Discussion

Emranul Sarkar et al.

December 22, 2020

1 Data selection process

Following reviewer's recommendation, we will make adequate data selection in the revised manuscript.

Data will be selected as:

1. Only data within the contour line 0.4 will be selected to remove bad data in the outer contour area. This will roughly correspond to the altitude range 85–95 km.
2. Zenith angle selection will be now within 50° (previously it was 60°)

All figures, table and numerical values will be updated accordingly, and shown in Appendix A.

2 Derivation in the Appendix

We apologise for few typos (line 513 and 526) in the Appendix. The statement *such that the expectation value of* $\mu^{'}$*, or* $< \mu^{'} > \neq 0$ should be *such that the expectation value of* $\mu^{'}$*, or* $< \mu^{'} >= 0$ (this is from Carroll and Rupper(1996)). The derivation is repeated below with additional explanation:

1. Motivation:

   The GM solution in the normalised coordinate is given by -

   $$\beta_W = \frac{\sqrt{s_{h'}}}{\sqrt{s_{d'}}} = 1 \tag{1}$$

   because $s_{h'} = s_{h'} = 1$ due to the normalisation with Eq. (18) (manuscript). We found out that this solution is over-estimated, and thus positively biased. Mathematically, this means we can correct for this bias by multiplying the $h^{'}$ with a scaling factor ν, such that $\nu \lesssim 1$ (this will reduce the numerator in the above equation).

   In other words, this reweighted normalised variable is now redefined as $\nu h^{'}$ or simply h^{*}. But this is just a mathematical treatment, the observed value of $h^{'}$ will be the same observed value. The difference between the variance of our hypothetical variable h^{*} and the real observable $h^{'}$ is some constant, say $s_{\mu'}$ (where $\mu^{'}$ is $h^{'} - h^{*}$),

   $$s_{h^*} = s_{h'} - s_{\mu'} = 1 - s_{\mu'} \tag{2}$$

2. But Eq. (2) above also implies,

   $$h^{'} = h^{*} + \mu^{'} \tag{3}$$

   In the manuscript, we have defined $h^{'}$ in line-139 as (but now in normalised coordinate, hence all with *prime* subscript, and $\alpha = 0$),

$$h_i' = \beta_W \xi_i' + \varepsilon_i' = \beta_W (d_i' - \delta_i') + \varepsilon_i' \tag{4}$$

Substitute (3) in (4) to get,

$$h_i^* + \mu_i' = \beta_W^{adj} \xi_i' + \varepsilon_i' = \beta_W^{adj} (d_i' - \delta_i') + \varepsilon_i' \tag{5}$$

Note that now we write the normalised slope as β_W^{adj}, where $\beta_W^{adj} \lesssim \beta_W$ (of course, $\beta_W = 1$). Equation (5) above is Eq. (A1) in the manuscript (line-512).

3. Solution: Assume $\beta_W^{adj} \approx \beta_W$

Write Eq. (5) as -

$$h_i^* - \beta_W^{adj} d_i' + \mu_i' = -\beta_W^{adj} \delta_i' + \varepsilon_i' \tag{6}$$

And so,

$$(\mu_i')^2 \approx (\varepsilon_i' - \delta_i')^2 \tag{7}$$

Equate the residual function of GM solution (line 251) with its perturbed value as (see Fig. 1 for a bit of visual understanding)-

$$\sum_{i=1}^{N} \frac{(h_i^* - \beta_W^{adj} d_i')^2}{1 + (\beta_W^{adj})^2} \approx \sum_{i=1}^{N} \frac{(h_i' - \beta_W d_i')^2}{1 + \beta_W^2} \tag{8}$$

Substitute Eq. (6) and Eq.(7) in Eq. (8), and with $\beta_W^{adj} \approx \beta_W = 1$, the left-hand side of Eq.(8)

$$\sum_{i=1}^{N} \frac{(-\delta_i' + \varepsilon_i' - \mu_i')^2}{2} \approx \sum_{i=1}^{N} \frac{(-\delta_i' + \varepsilon_i')^2 + (\mu_i')^2}{2} \approx \sum_{i=1}^{N} (\mu_i')^2 \tag{9}$$

Or Eq. (8) is now,

$$\sum_{i=1}^{N} (\mu_i')^2 = \sum_{i=1}^{N} \frac{(h_i' - d_i')^2}{2} \tag{10}$$

Divide by N and write in absolute form because of the 'square' dependence between LHS and RHS of Eq. (10)

$$\frac{1}{N} \sum_{i=1}^{N} \mid \mu_i' \mid = \frac{1}{N} \sum_{i=1}^{N} \sqrt{\frac{(h_i' - d_i')^2}{2}} \tag{11}$$

Bienaymé formula[1] states that for uncorrelated random variable Y_i :

$$Var\left(\sum_{i=1}^{N} Y_i\right) = \sum_{i=1}^{N} Var(Y_i) \tag{12}$$

Or,

[1] https://en.wikipedia.org/wiki/Variance

$$Var\left(mean(Y_i)\right) = Var\left(\frac{1}{N}\sum_{i=1}^{N} Y_i\right) = \frac{1}{N^2} Var\left(\sum_{i=1}^{N} Y_i\right) = \frac{1}{N} Var(Y_i) \tag{13}$$

Apply the formula in Eq. (13) to Eq. (11)

$$Var\left(\frac{1}{N}\sum_{i=1}^{N} \mid \mu_i^{'} \mid\right) = Var\left(\frac{1}{N}\sum_{i=1}^{N} \sqrt{\frac{(h_i^{'} - d_i^{'})^2}{2}}\right) \tag{14}$$

