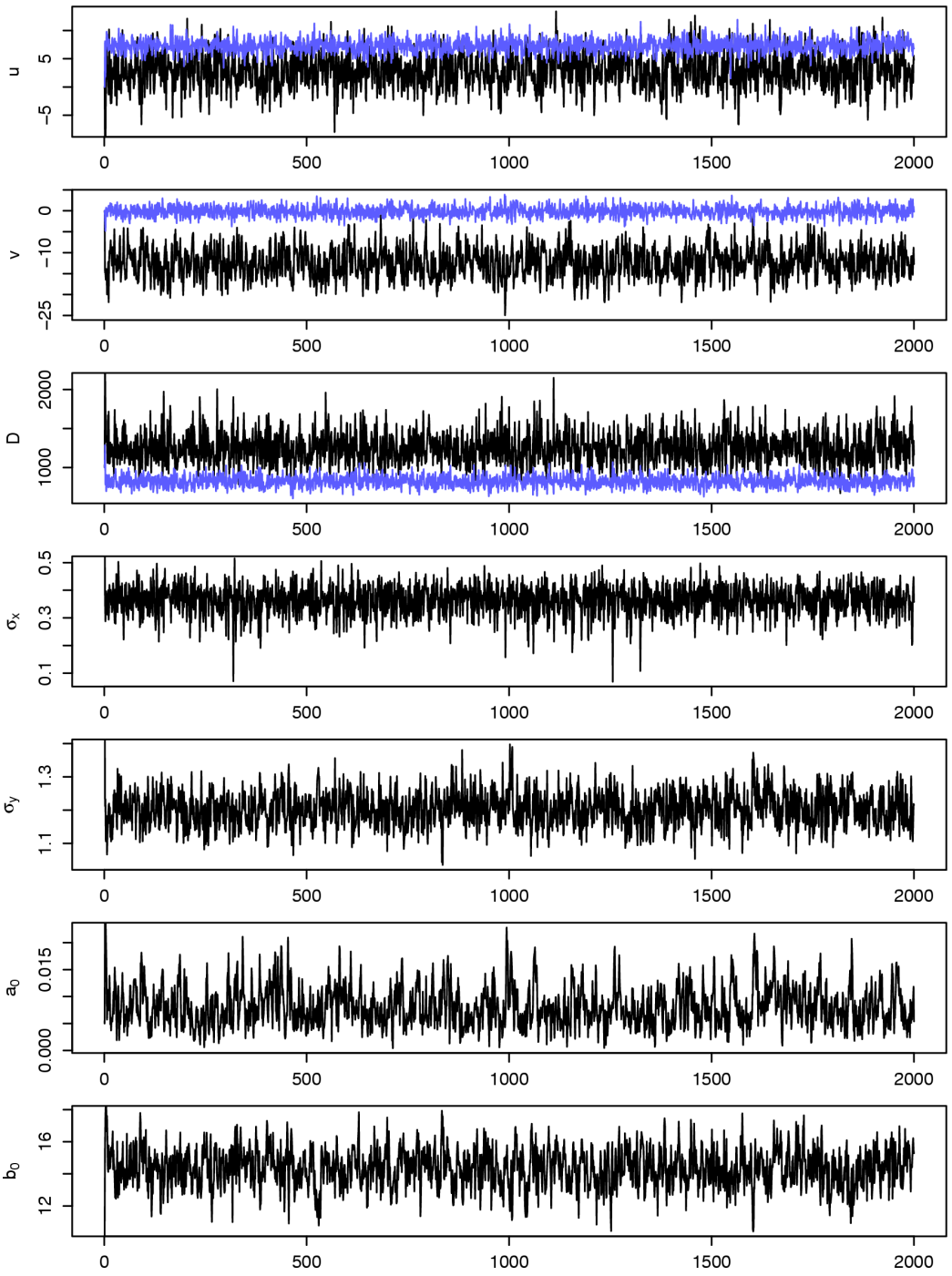

Figure 1: Samples of MCMC chain for 1999 tag releases.

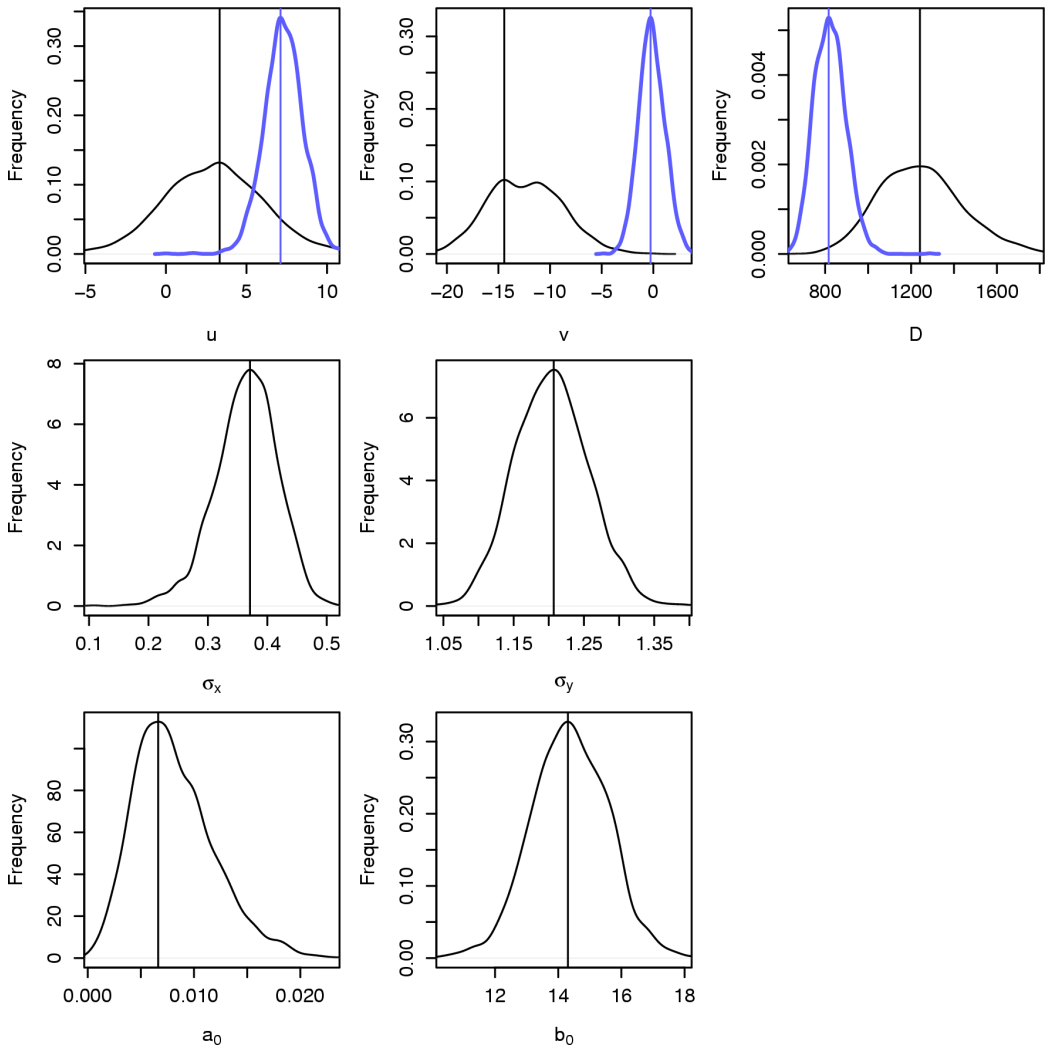

Figure 2: Frequency distribution of parameters from MCMC simulations for 1999 tag releases.

