Application of Multivariate Maxwellian Mixture Model to Plasma Velocity Distribution

Genta Ueno[1,3], Nagatomo Nakamura[2], Tomoyuki Higuchi[3], Takashi Tsuchiya[3], Shinobu Machida[1], and Tohru Araki[1]

[1] Department of Geophysics, Graduate School of Science, Kyoto University, Kyoto 606-8502, Japan
genta@kugi.kyoto-u.ac.jp
[2] Sapporo Gakuin University, Hokkaido 069-8555, Japan
[3] The Institute of Statistical Mathematics, Tokyo 106-8569, Japan

Abstract. Recent space plasma observations have provided us with three-dimensional velocity distributions having multiple peaks. We propose a method for analyzing such velocity distributions via a multivariate Maxwellian mixture model where each component of the model represents each of the multiple peaks. The parameters of the model are determined through the Expectation-Maximization (EM) algorithm. For the automatic judgment of the preferable number of components in the mixture model, we introduce a method of examining the number of extrema of a resulting mixture model. We show applications of our method to velocity distributions observed in the Earth's magnetotail.

1 Introduction

From direct measurements of space plasma by spacecraft we have obtained macroscopic physical quantities by calculating velocity moments of plasma particle velocity distributions (e.g., number density, bulk velocity and temperature). This macroscopic description assumes that the plasma is in a state of local thermal equilibrium. Under this assumption a particle velocity distribution is given as a normal distribution which is called the Maxwellian distribution in a scientific domain, plasma physics. The Maxwellian distribution is given by

$$g\left(\boldsymbol{v}\middle|\,\boldsymbol{V},\,\mathrm{T}\right)=\left(\frac{m}{2\pi\mathrm{T}}\right)^{3/2}\exp\left[-\frac{m\left|\boldsymbol{v}-\boldsymbol{V}\right|^{2}}{2\mathrm{T}}\right], \tag{1}$$

where $m\,[\mathrm{kg}]$ is the mass of the particle, $\boldsymbol{V}\,[\mathrm{m/s}]$ is the bulk velocity vector and $\mathrm{T}\,[\mathrm{J}]$ is the temperature.

Observational techniques are progressing notably today, and making it possible to measure detailed shapes of velocity distributions in the three-dimensional velocity space. These observations have revealed that there frequently happen cases in which space plasmas are not in a state of thermal equilibrium and their velocity distributions are not the Maxwellian but consist of multiple peaks.

This is because space plasmas are basically collisionless with large mean-free-path. Therefore we have to be aware that they may give the same velocity moments even if the shapes of distributions differ. For instance, when a plasma has two beam components whose velocity vectors are sunward and anti-sunward, and each component has the same numbers of particles, the bulk velocity becomes zero because their velocity vectors cancel out. On the other hand, when a stagnant plasma is observed, the bulk velocity also becomes zero. When we deal with two-beam distributions, we should separate the distributions into two beams and calculate the velocity moments for each beam. Such non-equilibrium multi-component distribution have been reported many times, and a kinetic description of space plasmas that accounts the shape of the velocity distribution have come to be required.

It has been difficult, however, to evaluate the shape of the velocity distribution. Some of the previous researchers have calculated the velocity moments of each component separated by their visual inspections; it would be expected to take time and have limitation. Moreover, resultant moment values will not be estimated accurately when more than one components partially overlap each other.

In this paper we develop a method of representing a three-dimensional distribution by a multivariate Maxwellian mixture model [7, 6] in which the parameter values are obtained by the Expectation-Maximization (EM) algorithm [7, 3, 5]. This method enables us to express the shape of the distribution and to find a feasible way to conduct a statistical analysis for many multi-component cases. The organization of this paper is the following. In Sect. 2, we describe the data of plasma velocity distribution. A fitting method with multivariate Maxwellian mixture model is described in Sect. 3, followed by considerations on how to judge the preferable number of components in the mixture model in Sect. 4. Two applications are demonstrated in Sect. 5. In Sect. 6, we discuss a problem of model selection. We conclude this paper in Sect. 7.