Or,

$$Var\left(\mu_i^{'}\right) = Var\left(\sqrt{\frac{(h_i^{'} - d_i^{'})^2}{2}}\right) \tag{15}$$

From Eq. (2), we therefore have,

$$\nu^2 = s_{h^*} = 1 - Var\left(\sqrt{\frac{(h_i^{'} - d_i^{'})^2}{2}}\right) \tag{16}$$

An alternative, and perhaps much straight-forward way to go from Eq. (11) to Eq. (15) is to utilise the property of half-normal distribution[2]. This states that if Y follows an ordinary normal distribution, with mean 0 and variance σ^2, then $X =\mid Y \mid$ follows a half-normal distribution, such that the variance of X is given by-

$$Var(X_i) = \sigma^2(1 - \frac{2}{\pi}) \tag{17}$$

And their expectation value is related as-

$$E(X) = \sigma\frac{\sqrt{2}}{\pi} \tag{18}$$

Since μ_i (left-hand side) and the residuals from GM solution (right-hand side) in Eq. (11) are assumed to have zero mean and normally distributed, their absolute value follows half-normal distribution. Applying the formula in Eq. (18) to Eq. (11) directly lead to Eq. (15).

Note:

- How is Eq. (5) related to the statistical model in line-305 of the manuscript?

  We have originally defined the normalised variable $h_i^{'}$ (observable) in terms of hypothetical true value $\eta_i^{'}$ in line 139, for which the normalised slope is β_W:

$$h_i^{'} = \eta_i^{'} + \varepsilon_i^{'} \tag{19}$$

  According to the model with 'equation error' in line-305, we need to replace $\eta_i^{'}$ by $\eta_i^{'} - \mu_i$ so that the variance of observed height $h_i^{'}$ is reduced to that of h^*.

$$h_i^* = \eta_i^{'} - \mu_i + \varepsilon_i^{'} \tag{20}$$

[2] https://en.wikipedia.org/wiki/Half-normal_distribution

Following reviewer comment, we have realised that this information above is quite redundant since Eq. (5) didn't specifically require this. In the revised manuscript, we will use adequate data selection and hence such extended discussion on 'natural variation' or 'equation error' (Sect. 3.2.2) is no longer necessary. Instead, we will simply refer to s_μ as asymmetric error.

- Numerical validation of the equality in Eq. (8) above:

  For the date 14 Nov 2015, the RHS of Eq. (8) is -

$$\sum_{i=1}^{N} \frac{(h_i^{'} - d_i^{'})^2}{2} \approx 85.87 \tag{21}$$

  In the current manuscript we have $\beta_W^{adj} \approx 0.75$ and $\nu \approx 0.87$, and so the LHS of Eq. (8) is -

$$\sum_{i=1}^{N} \frac{(h_i^{*} - \beta_W^{adj} d_i^{'})^2}{1 + (\beta_W^{adj})^2} \approx 75.50 \tag{22}$$

  In the revised manuscript (with adequate bad data removal) we have $\beta_W^{adj} \approx 0.95$ and $\nu \approx 0.96$ (see Fig. A.8), and now the LHS of Eq. (8) is -

$$\sum_{i=1}^{N} \frac{(h_i^{*} - \beta_W^{adj} d_i^{'})^2}{1 + (\beta_W^{adj})^2} \approx 82.35 \tag{23}$$

- *This manuscript do not make the default assumption that $\lambda = 1$ in the normalised coordinate is indeed the fundamental property of this distribution, and hence the motivation to develop a solution that allows for small variation in λ from 1. This is because the exact value of λ depends on the data selection process (and also probably on the instrument, but this study is just with one radar).*

- In the revised manuscript we will state the 3 possible cases:

  1. $\lambda = 1$. The GM solution is valid.
  2. $\lambda \gtrsim 1$. Eq. (37) is valid.
  3. $\lambda \lesssim 1$. Eq. (37) is valid but with h and d interchanged.

3 Specific

3.1 This implies that the core distribution could be a good way to get a better GM solution although the GM solution do not explicitly show up in Eq. (35) and Eq. (37)

.

Eq. (35) and Eq. (37) are indeed showing this convergence. Note that these equations are given in normalised coordinates ($d_i^{'}$ and $h_i^{'}$). Now, if we look closely in Fig. 1b (manuscript) or Fig. A.1b here, we see that as we approach the core of this distribution, both $d_i^{'}$ and $h_i^{'}$ gets smaller and converges to zero-point. This means both Eq. (35) and Eq. (37) converges to 1 near the peak. This is demonstrated in Fig. 1 (below) where we have estimated the slope for GM solution and with Eq. (37) using 6 different data selection criteria. The stability of the solution from Eq. (37) arises due to the correction term ν which is a constant that depends on the input data via Eq. (16) above (or, Eq. (35) in the manuscript).