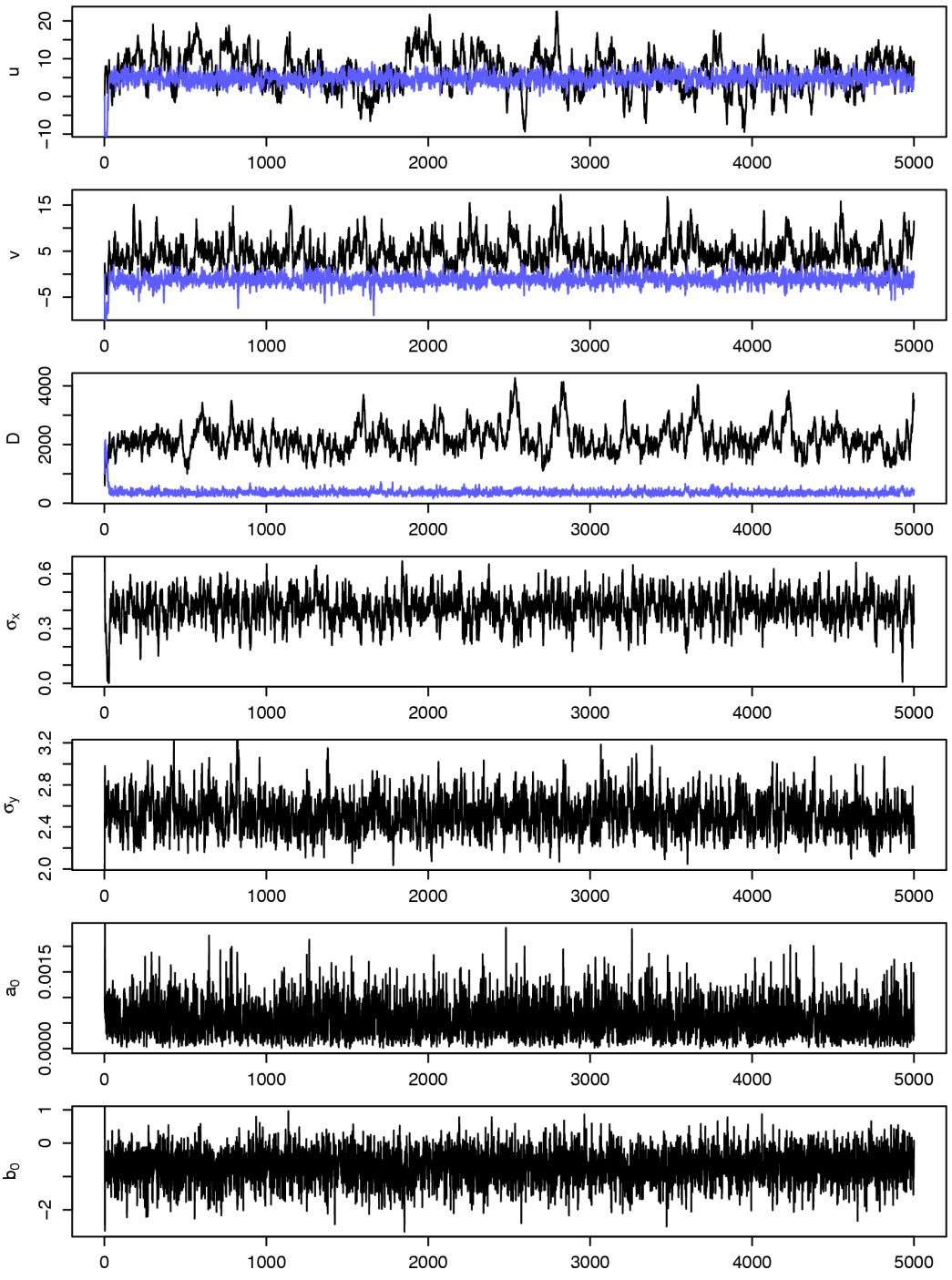

Figure 3: Samples of MCMC chain for 2000 tag releases.