2 Data

We used ion velocity distributions obtained by an electrostatic analyzer named LEP-EA on board the Geotail spacecraft. LEP-EA measured three-dimensional velocity distributions by classifying the velocity space into 32 for the magnitude of the velocity, 7 for elevation angles and 16 for azimuthal sectors (Fig. 1). Let us assume that LEP-EA detected the ion count $C\left(\boldsymbol{v}_{pqr}\right)$ [#] in a sampling time τ [s], where $\boldsymbol{v}_{pqr}$ [m/s] is the ion velocity. Subscription p, q and r are indicators of the magnitude of the velocity, elevation angle and azimuthal sector, and they take integers $p = 1, \cdots, 32$; $q = 1, \cdots, 7$; and $r = 1, \cdots, 16$. Thus we obtain the total ion count N [#]:

$$N = \sum_{p,q,r} C\left(\boldsymbol{v}_{pqr}\right). \tag{2}$$

Under the assumption that the incident differential ion flux is uniform within the energy and angular responses of the analyzer, the velocity distribution $f_0\left(\boldsymbol{v}_{pqr}\right)$

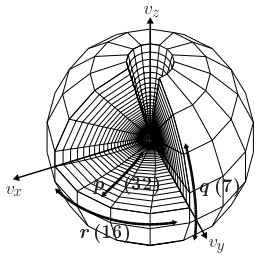

Fig. 1. Classes for observation of a velocity distribution with LEP-EA. Three orthogonal axes v_x, v_y, and v_z are taken in accordance with the spacecraft coordinates

$[\mathrm{s}^3/\mathrm{m}^6]$ is given by

$$f_0\left(\boldsymbol{v}_{pqr}\right) = 2 \times 10^4 \frac{1}{\tau \varepsilon_q \mathcal{G}_q} \frac{C\left(\boldsymbol{v}_{pqr}\right)}{\left(\boldsymbol{v}_{pqr}^T \boldsymbol{v}_{pqr}\right)^2}, \tag{3}$$

where ε_q is the detection efficiency and $\mathcal{G}_q\,[\mathrm{cm}^2\,\mathrm{sr\,eV/eV}]$ is the g-factor. Integrating $f_0\left(\boldsymbol{v}_{pqr}\right)$ over the velocity space, we obtain the number density $n\,[\#/\mathrm{m}^3]$:

$$n = \sum_{p,q,r} f_0\left(\boldsymbol{v}_{pqr}\right) d\boldsymbol{v}_{pqr}, \tag{4}$$

where $d\boldsymbol{v}_{pqr}$ is the class interval whose class mark is $\boldsymbol{v}_{pqr}$.

For convenience of statistical modeling, we deal with a probability function $f\left(\boldsymbol{v}_{pqr}\right)$ instead of $f_0\left(\boldsymbol{v}_{pqr}\right)$ defined by:

$$f\left(\boldsymbol{v}_{pqr}\right) = \frac{f_0\left(\boldsymbol{v}_{pqr}\right) d\boldsymbol{v}_{pqr}}{n}, \tag{5}$$

so that $\sum_{p,q,r} f\left(\boldsymbol{v}_{pqr}\right) = 1$. Since $f\left(\boldsymbol{v}_{pqr}\right)$ is a function of discrete variables, it is necessary to consider the mixture of probability functions, but we approximate it by the mixture of the Maxwellian that is a probability *density* function.

3 Method

3.1 Multivariate Maxwellian Mixture Model

To deal with multiple components seen in the probability function (Eq. (5)), we fit Eq. (5) by the mixture model composed of the sum of s multivariate Maxwellian distributions:

$$f\left(\boldsymbol{v}_{pqr}\right) \simeq \sum_{i=1}^{s} n_i g_i\left(\boldsymbol{v}_{pqr} | \boldsymbol{V}_i,\, \mathbf{T}_i\right), \tag{6}$$

where n_i is the mixing proportion of the Maxwellians ($\sum_{i=1}^{s} n_i = 1,\ 0 < n_i < 1$). Each multivariate Maxwellian g_i is written as

$$\begin{aligned} & g_i\left(\boldsymbol{v}_{pqr} | \boldsymbol{V}_i,\ \mathbf{T}_i\right) \\ & = \left(\frac{m}{2\pi}\right)^{3/2} \frac{1}{\sqrt{|\mathbf{T}_i|}} \exp\left[-\frac{m}{2}\left(\boldsymbol{v}_{pqr} - \boldsymbol{V}_i\right)^T {\mathbf{T}_i}^{-1}\left(\boldsymbol{v}_{pqr} - \boldsymbol{V}_i\right)\right], \end{aligned} \tag{7}$$

where $\boldsymbol{V}_i$ [m/s] is the bulk velocity vector and $\mathbf{T}_i$ [J] is the temperature matrix of i-th multivariate Maxwellian, and superscript T denotes transposition operation. The multivariate Maxwellian (7) is a generalized form of the Maxwellian (1) as a thermal equilibrium velocity distribution, and can also deal with anisotropic temperature. Plasma with anisotropic temperature is often observed in space because of its collisionless property; it takes a long time for non-equilibrium plasma to relax into a state of thermal equilibrium due to rare particle interactions.