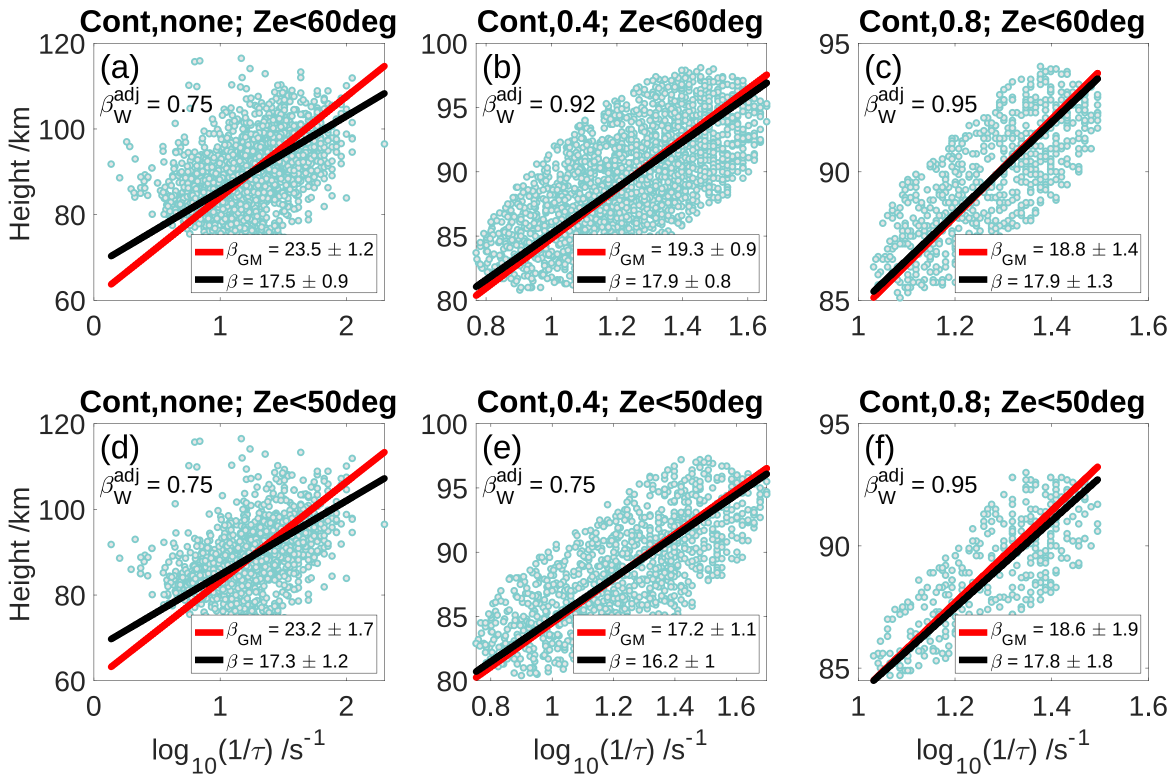

Figure 1: Comparison of the slope estimate of GM solution and with Eq. (37) for the date 14 Nov 2015 with six different data selection criteria. The contours (or Cont in figure above) are from Fig. 1b in the manuscript.

3.2 The authors think that the distribution in these orange areas are affected by natural variation and should not be rejected in a arbitrary way

.

In line 329 we have specifically stated that "*we therefore no longer stringently distinguish between different types of error in this data. In other words, the error variances, $s_{\delta'}$ and $s_{\varepsilon'}$, now consist of both measurement errors and partly the natural geophysical variation.*"

The word '*arbitrary*' in line-290 was referring to the way we have defined these contour lines. In the revised manuscript we will rephrase the word 'arbitrary' and perform adequate data selection.

3.3 The dates at the tick marks seem shifted by 15 days in Fig. 2, 3 and 6 axis

The ticks are labelled as first day of month and year (2015 or 2016). We will add X-axis label in the revised manuscript to clarify this.

3.4 Line 276-277 "the standard errors in these temperatures, which is on average 19 K": Is this a value estimated using using season of the CORAL data

No. This value is estimated using Bootstrap analysis as explained in Sect. 3.2.3 (we used the MATLAB function *bootci*). This relatively large value of 19K is mainly due to low N with this radar (since standard error is inversely related to $\sqrt{N}$).

3.5 Line 366-267 "does this mean that the resampling was made 20000 times for every 24 hr"

Yes (line 363).

3.6 Line 342 "Two of the four ν redundant?"

No (unless, we misunderstood this question).

3.7 Since it is the decay time that is mostly affected in the lower and upper distribution rather than height as seen in Fig 1, it seems natural to use a correction term to decrease the effective variance of d_i instead that of h_i. Is such an approach possible? I presume it will give an equivalent result.

Looking at Eq. (1) and (2) above, this implies that now it will be something like,

$$s_{d^*} = s_{d'} + s_{\mu'} = 1 + s_{\mu'} \tag{24}$$

The 'positive' sign on the right indicates that we want to increase the effective variance of d_i to account for the over-estimate in GM solution. Note that Eq. (7) and Eq. (8) above in the the derivation of $s_{\mu'}$ will be still the same if correction term was applied to d_i.

On the other hand, if other authors find that for their data the GM solution is under-estimate, Eq. (37) will be still valid but with h and d interchanged.

In Sect. 3.1 of the manuscript, we have demonstrated a simple way to find out the direction of asymmetric error effect in this data as follows:

1. First we have performed two OLS fittings with h and d as independent variable and estimated temperatures. Comparison of this temperature with lidar data resulted in the histograms shown in Fig. 3b in the manuscript (or, Fig. A.3b below). We see that the blue histogram is located much farther to right side of origin than the green one. This asymmetric location around the origin is a clear indication that the overall effect of asymmetry in this data is along the Y-axis. Likewise, the geometric mean of the green and the blue is the red histogram which is located on the right-hand side of the origin, thereby indicating that $\lambda > 1$.
2. To confirm the observation above, next we have applied SCT calibration (and Eq. (17) and Eq. (18) in the manuscript) to obtain an experimental estimate of the error variance in this data. This is reported in line 246-247, where we have stated that *"the mean values of $s_{\varepsilon'}$ and $s_{\delta'}$ are found to be 0.64 ± 0.04 and 0.38 ± 0.06 respectively"* (where $s_{\varepsilon'}$ and $s_{\delta'}$ are error variances of normalised height and $log_{10}(1/\tau)$). Because $s_{\varepsilon'} > s_{\delta'}$, the overall direction of asymmetric error must be along the h data.
3. Finally in Sect. 4, we have demonstrated another method to obtain a numerical estimate of λ. This was done by equating Eq. (26) with Eq. (37). This method is completely independent since it didn't require lidar data, and yet agreed very well with it (Table 1)

How does the analysis discussed above change when we improve the data selection criteria as mentioned in Sect. 1 above?