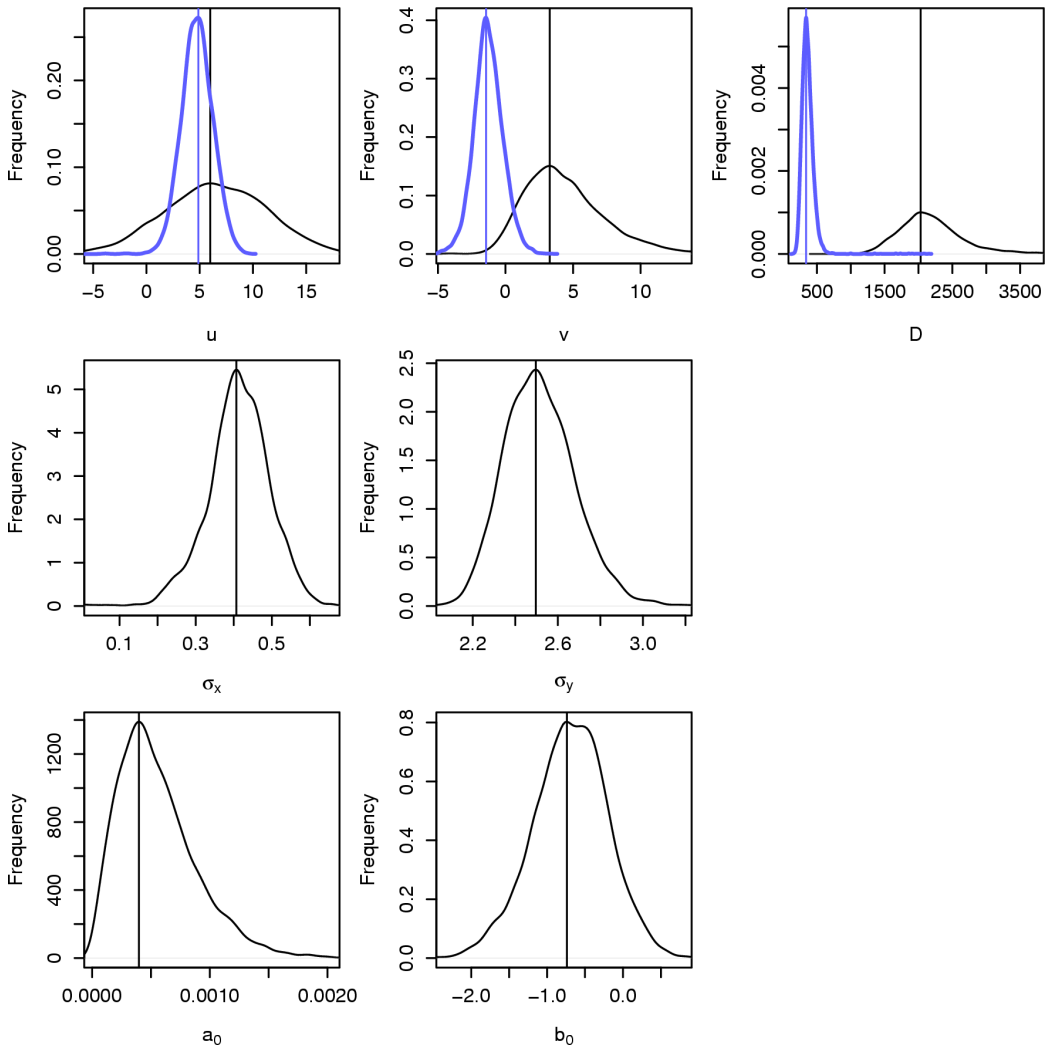

Figure 4: Frequency distribution of parameters from MCMC simulations for 2000 tag releases.

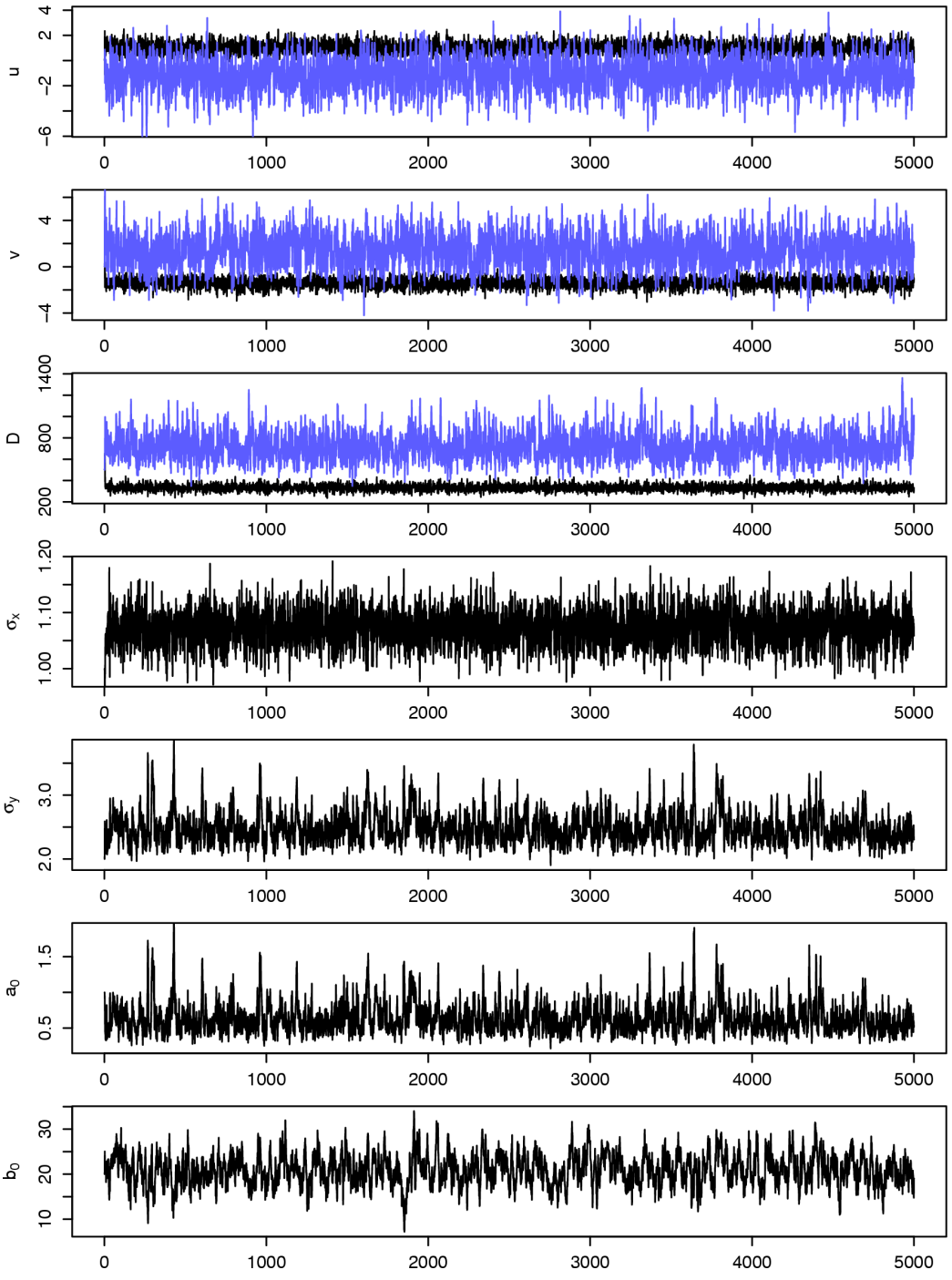

Figure 5: Samples of MCMC chain for 2002 tag releases.