3.2 Parameter Estimation Procedure

The log-likelihood of this mixture model (Eq. (6)) becomes

$$l\left(\theta\right) = N \sum_{p,q,r} f\left(\boldsymbol{v}_{pqr}\right) \log \sum_{i=1}^{s} n_i g_i\left(\boldsymbol{v}_{pqr} | \boldsymbol{V}_i,\ \mathbf{T}_i\right) \tag{8}$$

where $\theta = (n_1,\ n_2,\ \cdots,\ n_{s-1},\ \boldsymbol{V}_1,\ \boldsymbol{V}_2,\ \cdots,\ \boldsymbol{V}_s,\ \mathbf{T}_1,\ \mathbf{T}_2,\ \cdots,\ \mathbf{T}_s)$ denotes the all unknown parameters. These parameters directly correspond to the velocity momemts for each component.

Partially differentiate (8) with respect to $\boldsymbol{V}_i$, $\mathbf{T}_i^{-1}$ $(i = 1,\ 2, \cdots,\ s)$ and put them equal to zero, maximum likelihood estimators (denoted by $\hat{}$) of the mixing proportion, the bulk velocity vector and the temperature matrix of each Maxwellian are given by

$$\hat{n}_i = \sum_{p,q,r} f\left(\boldsymbol{v}_{pqr}\right) \hat{P}_i\left(\boldsymbol{v}_{pqr}\right), \tag{9}$$

$$\hat{\boldsymbol{V}}_i = \frac{1}{\hat{n}_i} \sum_{p,q,r} f\left(\boldsymbol{v}_{pqr}\right) \hat{P}_i\left(\boldsymbol{v}_{pqr}\right) \boldsymbol{v}_{pqr}, \tag{10}$$

$$\hat{\mathbf{T}}_i = \frac{1}{\hat{n}_i} \sum_{p,q,r} f\left(\boldsymbol{v}_{pqr}\right) \hat{P}_i\left(\boldsymbol{v}_{pqr}\right) \frac{1}{2} m \left(\boldsymbol{v}_{pqr} - \hat{\boldsymbol{V}}_i\right) \left(\boldsymbol{v}_{pqr} - \hat{\boldsymbol{V}}_i\right)^T, \tag{11}$$

where

$$\hat{P}_i\left(\boldsymbol{v}_{pqr}\right) = \frac{\hat{n}_i g_i\left(\boldsymbol{v}_{pqr} | \hat{\boldsymbol{V}}_i,\ \hat{\mathbf{T}}_i\right)}{\sum_{j=1}^{s} \hat{n}_j g_j\left(\boldsymbol{v}_{pqr} | \hat{\boldsymbol{V}}_j,\ \hat{\mathbf{T}}_j\right)} \tag{12}$$

is an estimated posterior probability.

We should note that when applying a single-Maxwellian model ($s = 1$), we will obtain the parameters identical to the usual velocity moments. This is because the number density is obtained by the usual moment calculation (Eq. (4)), and because the bulk velocity and the temperature matrix as maximum likelihood estimators prove to be identical with those obtained from the moment calculation [8].

On the basis of Eqs. (9)–(12), we estimate the unknown parameters by the EM algorithm [7]. In the following procedure, t denotes an iteration counter of the EM algorithm. Suppose that superscript (t) denotes the current values of the parameters after t cycles of the algorithm for $t = 0, 1, 2, \cdots$.

Setting Initial Value: $t = 0$. We classify the observed velocity space in s groups (G_i; $i = 1, 2, \cdots, s$) using the k-means algorithm, and set the initial value of the posterior probability as

$$P_i^{(0)}(\boldsymbol{v}_{pqr}) = \begin{cases} 1 & (\boldsymbol{v}_{pqr} \in G_i) \\ 0 & (\boldsymbol{v}_{pqr} \notin G_i) \end{cases}, \tag{13}$$

where $i = 1, 2, \cdots, s$. With $P_i^{(0)}(\boldsymbol{v}_{pqr})$, we calculate $\hat{n}_i^{(0)}$, $\hat{\boldsymbol{V}}_i^{(0)}$, and $\hat{\mathbf{T}}_i^{(0)}$ by Eqs. (9), (10) and (11).

Parameter Estimation by EM Algorithm: $t \geq 1$. On the t-th iteration ($t \geq 1$), we compute $n_i^{(t)}$ and $P_i^{(t)}(\boldsymbol{v}_{pqr})$ by Eqs. (9) and (12) as the E-step. At the M-step, we choose $\boldsymbol{V}_i^{(t)}$ and $\mathbf{T}_i^{(t)}$ as maximum likelihood estimators by Eqs. (10) and (11).

