We thank this referee for reviewing our manuscript and providing us the valuable comments. We reply to the referee's comments as follows and have made appropriate changes to the manuscript.

Moderate comments

Comment 1

I follow most of the method, but am not sure I have a correct understanding of the difference between d~ and d. At the end of 2.3 it says the finally obtained modelled data are shown as d~ and g~. Are they the same as d and g, but with noise added? If so this could be more explicitly stated.

Reply 1:

Yes, they are. $\tilde{\mathbf{d}}$ and $\tilde{\mathbf{g}}$ were made by just adding noise to **d** and **g**, respectively. We modified the last paragraph of the section 2.3 to make that clear, as shown below.

We added the noise to **d** and **g** and finally obtained modelled data, $\tilde{\mathbf{d}}$ and $\tilde{\mathbf{g}}$. Gaussian noise with a standard deviation of 5% of the electron density was added to the electron density data. The offset of 300 R was added to the gray level data and then Gaussian noise with a standard deviation of $\sqrt{\mathbf{g}+300}$ R were added. Figure 4a and 4b show the modelled ionospheric electron density that should be obtained by the EISCAT_3D radar and the modelled auroral images at five ALIS stations.

Comment 2

The description of the inversion from line 187 onwards could be explained more clearly. I think the main issue is the words "as shown below:" on line 187 – does that refer to equation 13, or all of section 3 after line 187? I suggest adding some words before equation 13 to explain what the equation is for, and/or reordering the description. Perhaps you could add a flow chart showing all of the steps for maximising equation 12, to help the reader to understand?

Reply 2:

We revised the description after line 180, as shown below. In addition, we summarized the flow of the inverse analysis at the end of this paragraph.

$$P(\tilde{\mathbf{b}}\,|\,\mathbf{f}) \propto \exp\left\{-\sum_j \frac{1}{2}\left(\tilde{\mathbf{b}}_j - \mathbf{b}_j(\mathbf{f})\right)^T \boldsymbol{\Sigma}_j^{-1}\left(\tilde{\mathbf{b}}_j - \mathbf{b}_j(\mathbf{f})\right)\right\} \tag{11}$$

where σ^2 is the variance of $\nabla^2\mathbf{f}$, $\boldsymbol{\Sigma}_j^{-1}$ is the inverse covariance matrix, and j means the kind of data. It was assumed that the modelled data are independent from each other, so $\boldsymbol{\Sigma}_j^{-1}$ has zero off-diagonal elements and the inverse of the variance of $\tilde{\mathbf{b}}_j$ in the diagonal elements. $\tilde{\mathbf{b}}_j$ corresponds to the modelled data $\tilde{\mathbf{g}}$ and $\tilde{\mathbf{d}}$ for j = 1 and 2, respectively, and they include the noise. $\mathbf{b}_j(\mathbf{f})$ corresponds to $\mathbf{g}$ and $\mathbf{d}$ in (7) and (8). Equation (10) indicates the smoothness constraint on $\mathbf{f}$ with respect to x, y, and E. In (11), it was assumed that the modelled data $\tilde{\mathbf{b}}$ has Gaussian errors. By substituting (10) and (11) into (9), $P(\mathbf{f}\,|\,\tilde{\mathbf{b}})$ is given by

$$P(\mathbf{f}\,|\,\tilde{\mathbf{b}}) \propto \exp\left[-\frac{1}{2\sigma^2}\left\{\sum_j w_j^2\left(\tilde{\mathbf{b}}_j - \mathbf{b}_j(\mathbf{f})\right)^T \boldsymbol{\Sigma}_j^{-1}\left(\tilde{\mathbf{b}}_j - \mathbf{b}_j(\mathbf{f})\right) + \left\|\nabla^2\mathbf{f}\right\|^2\right\}\right] \tag{12}$$

where w_j is a hyper-parameter, which is a constant corresponding to the weighting factor for each instrument data.

Maximization of the posterior probability is equivalent to minimization of the function inside the curly brackets of (12), which is given by

$$\varphi(\mathbf{f}; w_j) = \sum_j w_j^2\left(\tilde{\mathbf{b}}_j - \mathbf{b}_j(\mathbf{f})\right)^T \boldsymbol{\Sigma}_j^{-1}\left(\tilde{\mathbf{b}}_j - \mathbf{b}_j(\mathbf{f})\right) + \left\|\nabla^2\mathbf{f}\right\|^2. \tag{13}$$

Here, we define $\mathbf{r}(\mathbf{f}; w_1, w_2)$ as follows;

$$\mathbf{r}(\mathbf{f}; w_1, w_2) = \begin{pmatrix} w_1\boldsymbol{\Sigma}_1^{-\frac{1}{2}}\left(\tilde{\mathbf{g}} - \mathbf{g}(\mathbf{f})\right) \\ w_2\boldsymbol{\Sigma}_2^{-\frac{1}{2}}\left(\tilde{\mathbf{d}} - \mathbf{d}(\mathbf{f})\right) \\ \nabla^2\mathbf{f} \end{pmatrix}. \tag{14}$$