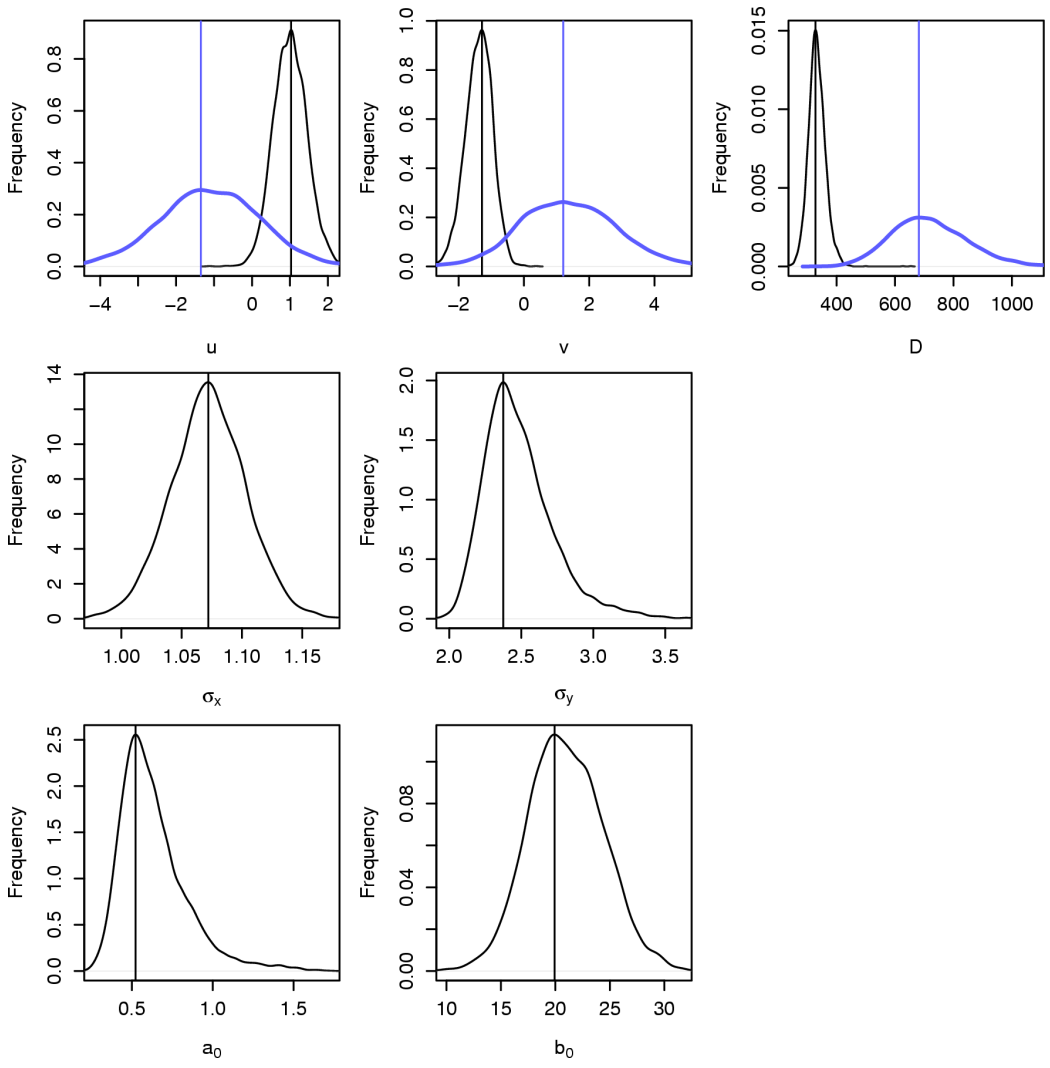

Figure 6: Frequency distribution of parameters from MCMC simulations for 2002 tag releases.

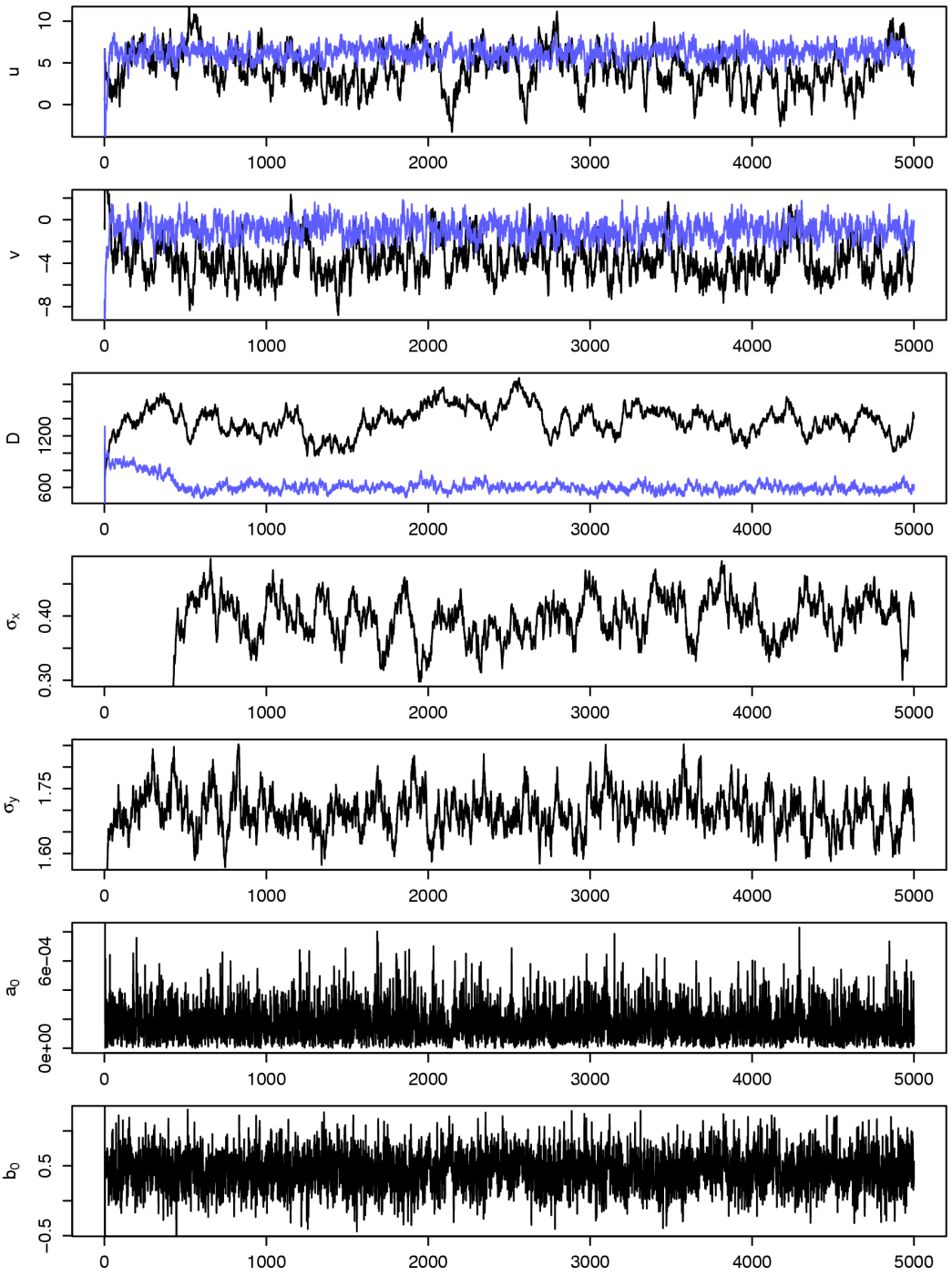

Figure 7: Samples of MCMC chain for combined 1999 and 2000 tag releases.