Judgment of Convergence. We finish the iteration if

$$\left| l\left(\hat{\theta}^{(t)}\right) - l\left(\hat{\theta}^{(t-1)}\right) \right| < \epsilon \quad \text{and} \quad \left\| \hat{\theta}^{(t)} - \hat{\theta}^{(t-1)} \right\| < \delta, \tag{14}$$

where ϵ and δ are sufficiently small positive number. If the above convergence condition is not satisfied, return to the E-step with replacing t by $t + 1$.

4 Preferable Number of Components

When we fit a velocity distribution by the Maxwellian mixture model, we should examine how reasonable the fit is. That is, for instance, it is inappropriate to approximate a unimodal observation by a multi-Maxwellian mixture model. Here we introduce a method of judging which of two models, i.e., a single-Maxwellian model or a two-Maxwellian mixture model, is preferable for each observation. We adopt the following principle. If a two-Maxwellian mixture model which resulted from an observation shows two peaks, the observation will also have two peaks. Hence we conclude that the two-Maxwellian mixture model is reasonable to use.

On the other hand, if the resulting two-Maxwellian mixture model has only one peak, the observation will be a unimodal distribution: We should use a usual single-Maxwellian fitting.

To judge whether the fitting result is reasonable or not, we enumerate the number of peaks of the resulting fitted model. Actually, to count the number of peaks, we enumerate the number of extrema of the model. Let us consider when we fit some data $f(\boldsymbol{v})$ by a two-Maxwellian mixture model and the fitting result is computed as

$$f(\boldsymbol{v}) \simeq n_1 g_1(\boldsymbol{v}|\boldsymbol{V}_1, \mathbf{T}_1) + n_2 g_2(\boldsymbol{v}|\boldsymbol{V}_2, \mathbf{T}_2). \tag{15}$$

To count the number of peaks, we need to count the number of $\boldsymbol{v}$ satisfying

$$\frac{d}{d\boldsymbol{v}}\left[n_1 g_1(\boldsymbol{v}|\boldsymbol{V}_1, \mathbf{T}_1) + n_2 g_2(\boldsymbol{v}|\boldsymbol{V}_2, \mathbf{T}_2)\right] = 0. \tag{16}$$

It is difficult, however, to treat the three-dimensional variable $\boldsymbol{v}$. We then reduce this three-dimensional problem to one set of simultaneous equations of one-dimensional variables (details are given in Ref.[11, 12]):

$$\eta(\xi) = \xi, \tag{17}$$

$$\eta(\xi) = \frac{n_1 g_1(\boldsymbol{w}(\xi)|0, \mathbf{I})}{n_2 g_2(\boldsymbol{w}(\xi)|\boldsymbol{W}, \mathbf{M}^{-1})}, \tag{18}$$

whose number of solutions is equivalent to the number of $\boldsymbol{v}$ satisfying Eq. (16). Here we put $\boldsymbol{w}(\xi) = (\mu_1 W_1/(\xi+\mu_1),\ \mu_2 W_2/(\xi+\mu_2),\ \mu_3 W_3/(\xi+\mu_3))$, $\boldsymbol{W} = (\boldsymbol{y}_1, \boldsymbol{y}_2, \boldsymbol{y}_3)^{-1}\mathbf{L}^T(\boldsymbol{V}_2 - \boldsymbol{V}_1)$, where $\mathbf{L}$ is a matrix that satisfies $\mathbf{L}\mathbf{L}^T = \mathbf{T}_1^{-1}$, and μ_1, μ_2 and μ_3 are the eigenvalues of $\left(\mathbf{L}^T\mathbf{T}_2\mathbf{L}\right)^{-1}$ whose corresponding unit eigenvectors are $\boldsymbol{y}_1$, $\boldsymbol{y}_2$ and $\boldsymbol{y}_3$. Consequently, we need to count the nodes of the line (17) and the curve (18) in the ξ-η plane.

Utilizing Eqs. (17) and (18), we can also evaluate a mixture model with three or more Maxwellians. That is, we first pick up all the combinations of two Maxwellians from the multiple Maxwellians, and then apply Eqs. (17) and (18) for each combination.

5 Application

The left-hand panel of Fig. 2(a) show an ion velocity distribution observed in the plasma sheet boundary layer of the Earth's magnetotail. Displayed distribution is a slice by the v_x-v_y plane, whose values are black-to-white-coded as shown in the bar. Used coordinate system is taken in accordance with the spacecraft coordinate system: The v_z axis is parallel to the spacecraft spin axis and positive northward, the v_x axis is parallel to the projection of the spacecraft-Sun line on the plane whose normal vector is the v_z axis and is positive sunward, and the v_y axis completes a right-hand orthogonal system. We can observe a hot component and a cold component whose bulk velocities are $(v_x, v_y) \simeq (1000, 0)$ km/s and $(v_x, v_y) \simeq (-200, -500)$ km/s, respectively.