After applying an improved data selection criteria, we have repeated the full analysis in the manuscript. The figures and table are given in appendix A.

In Fig. A.3, we still see that the *blue* and the *green* histograms are asymmetrically located around the origin. This implies that the effect of error in height data in more than that in $log_{10}(1/\tau)$[3]. The results from SCT analysis is given in Table A.1, showing that the error variances in normalised height and $log_{10}(1/\tau)$ are unequal, and $s_{\varepsilon'} > s_{\delta'}$. The SCT estimate of the effective value of λ is now 1.24, and thus $\lambda > 1$ with the consequence of over-estimation in GM solution.

[3]This value has decreased now (Table A) as we got rid of those annoying region as indicated by the reviewer.

Comparison with lidar data is shown in Fig. A.7. Although the GM solution has improved significantly, the adjusted GM solution is still a better estimate for the following reason-

There is an additional source of uncertainty in this method that we haven't discussed explicitly. The temperature gradient taken from MSIS model values are on average $-2.4K/km$ for this season. The temperature gradient estimated by other authors with real observation indicates that the temperature gradient during winter time is on average $-1.5K/km$. Hence, with a more realistic temperature-gradient value, the histograms in Fig. A.7a and Fig. A.7c will shift to the right by another 7–10 K. This will increase the offset in GM temperature, whereas the temperature with adjusted GM solution will get even better. In fact, the temperature estimated by Eq. (37) will be always better or same as GM solution, since an exact match of $\lambda = 1$ is unlikely. However, whether this offset in GM temperature is significant or not must be determined by respective investigators depending on their own science goal.

3.8 A strict mathematical treatment of μ_i is beyond my understanding. But is such a practical approach that a constant value of $s_{\mu'}$ represents the while equation error mathematically acceptable? Or simply practical?

Variable $s_{\mu'}$ is of course plausible, but in that case the system will no longer admit an analytic solution (line 215-216).

Note that $s_{\mu'}$ is not an arbitrary constant, but rather a constant which depends directly on the input data.

We understand that this over-emphasize on 'equation error' and 'natural variation' in the current version of the manuscript has unnecessarily complicated the physical meaning of $s_{\mu'}$. In essence, $s_{\mu'}$ is just a measure of 'bias' in GM solution as a result of asymmetric error in the data. Remember that GM solution, by its own definition, make the default assumption that the effect of error in normalised h and d are exactly equal. In case of unequal value of $s_{\varepsilon'}$ and $s_{\delta'}$, that elliptical scattered plot will get tilted either in Y direction (if $s_{\varepsilon'} > s_{\delta'}$) or along the X-direction (if $s_{\varepsilon'} < s_{\delta'}$). But GM solution doesn't see that apparent tilting, and will always try to bisect this data symmetrically, hence either this solution is over-estimated or under-estimated. The idea of $s_{\mu'}$ is to correct for this tilt in the statistical sense.

3.9 What are bars over h and d in line 353?

The mean value of original h and d. We will add a line here to mention this.

3.10 Could you explain the more about "a priori knowledge" in line 358

The EIV slope (in normalised coordinate) is given by (Eq. (25) in the manuscript) -

$$\beta_W = \frac{s_{h'} - \lambda s_{d'} + \sqrt{(s_{h'} - \lambda s_{d'})^2 + 4\lambda s^2_{h'd'}}}{2 s_{h'd'}} \tag{25}$$

The main problem here is that we need to know the exact value of λ, otherwise we run into the risk of over-parameterization. For this data we have found that λ is of the order of 1. Now, we could have picked up that value of $\lambda = 1.7$[4] (which we have estimated using OLS+SCT calibration in Sect. 3.1) and directly use it in the equation above to estimate EIV slope (irrespective of data selection process since, in theory, the parameter λ is suppose to contain all error information anyway). But this would imply that we would always need lidar or satellite data to estimate these temperatures. Alternatively, we can stick to $\lambda = 1$, and instead introduced a correction term in the statistical model such that there is a negative sign in Eq. (35) to correct the 'positive bias' in GM solution. These two processes are equivalent as stated in line 291-296.

[4] with the improved data selection in the revised manuscript this value is now 1.24 as presented in Table A.1.

4 Summary

We thank the anonymous reviewer for providing interesting insight into this work and recommendation for further improvement. These will be addressed accordingly in the revised manuscript.