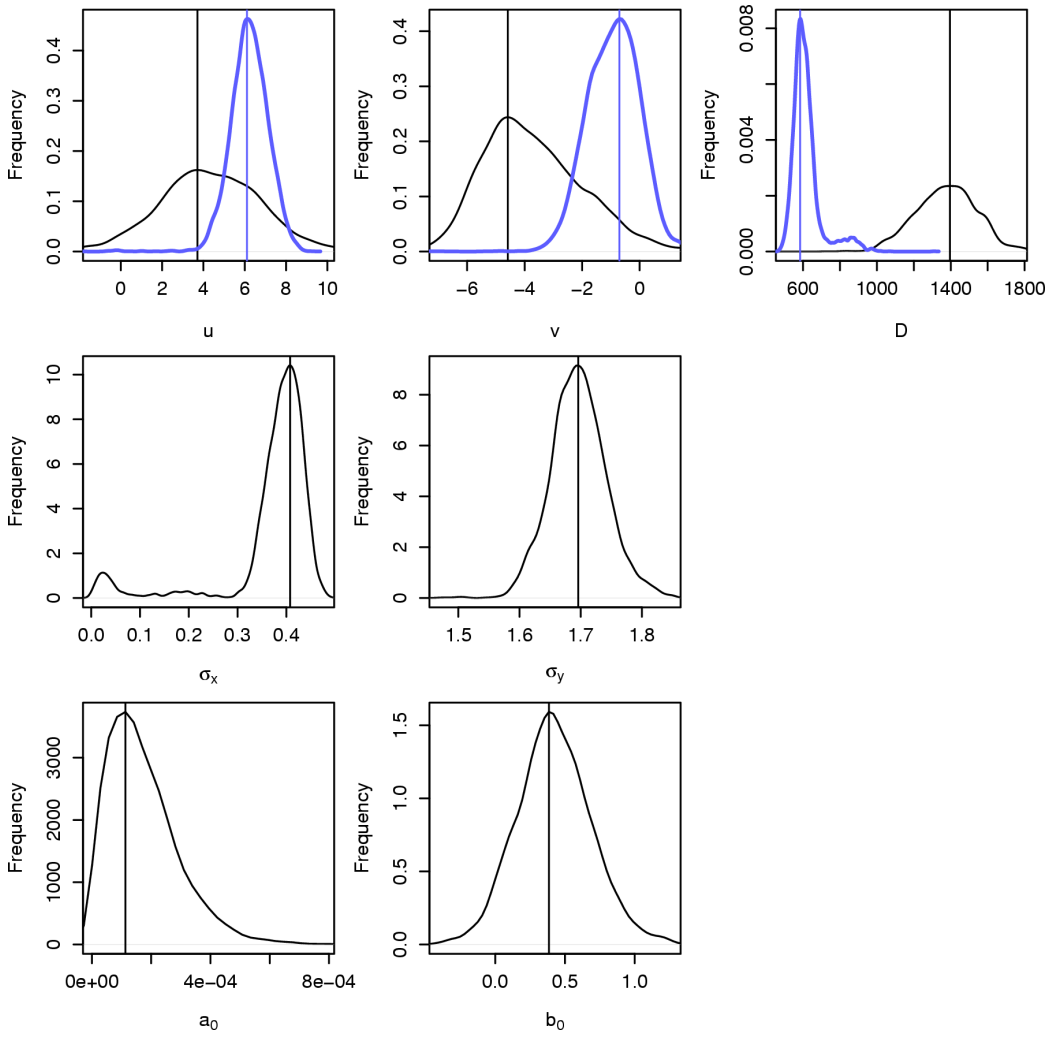

Figure 8: Frequency distribution of parameters from MCMC simulations for combined 1999 and 2000 tag releases.