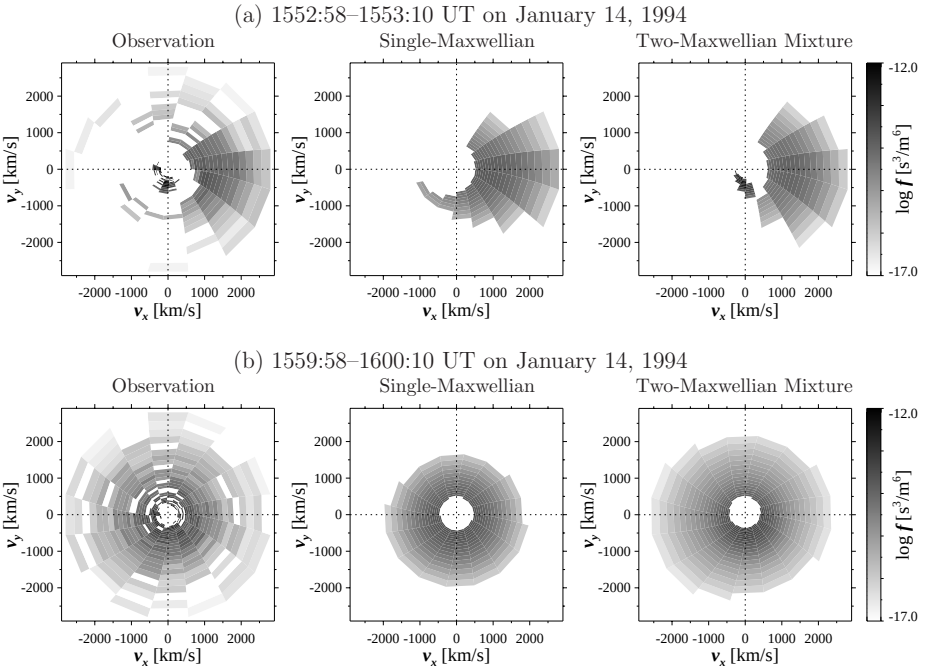

Fig. 2. (a) Observation of an ion velocity distribution on the v_x-v_y plane in the time interval 1552:58–1553:10 UT on January 14, 1994 (left), fitting functions by a single-Maxwellian model (center) and a two-Maxwellian mixture model (right). (b) For the observation between 1559:58–1600:10 UT on the same day

When we fit this data with a single-Maxwellian, we obtain the estimated parameters given in the second row of Table 1(a). A single-Maxwellian model with these parameters produces the distribution shown in the central panel of Fig. 2(a). This corresponds to what we deal with by the usual velocity moments, and it is hot and has a shifted bulk velocity compared with the observation.

This problem, however, is easily resolved by applying a two-Maxwellian mixture model. Similarly, we give the estimated parameters in the third and fourth rows of Table 1(a), and display the produced distribution in the right-hand panel of Fig. 2(a). It can be seen that both the hot and cold components existing in the observed distribution were properly reproduced.

For this example, we find that the fitting by the two-Maxwellian model is preferable to the single-Maxwellian model by counting the number of solutions with simultaneous equations (17) and (18). The two-Maxwellian fitting is therefore justified, which agrees with our inspection of the observed distribution.

The other example is a distribution observed in the central plasma sheet. As can be seen in the left-hand panel of Fig. 2(b), it is appropriate to expect that this consists of a single hot component whose bulk velocity is located near

Table 1. Estimated parameters for single-Maxwellian and two-Maxwellian mixture models. The first column n is the number density, that is, the mixing proportion multiplied by the number density

(a) 1552:58–1553:10 UT on January 14, 1994

	n [/cc]	V_x [km/s]	V_y	V_z	T_{xx} [eV]	T_{xy}	T_{xz}	T_{yy}	T_{yz}	T_{zz}
1	0.040	793	−104	6	5515	327	−216	2177	−129	2089
1	0.012	−101	−260	47	120	−138	23	326	−31	95
2	0.028	1167	−38	−11	2815	−347	−90	2800	−130	2912

(b) 1559:58–1600:10 UT on January 14, 1994

	n [/cc]	V_x [km/s]	V_y	V_z	T_{xx} [eV]	T_{xy}	T_{xz}	T_{yy}	T_{yz}	T_{zz}
1	0.087	−5	−121	−135	3203	16	44	2936	−174	3185
1	0.029	−94	−46	−27	6495	192	187	5676	−275	6359
2	0.058	39	−158	−189	1512	−20	48	1536	−187	1523

the origin of the velocity space. Hence, when we fit the data, we should adopt a single-Maxwellian model rather than a two-Maxwellian mixture model.