Then, (13) is given by

$$\varphi(\mathbf{f}; w_1, w_2) = \left\|\mathbf{r}(\mathbf{f}; w_1, w_2)\right\|^2. \tag{15}$$

Here, we change the variables by letting $\mathbf{f} = \exp(\mathbf{x})$ to take advantage of the non-negative constraint of **f** (i.e., $\mathbf{f} \geq \mathbf{0}$). Then, the minimization of $\varphi(\mathbf{x}; w_1, w_2)$ becomes a non-linear least squares problem with respect to $\mathbf{x}$, so we solved it by the Gauss-Newton algorithm.

In the Gauss-Newton method, the parameter $\mathbf{x}$ proceeds by the iteration, $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \Delta\mathbf{x}^{(k)}$, where the increment $\Delta\mathbf{x}^{(k)}$ at the k^{th} step is a solution of the following equation:

$$\left(\mathbf{J}^T(\mathbf{x}^{(k)})\mathbf{J}(\mathbf{x}^{(k)})\right)\Delta\mathbf{x}^{(k)} = -\mathbf{J}^T(\mathbf{x}^{(k)})\mathbf{r}(\mathbf{x}^{(k)}), \qquad (16)$$

where $\mathbf{J}(\mathbf{x})$ is the Jacobian matrix of $\mathbf{r}(\mathbf{x})$ with respect to $\mathbf{x}$. Since Eq. (16) is a normal equation with a large sparse matrix, we solved it by the Conjugate Gradient (CG) method.

The initial values $\mathbf{x}^{(0)}$ was obtained in advance from only gray level data $\tilde{\mathbf{g}}$ by solving $\min[\varphi(\mathbf{f})]$ with respect to **f**. We solved the linear least squares problem (i.e., $\min[\varphi(\mathbf{f})]$) by the Simultaneous Iterative Reconstruction Technique (SIRT) method (Aso et al., 1998) with $\mathbf{f}^{(0)} = 10^7$ [m^{-2} s^{-1} eV^{-1}] and used the solution $\mathbf{f}^*$ for the initial value of the Gauss-Newton algorithm (i.e., $\mathbf{x}^{(0)} = \log(\mathbf{f}^*)$). The hyper-parameters (w_1, w_2) were determined by using the 5-fold cross-validation (Stone, 1974).

The flow of the inverse analysis is summarized as follows.

1. Calculate the initial value of **x**, $\mathbf{x}^{(0)}$. $\mathbf{x}^{(0)}$ is given by $\mathbf{x}^{(0)} = \log(\mathbf{f}^*)$ where **f*** is the solution of $\min[\varphi(\mathbf{f})]$, which is solved by the SIRT method. Only auroral images are used for this step (i.e., $\varphi(\mathbf{f}) = \left\|\boldsymbol{\Sigma}_1^{-1/2}(\tilde{\mathbf{g}} - \mathbf{g}(\mathbf{f}))\right\|^2$) and the initial value of **f** is set to 10^7 [m^{-2} s^{-1} eV^{-1}].
2. Determine the hyper-parameters (w_1, w_2) so as to minimize $\varphi(\mathbf{x}; w_1, w_2)$ (= $\left\|\mathbf{r}(\mathbf{x}; w_1, w_2)\right\|^2$) by the 5-fold cross-validation method. In this step, the values of w_1 and w_2 are selected from pre-created lists, and the same algorithm as shown in the step 3 is used to solve $\min[\varphi(\mathbf{x}; w_1, w_2)]$.
3. Solve $\min[\varphi(\mathbf{x}; w_1, w_2)]$ with w_1 and w_2 determined in the step 2 by the Gauss-Newton algorithm. In the Gauss-Newton algorithm, **x** proceeds by the iteration, $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \Delta\mathbf{x}^{(k)}$, where $\Delta\mathbf{x}^{(k)}$ is obtained by solving the normal equation $\left(\mathbf{J}^T(\mathbf{x}^{(k)})\mathbf{J}(\mathbf{x}^{(k)})\right)\Delta\mathbf{x}^{(k)} = -\mathbf{J}^T(\mathbf{x}^{(k)})\mathbf{r}(\mathbf{x}^{(k)})$ by the Conjugate Gradient method. The

reconstructed differential flux is obtained by substituting the solution into $\mathbf{f} = \exp(\mathbf{x})$.

Comment 3

The study found that ACT underestimates the electron density and electron flux in this case. Could you comment in the discussion on why that might be? Is it expected to be a common situation, or would the density/flux be overestimated just as often?