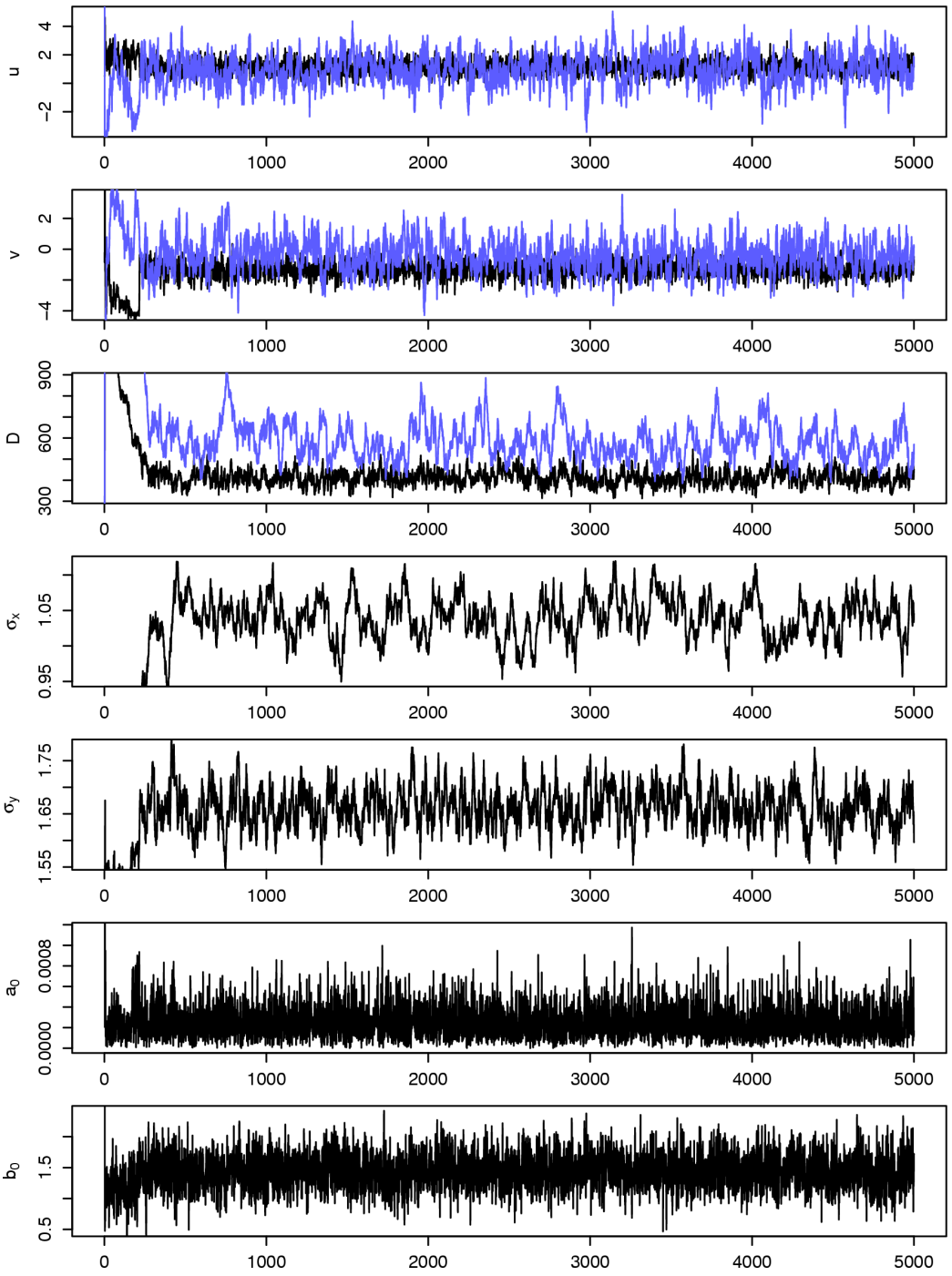

Figure 9: Samples of MCMC chain for combined 2000 and 2002 tag releases.

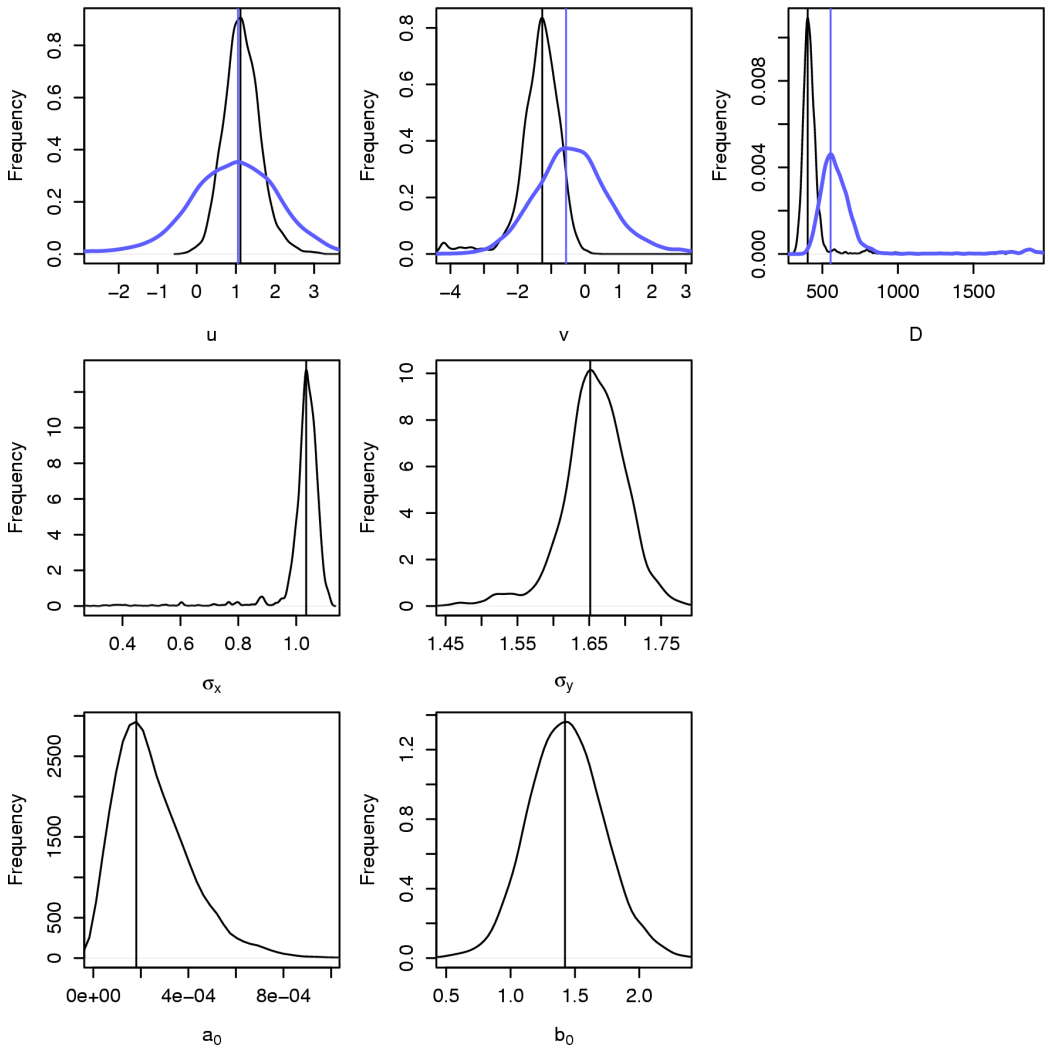

Figure 10: Frequency distribution of parameters from MCMC simulations for combined 2000 and 2002 tag releases.