A Updated figures and table for the revised manuscript

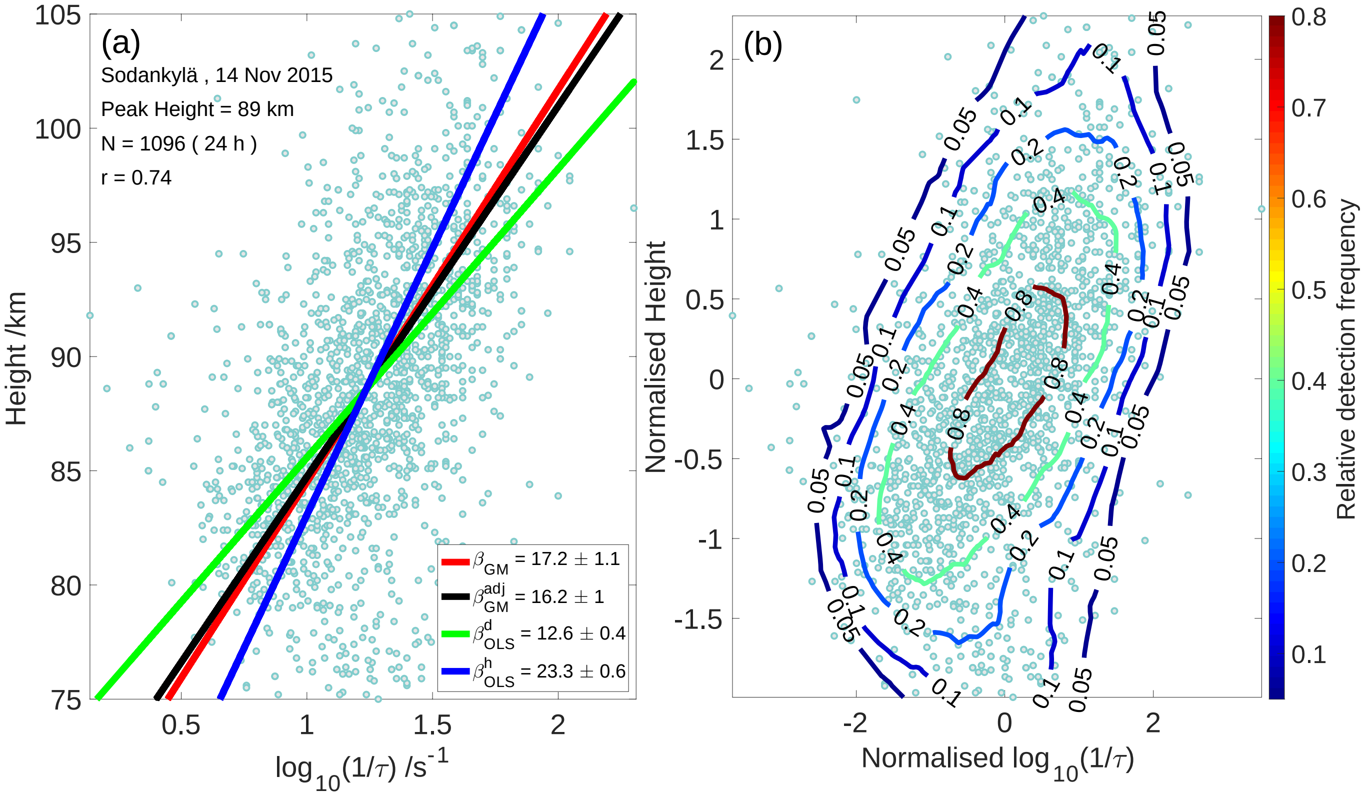

Figure A.1: (a) Typical scatter plot of $log_{10}(1/\tau)$ and height. The lines correspond to best fit models using different regression methods described in the text. The *green* and *blue* line corresponds to 'ordinary least-squares method (OLS)' with $log_{10}(1/\tau)$ and height as independent variable respectively. The *red* line correspond to the geometric mean (GM) of β^d_{OLS} and β^h_{OLS}. The *black* line is the bias-corrected slope obtained by using Eq. (37). (b) The bivariate distribution of the data. The measured height and $log_{10}(1/\tau)$ are converted to dimension free coordinates using Eq. (18). The relative density contours are obtained by counting the number of detections in a radius of 0.5 relative to the density at the height of peak meteor occurrences at the center. All slopes in (a) are estimated at contour level 0.4.