In the central panel of Fig. 2(b), we show the calculated distribution with the single-Maxwellian model. The parameters used are given in the second row of Table 1(b). In this case the single-Maxwellian fitting appears to be sufficient. Furthermore, we display the result with the two-Maxwellian mixture model. The right-hand panel shows the calculated distribution with the estimated parameters given in the third and fourth rows of Table 1(b).

By examining the number of solutions of the simultaneous equations for ξ and η, we find that they have only one solution. We therefore adopt the usual velocity moments obtained by the single-Maxwellian fitting.

6 Discussion

To select the preferable number of components, we adopt in this study an empirical approach; we first fit the data with a two-Maxwellian mixture model, then examine whether there is a saddle point on the segment between the bulk velocities of the model. Generally, AIC (Akaike Information Criterion [1]) defined by

$$\mathrm{AIC} = -2 \max l\left(\theta\right) + 2 \dim \theta, \tag{19}$$

has been frequently employed for this problem [6]. However, AIC selected a mixture model having a larger number of components compared with our intuition. According to AIC, the best number of components was found to be six or more for both distributions shown in Fig. 2.

This is expected to be due to following three reasons. First, it is not so appropriate in our data set to adopt the Maxwellian distribution as a component

distribution of a mixture model. While the Maxwellian distribution can well represent the observed distribution near the peak, it cannot follow the distribution in the large-velocity range. Since the observed distribution has a heavy tail in the large-velocity range, it is necessary to have many components for fitting such a tail accurately. In fact, the two-Maxwellian mixture model shown in Fig. 2(b) consists of two components which present a peak and a heavy tail, respectively. This problem would be solved when using a mixture model which consists of heavy tail distributions instead of the Maxwellian distributions.

One of the heavy tail distributions is the κ distribution defined by

$$g_i\left(\boldsymbol{v}_{pqr}|\,\boldsymbol{V}_i,\,\mathbf{T}_i,\,\kappa_i\right)=\left(\frac{m}{2\pi}\right)^{3/2}\frac{\Gamma\left(\kappa_i\right)}{\Gamma\left(\kappa_i-3/2\right)}\frac{1}{\sqrt{|\mathbf{T}_i|}}\cdot\left[1+\frac{m}{2}\left(\boldsymbol{v}_{pqr}-\boldsymbol{V}_i\right)^T{\mathbf{T}_i}^{-1}\left(\boldsymbol{v}_{pqr}-\boldsymbol{V}_i\right)\right]^{-\kappa_i}. \quad (20)$$

This converges to the Maxwellian distribution in the limit of $\kappa\to\infty$, so it can give a more comprehensive treatment of the data. When we select the κ distribution as a component distribution, the algorithm presented in Sect. 3 can work by including the κ renewing step. However, our experiments found that the estimated κ value was the order of 10^1, which means that the resultant distribution is practically the Maxwellian.

The second reason is that the observation has a large total ion count ($N=2925$ and 4762 for the Figs. 2 (a) and (b), respectively). The log-likelihood $l\left(\theta\right)$ is multiplied by N as defined in Eq. (8), and $\max l\left(\theta\right)$ was on the order of 10^4–10^5 in the two examples. On the other hand, the dimension of free parameters θ was on the order of 10^0–10^1. AIC was determined practically by $\max l\left(\theta\right)$ and was not affected by $\dim\theta$ as a penalty term.

To take large N into account, we evaluated the competing models by BIC (Bayesian Information Criterion [9]) instead of AIC. BIC is an information criterion such that posterior probability is maximized and defined as

$$\mathrm{BIC}=-2\max l\left(\theta\right)+\log N\dim\theta. \quad (21)$$

BIC, however, yielded the same result as with AIC in our cases.

Finally, we should notice that there exist some classes of $\boldsymbol{v}_{pqr}$ around $\boldsymbol{v}=0$ such that $f\left(\boldsymbol{v}_{pqr}\right)=0$, which is identified as a white region around $v_x=v_y=0$ in the two observations displayed in Fig. 2. This is due to the instrument, an electrostatic analyzer. That is, when we observe the ambient velocity distribution as shown in the left-hand panel of Fig. 3 by the electrostatic analyzer, we obtain the count $C\left(\boldsymbol{v}_{pqr}\right)$ as in the central panel. Since $C\left(\boldsymbol{v}_{pqr}\right)$ is a count data, it becomes zero if it is less than unity (under the one count level presented by the dotted line). The zero count is then converted to zero probability $f\left(\boldsymbol{v}_{pqr}\right)$ through Eqs. (3) and (5). Namely, probability below the one-count level curve becomes zero in the observation (see the right-hand panel). This cut off effect with quantization occurs especially around $\boldsymbol{v}=0$, which produces the observed distribution, $f\left(\boldsymbol{v}_{pqr}\right)$, having a "hole" around $\boldsymbol{v}=0$ as in the right-hand panel.