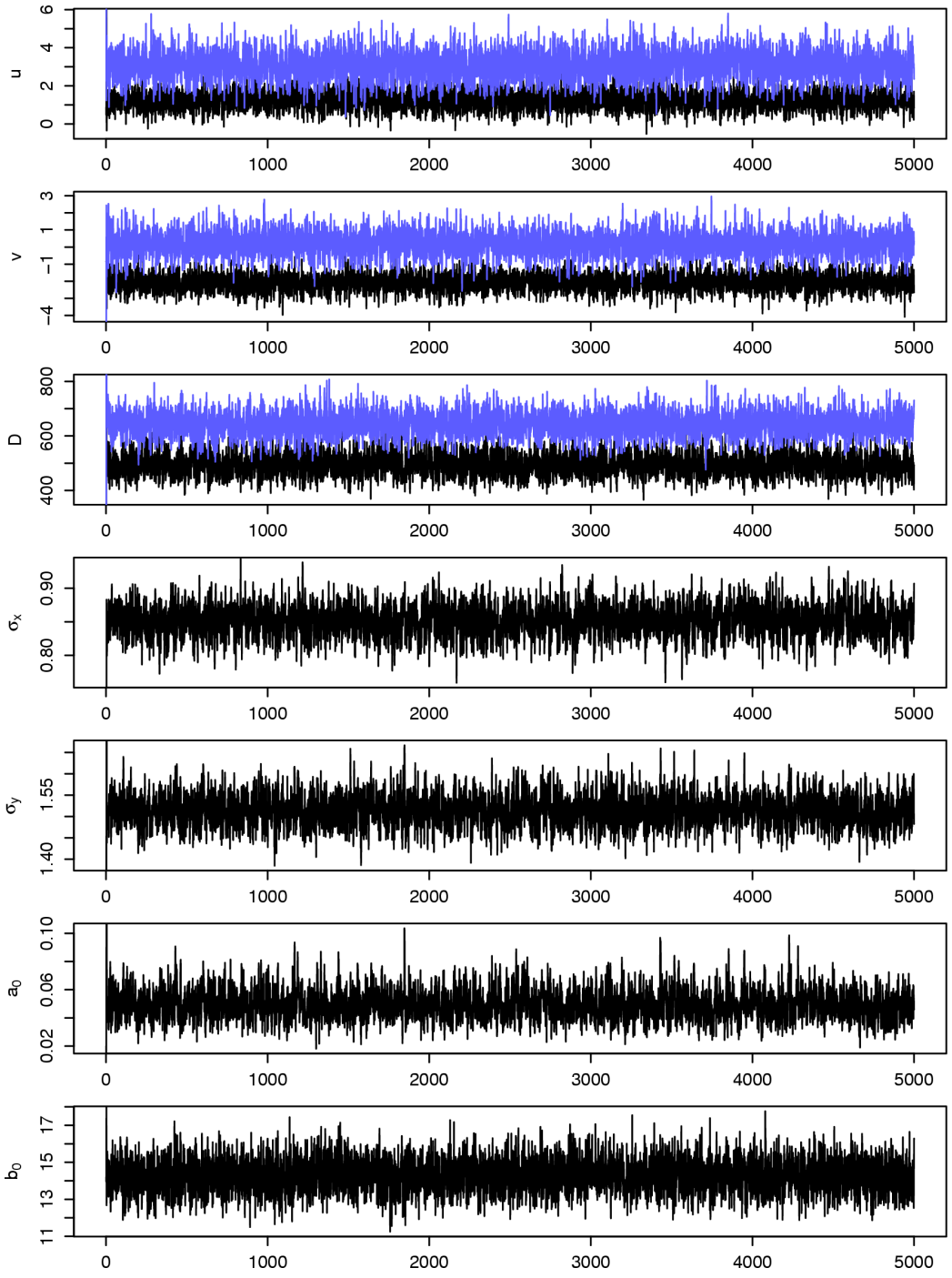

Figure 11: Samples of MCMC chain for combined 1999 and 2002 tag releases.

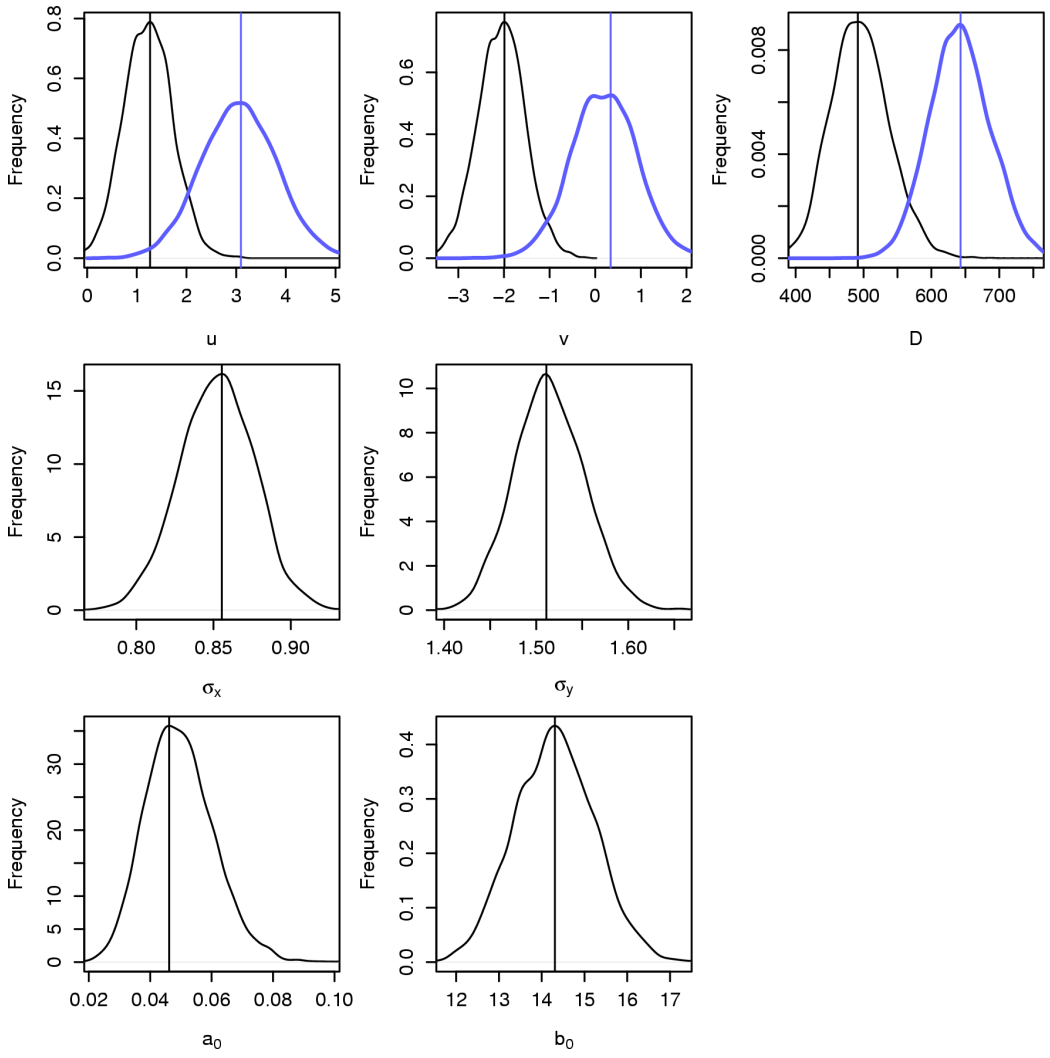

Figure 12: Frequency distribution of parameters from MCMC simulations for combined 1999 and 2002 tag releases.