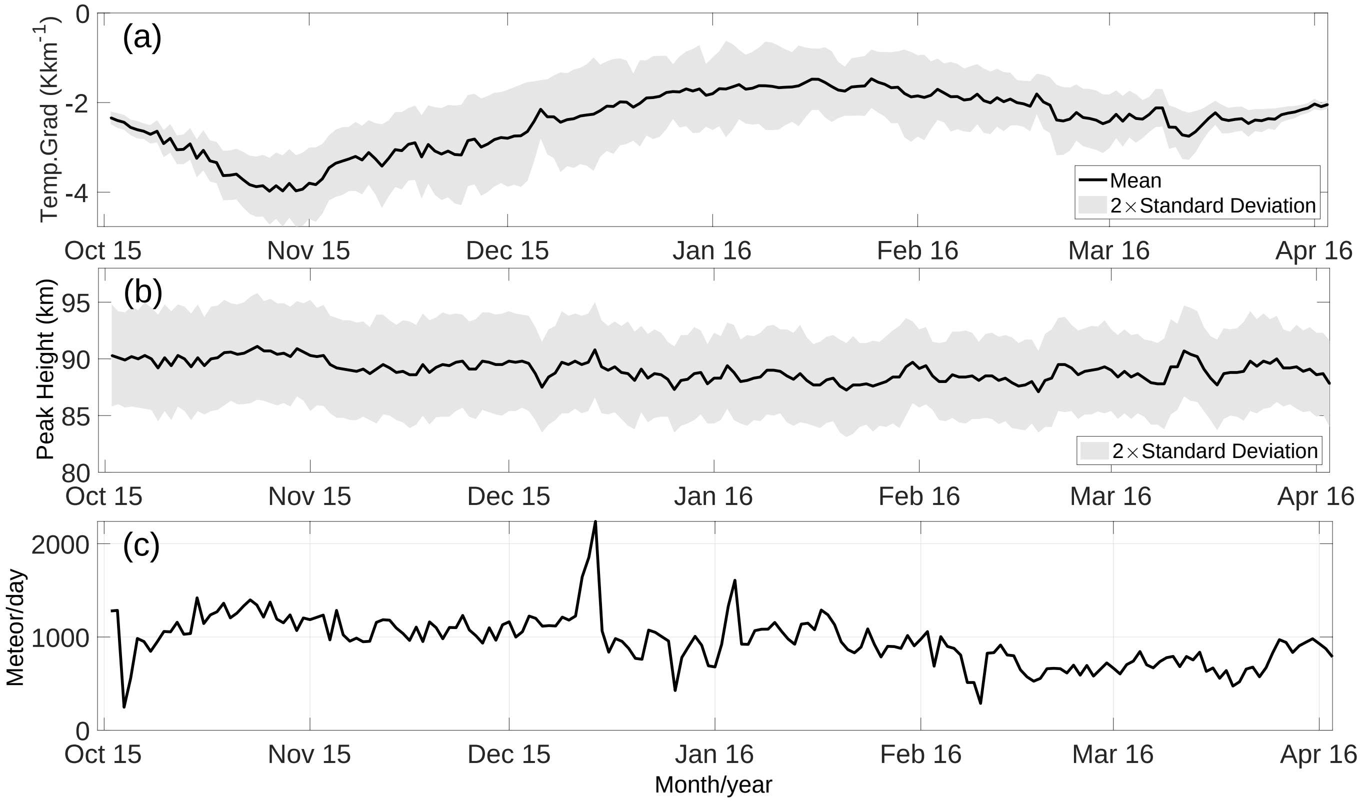

Figure A.2: (a) Temperature gradient model derived from MSIS90. (b) Peak meteor heights for the data used in this work, and (c) the daily meteor detection for zenith angle less than 50°, velocity in the range ±100 m/s and at contour level 0.4.

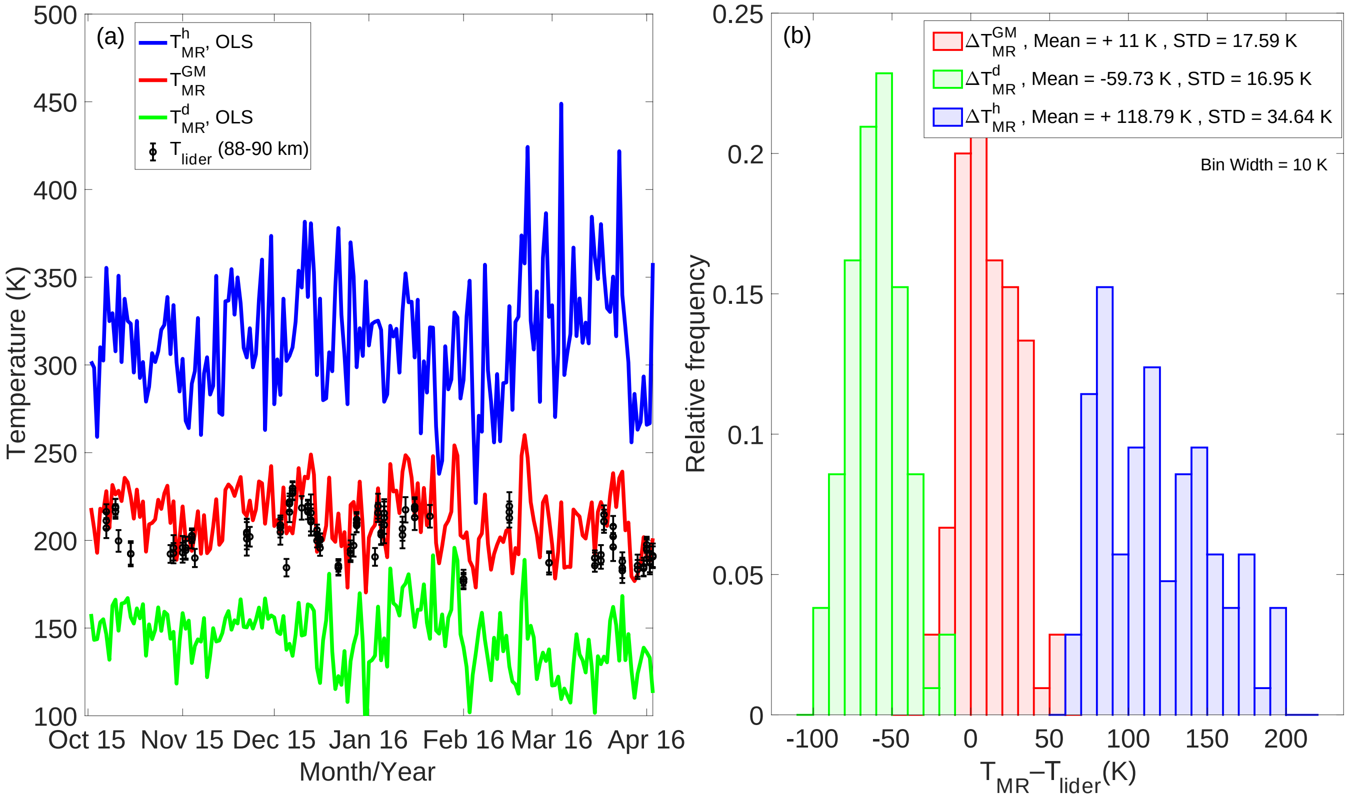

Figure A.3: (a) Temperature estimated in OLS method using $log_{10}(1/\tau)$ (*green*) and height (*blue*) as independent variable. Also, showing (*red*) the temperatures obtained using the geometric mean (GM) fitting. (b) The offset between the lidar (T_{lidar}) temperatures and the estimated MR temperatures (T_{MR}) using OLS fitting and GM fitting. .