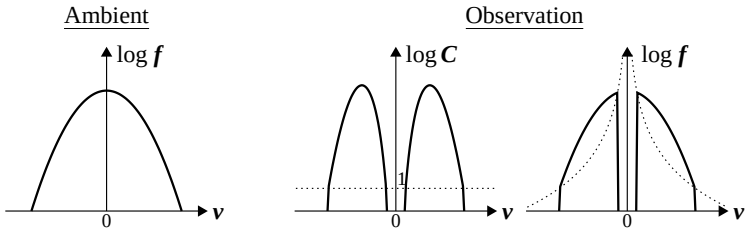

Fig. 3. Creation of a "hole" in the observed velocity distribution near the origin. When the ambient velocity distribution (left) is observed, the obtained count and velocity distribution become the ones as shown in the central and right-hand panels. The dotted line and curve present one count level

AIC is expected to choose the model with multi-components to present the edge of the "hole" precisely.

This "hole" effect will be reduced if we give a probabilistic description to the ion count $C(\boldsymbol{v}_{pqr})$. We assume that the ion count observation is generated from a multinominal distribution in which a detection probability of an ion of velocity $\boldsymbol{v}_{pqr}$ is $h(\boldsymbol{v}_{pqr})$. When an ion whose velocity is $\boldsymbol{v}_{pqr}$ is detected $C(\boldsymbol{v}_{pqr})$ times in N trials, the likelihood function becomes

$$\lambda\left(C\left(\boldsymbol{v}_{pqr}\right)|h\left(\boldsymbol{v}_{pqr}\right)\right) = N!\prod_{p,q,r}\frac{h\left(\boldsymbol{v}_{pqr}\right)^{C\left(\boldsymbol{v}_{pqr}\right)}}{C\left(\boldsymbol{v}_{pqr}\right)!}. \tag{22}$$

An expected count of the ion of velocity $\boldsymbol{v}_{pqr}$ is $Nh(\boldsymbol{v}_{pqr})$. We also assume that the velocity distribution as a function of $h(\boldsymbol{v}_{pqr})$ can be approximated by a mixture model. This approximation can be realized by considering the following prior distribution for $h(\boldsymbol{v}_{pqr})$ [4, 10]

$$\pi\left(h\left(\boldsymbol{v}_{pqr}\right)|\theta\right) = \prod_{p,q,r}\left[\sum_{i=1}^{s} n_i g_i\left(\boldsymbol{v}_{pqr}|\boldsymbol{V}_i, \mathbf{T}_i\right)\right]^{Nf'\left(\boldsymbol{v}_{pqr},h\left(\boldsymbol{v}_{pqr}\right)\right)}, \tag{23}$$

where we set

$$f'\left(\boldsymbol{v}_{pqr}, h\left(\boldsymbol{v}_{pqr}\right)\right) = \frac{f'_0\left(\boldsymbol{v}_{pqr}, h\left(\boldsymbol{v}_{pqr}\right)\right)d\boldsymbol{v}_{pqr}}{\sum_{p,q,r} f'_0\left(\boldsymbol{v}_{pqr}, h\left(\boldsymbol{v}_{pqr}\right)\right)d\boldsymbol{v}_{pqr}}, \tag{24}$$

$$f'_0\left(\boldsymbol{v}_{pqr}, h\left(\boldsymbol{v}_{pqr}\right)\right) = 2\times 10^4 \frac{1}{\tau\varepsilon_q \mathcal{G}_q}\frac{Nh\left(\boldsymbol{v}_{pqr}\right)}{\left(\boldsymbol{v}_{pqr}^T\boldsymbol{v}_{pqr}\right)^2}. \tag{25}$$

With Eqs. (22) and (23), the log-likelihood function is defined by

$$l'\left(C\left(\boldsymbol{v}_{pqr}\right)|\theta\right) = \log\int \lambda\left(C\left(\boldsymbol{v}_{pqr}\right)|h\left(\boldsymbol{v}_{pqr}\right)\right)\pi\left(h\left(\boldsymbol{v}_{pqr}\right)|\theta\right)dh\left(\boldsymbol{v}_{pqr}\right). \tag{26}$$