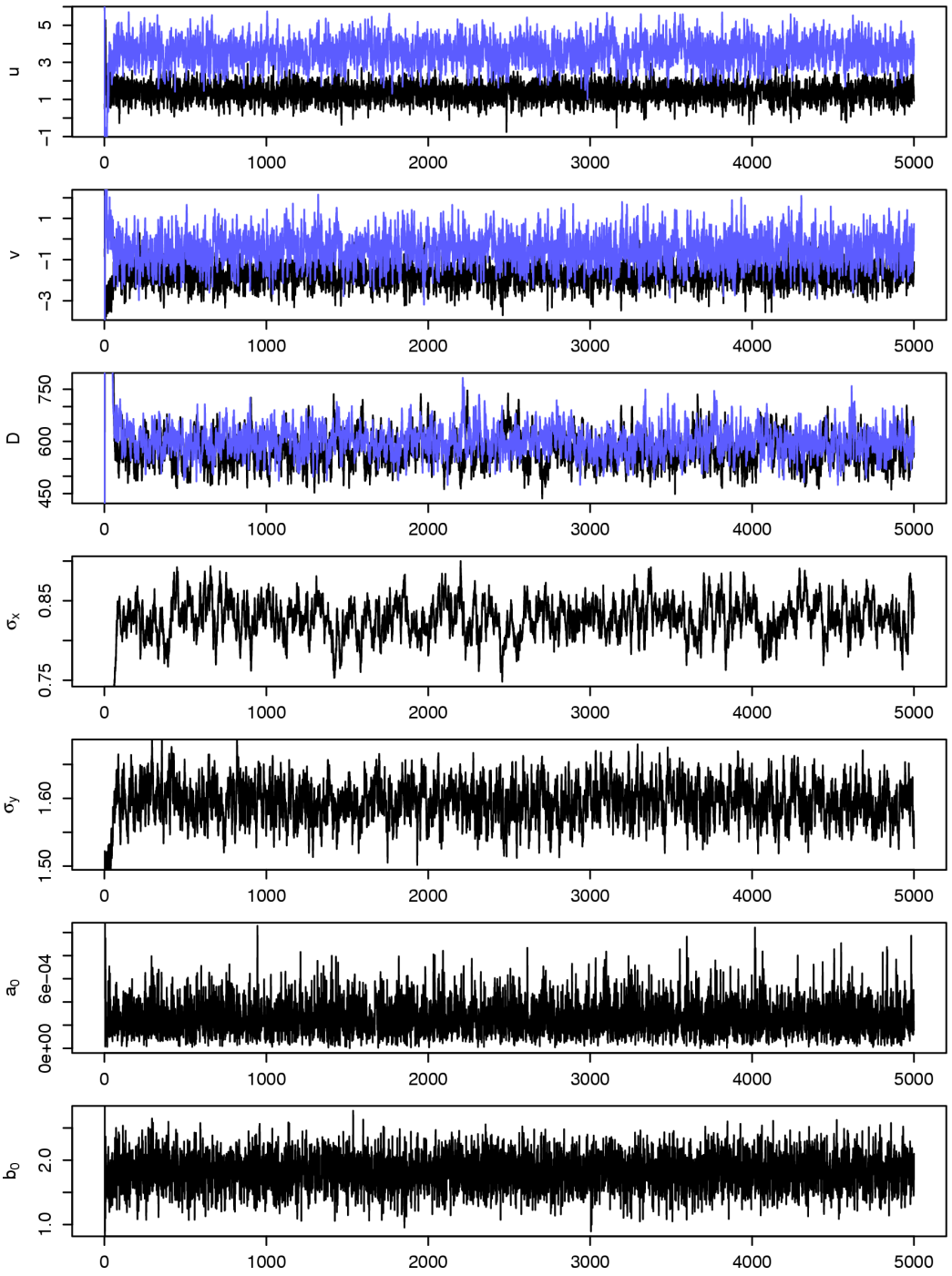

Figure 13: Samples of MCMC chain for all tag releases combined.

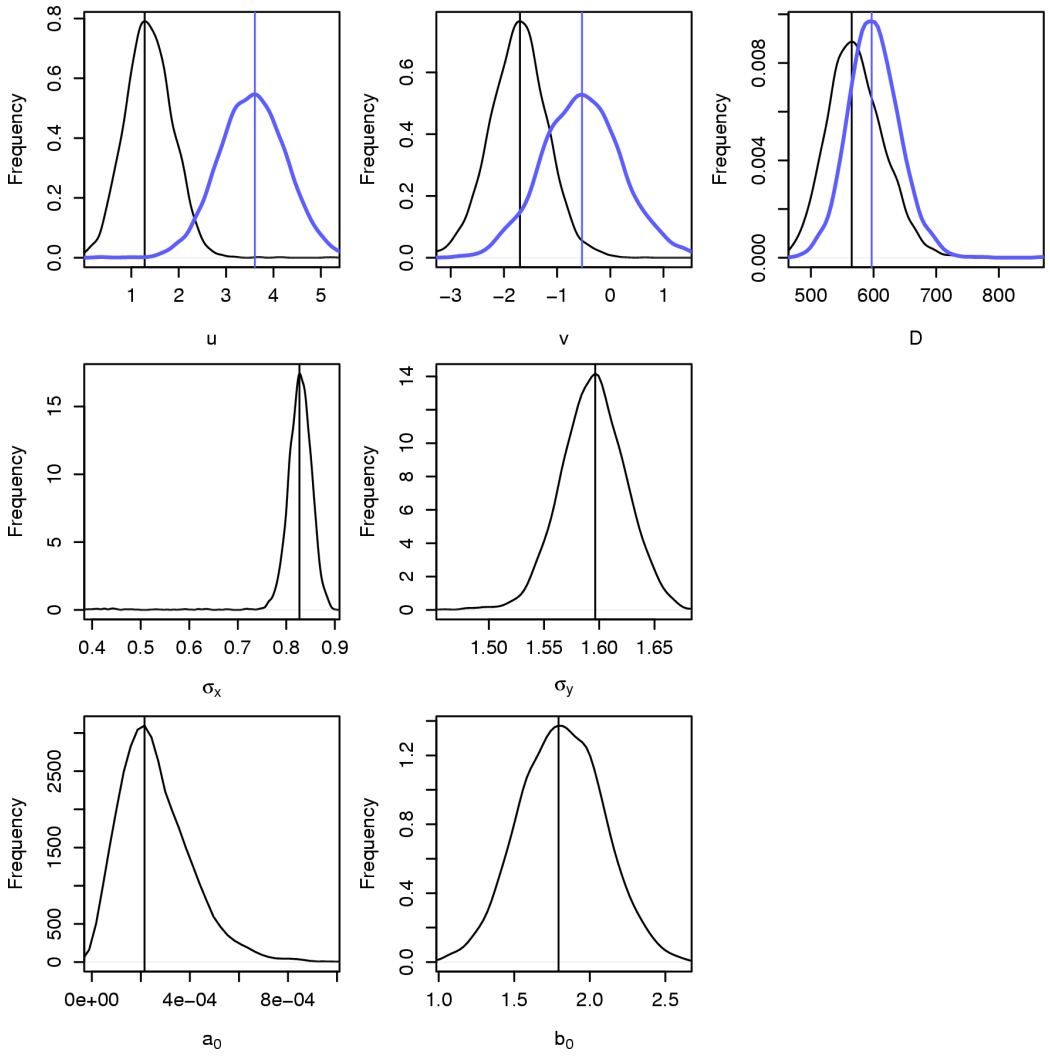

Figure 14: Frequency distribution of parameters from MCMC simulations for all tag releases combined.