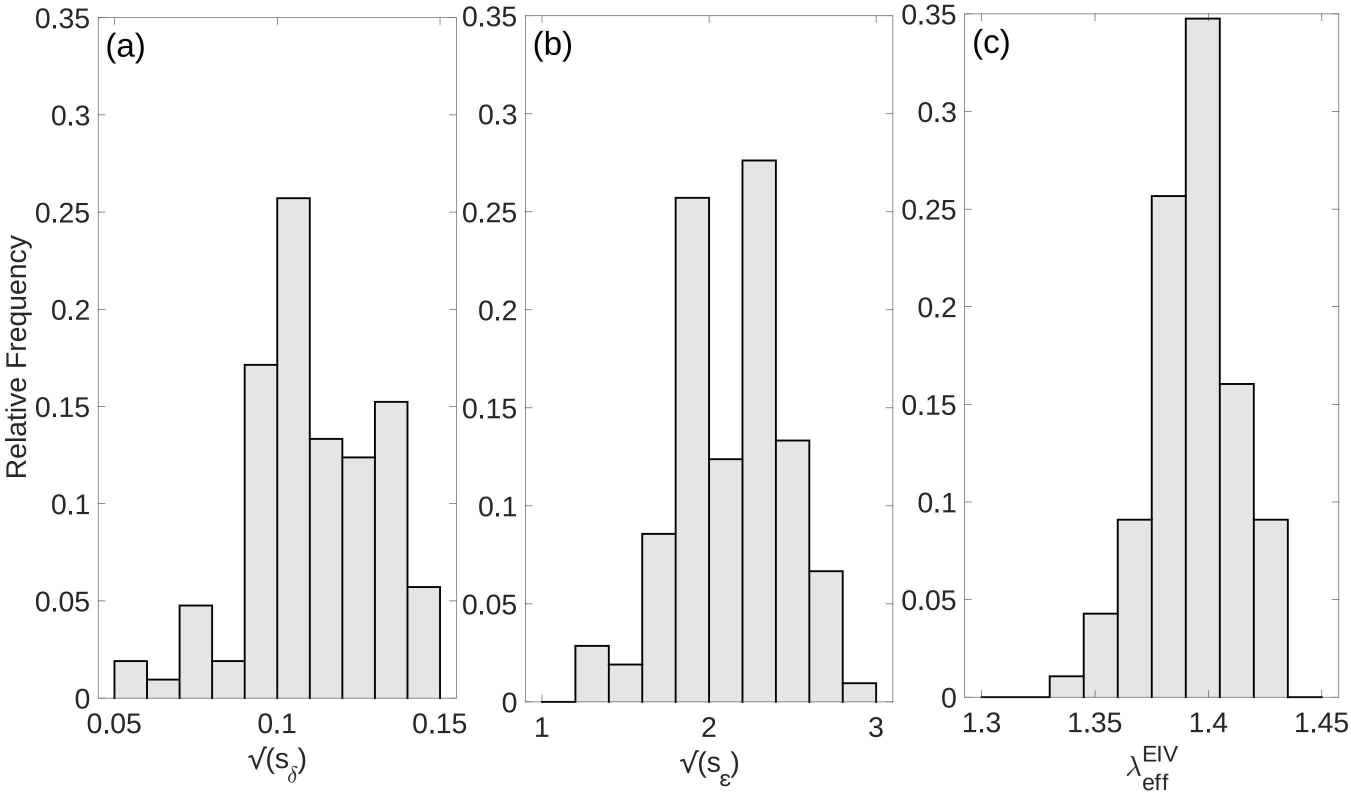

Figure A.4: Statistical distribution of (a) $\triangle log_{10}(1/\tau)$ and (b) $\triangle height$ obtained using SCT method with colocated lidar data. The variance ratios of these parameters in the normalised coordinate system is shown in (c). The properties of these histograms are presented in Table A.1.