Reply 3:

As for the multiple arcs assumed in this study, the total energy flux of the precipitating electrons reconstructed by ACT were underestimated inside the discrete arcs. However, different conditions such as the relative position of the aurora and the imagers, the noise level, and the shape of the aurora cause the reconstructed electron flux to be overestimated. I revised the paragraph from line 308, as shown below.

Since the auroral images usually include observational noise, it is often difficult to reconstruct the auroral 3D distribution precisely by using the ACT method. As for the multiple arcs assumed in this study, the total energy flux of the precipitating electrons reconstructed by ACT was underestimated inside the discrete arcs. This is because that the two neighboring arcs overlapped when viewed from several imagers and were difficult to perfectly separate. In Figure 5b, the reconstructed electron flux between the arcs was greater than the modelled flux, and instead, the flux inside the arcs decreased. When the multiple arcs overlapped from all imagers, it was quite difficult for the ACT to distinguish them from each other (Figure 8). Of course, the reconstruction result depends on the condition such as the relative position of the aurora and the imagers, the noise level, and the shape of the aurora, and different conditions cause the reconstructed electron flux to be overestimated. We demonstrated that the G-ACT can significantly reduce the reconstruction errors caused by the ACT.

Minor comments

Is there a reason why you use x for the approximately north-south direction and y for the approximately east-west direction? It would be natural to me to name the directions the other way around, and then the axes in figure 3a would be more conventional.

The coordinate system with x for the magnetically north-south direction and y for the east-west direction was used by Tanaka et al. [2010]. Since we followed the analysis method described in Tanaka et al. [2010], we would like to use the same coordinate system.

Line 40 is missing a closing parenthesis).

We modified it.

Line 48/49: "It can measure… even though they can detect…" - This sentence mixes singular "it" (an optical imager) with plural "they", so needs fixing.

We replaced "they" with "it".

Line 62: I believe the set of authors is not identical between the submitted paper and the papers cited on this line, and for clarity I suggest rewording to remove the use of "we", e.g. "Aso et al., 2008 and Tanaka et al., 2011 extended the ACT method to generalized ACT (G-ACT)."

We revised this sentence according to the reviewer's suggestion.

Line 81: Here (and other places) "field of views" should be "fields of view". Also "each instrument's field of views" should be "the instruments' fields of view" (other wordings possible but the current wording is not quite correct).

We revised them according to the reviewer's comment.

Line 93: missing unit "pixels" after 256 x 256.

We added "pixels" after 256 x 256.

Line 105: "size" should be plural "sizes".

We modified it.

Line 168: I think you are adding Gaussian noise to the auroral images, but I don't think this is explicitly stated, and should be.

We clearly stated that Gaussian noise was added to the auroral images, as shown in Reply 1.

Figure 5: Did you try plotting this with a log scale for the color axis? It might show the electron flux between the arcs produced by the ACT method more clearly.

We tried plotting Figure 5 with a log scale for the color axis. However, the plots emphasize too much the region where the total energy flux is small and we do not focus on. Thus, we left Figure 5 as it is and added the following description to the paragraph from L205.

It appears that Q_0 was reconstructed well; however, there are two points to be noted: one is an underestimation of Q_0 at the peak location of each discrete arc and the other is an overestimation between the two arcs. The energy flux at the center of the reconstructed arcs is slightly smaller than the input flux. On the other hand, the energy flux between the two arcs is greater than the input flux, particularly at $y < 0$. For example, Q_0 at (x, y)=(45km, -20km) is 1.47 mW/m^2 for the input flux and 7.30 mW/m^2 for the reconstructed one by the ACT. The Q_0 at the location was improved to 4.36 mW/m^2 (2.29 mW/m^2) by the G-ACT method with the electron density from 10×10 beams (21×21 beams) of the EISCAT_3D radar.

Line 217: "better improved" could be "more accurate".

We modified it. In addition, we calculated the Mean Absolute Error (MAE) or the Mean Absolute Percentage Error (MAPE) to quantify the performance of the reconstruction methods and added these values to Figures 5-9 and the text. Please see "Reply on RC1" more in detail.

Line 241: I suggest "especially" instead of "significantly" (last word of the line).

We modified it.

Line 271: "the both two arcs" should just be "both arcs".

We modified it.

Line 318: Could you comment on how the radar temporal resolution (scan time over all *beams) compares to the optical exposure time? Is it relevant?*

The temporal resolution of optical imager depends on the performance of the imager, the wavelength of the filter, the auroral emission intensity, etc. Since the monochromatic images are required for the G-ACT analysis, the temporal resolution of high-sensitivity imagers (e.g., electron-multiplying CCD (EMCCD) imagers) with the monochromatic filters is usually a few seconds and can be higher than that of the 10×10 beam scan of the radar. For example, Fukizawa et al. (2022) reconstructed the 3D distribution of pulsating aurora every 2 seconds by the ACT using the 427.8-nm auroral images.
We added the above description in Discussion.

Line 330: "increasing electron density" should be "the electron density increases".

We modified it.