An optimal θ can be also obtained by maximizing $l'\left(C\left(\boldsymbol{v}_{pqr}\right)|\theta\right)$. An evaluation of the competing models is again carried out by comparing AIC which is defined

by

$$\mathrm{AIC} = -2 \max l' \left(C \left(\boldsymbol{v}_{pqr} \right) | \theta \right) + 2 \dim \theta. \tag{27}$$

In this framework, AIC defined by Eq. (27) is sometimes called ABIC (Akaike Bayesian Information Criterion [2]). In this approach, an evaluation of model is conducted by considering two distribution of $C\left(\boldsymbol{v}_{pqr}\right)$ and $f\left(\boldsymbol{v}_{pqr}\right)$, namely a Bayesian approach. The low count around $\boldsymbol{v} = 0$ yields a small "weight" so that the fitting of $f\left(\boldsymbol{v}_{pqr}\right)$ is allowed to be done less precisely around $\boldsymbol{v} = 0$. Moreover, a Bayesian approach will treat a problem of a heavy tail in the high-velocity range. Since the count of ions of high-velocity is also small, modeling and model comparison with a smaller effect of the heavy tail will be tractable.

7 Conclusion

We have proposed a method for analyzing a multi-peaked velocity distribution via a multivariate Maxwellian mixture model. With the fitting of this model, we can extract the velocity moments for each component of the multiple peaks automatically. The parameters of the model are determined through the EM algorithm. For the automatic judgment of the preferable number of components of the mixture model, we introduced a method of examining the number of extrema of the resulting mixture model. When we use the method, we can adopt an appropriate fitting result and obtain a tool for a kinetic description of the plasma dynamics; this is especially effective for dealing with a large data set. Application of our method to observations confirmed that the method works well as shown in Fig. 2. More scientific application is presented in our recent work [12].

Acknowledgment. We would like to thank T. Mukai for providing us with Geotail/LEP data. This work was carried out under the auspices of JSPS Research Fellowships for Young Scientists.

References

1. Akaike, H.: A New Look at the Statistical Model Identification. IEEE Trans. Autom. Control **AC-19** (1974) 716–723
2. Akaike, H.: Likelihood and the Bayes procedure. In: Bernardo, J. M., De Groot, M. H., Lindley, D. V., Smith, A. F. M. (eds.): Bayesian Statistics. University Press, Valencia Spain (1980) 1–13
3. Dempster, A. P., Laird, N. M., Rubin, D. B.: Maximum Likelihood from Incomplete Data via the *EM* Algorithm. J. Roy. Statist. Soc. B **39** (1977) 1–38
4. Ishiguro, M., Sakamoto, Y.: A Bayesian Approach to Binary Response Curve Estimation. Ann. Inst. Statist. Math. **35** (1983) 115–137
5. McLachlan, G. J., Krishnan, T.: The EM Algorithm and Extensions. Wiley Series in Probability and Statistics. John Wiley and Sons, New York (1997)

6. McLachlan, G. J., Peel, D.: Finite Mixture Models. Wiley Series in Probability and Statistics. John Wiley and Sons, New York (2000)
7. Nakamura, N., Konishi, S., Ohsumi, N.: Classification of Remotely Sensed Images via Finite Mixture Distribution Models (in Japanese with English abstract). Proceedings of the Institute of Statistical Mathematics **41** (1993) 149–167
8. Sakamoto, Y., Ishiguro, M., Kitagawa G.: Akaike Information Criterion Statistics. KTK Scientific Publishers. D. Reidel, Tokyo (2000)
9. Schwarz, G.: Estimating the Dimension of a Model. Ann. Statist. **6** (1978) 461–464
10. Tanabe, K., Sagae, M.: An Exact Cholesky Decomposition and the Generalized Inverse of the Variance-Covariance Matrix of the Multinominal Distribution, with Applications. J. R. Statist. Soc. B. **54** (1992) 211–219
11. Ueno, G., Nakamura, N., Higuchi, T., Tsuchiya, T., Machida, S., Araki, T: Application of Multivariate Maxwellian Mixture Model to Plasma Velocity Distribution Function. In: Arikawa, S., Morishita, S. (eds.): Discovery Science. Lecture Notes in Computer Science, Vol. 1967. Springer-Verlag, Berlin Heidelberg New York (2000) 197–211
12. Ueno, G., Nakamura, N., Higuchi, T., Tsuchiya, T., Machida, S., Araki, T., Saito, Y., Mukai, T.: Application of Multivariate Maxwellian Mixture Model to Plasma Velocity Distribution Function. to appear in J. Geophys. Res. (2001)