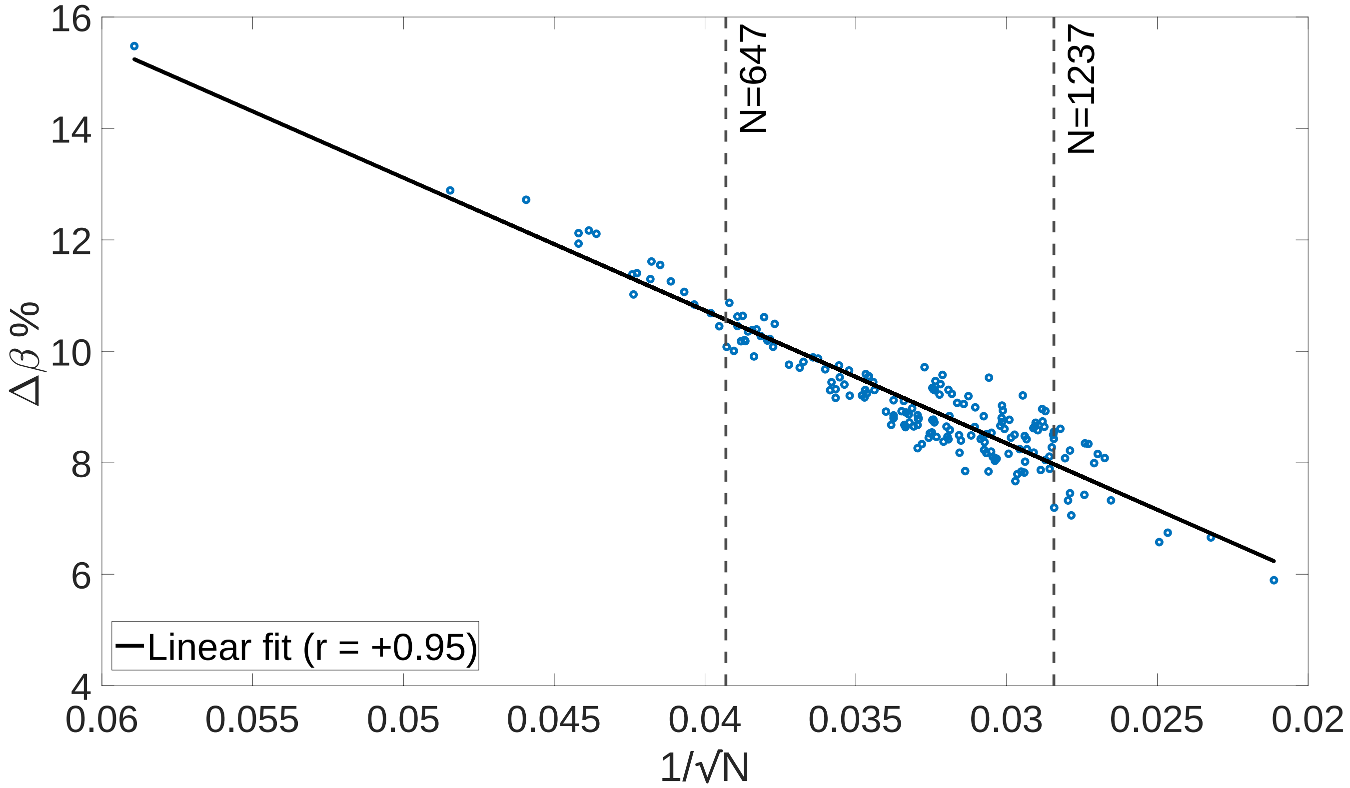

Figure A.5: Estimated percentage errors in the slope using bootstrap resampling. The two dashed lines indicate the range of N corresponding to 80% or 148 days out of the total of 185 days of data used in this work.

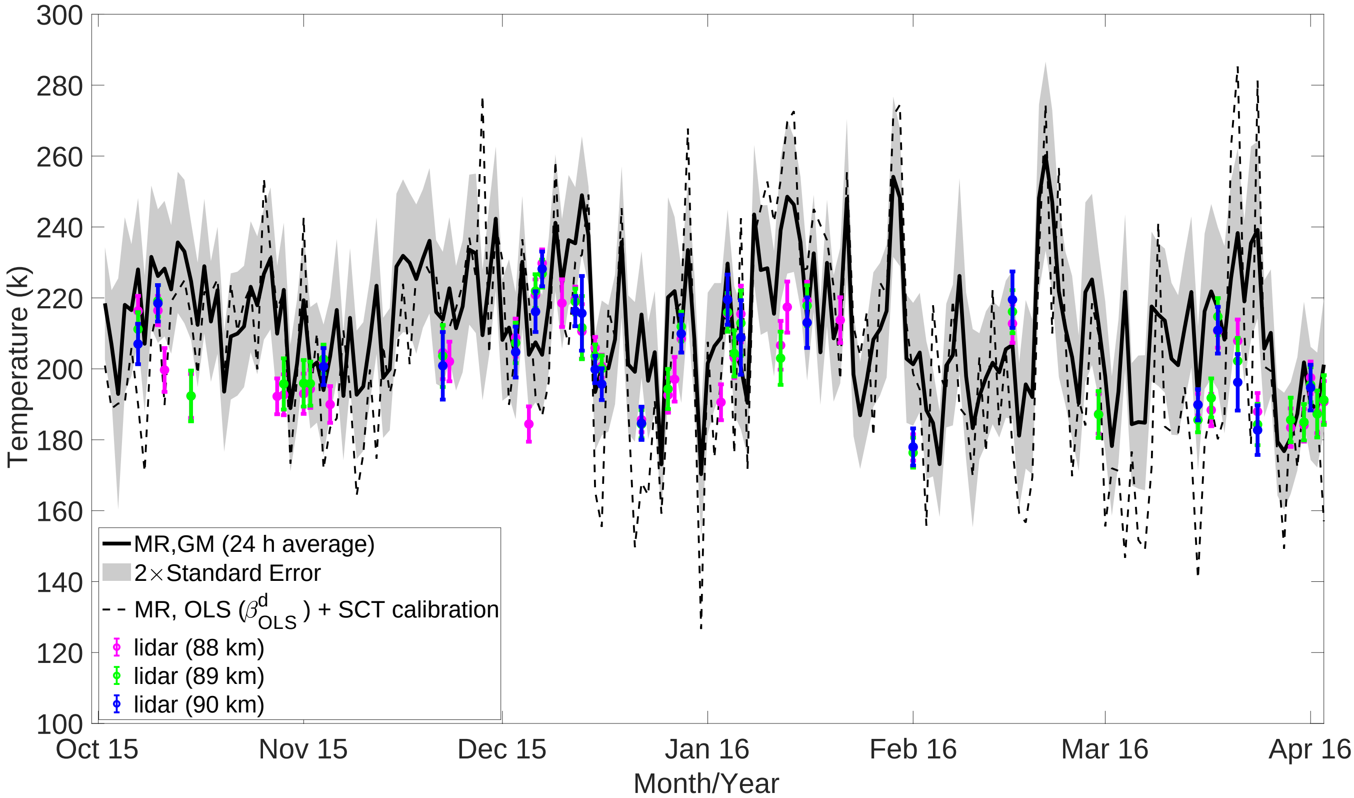

Figure A.6: Comparison of the bias-corrected MR temperatures with lidar data for the winter 2015–2016. The *solid* line corresponds to the temperature estimated using the GM solution described in Sect. 3.2.1. The *dashed* line corresponds to the SCT calibrated temperatures using the colocated lidar measurements. The OLS estimates are obtained with $log_{10}(1/\tau)$ as independent variable. The errors in lidar temperatures are 5–10 K and the standard error (grey shade) of the temperature from EIV analysis is on average 19 K. The differences between lidar and MR temperatures are presented in Fig. A.7.

Table A.1: Representative value of the (square root of) error variances and their normalised ratio for SGO's meteor radar obtained by SCT method (Fig. A.4) for winter 2015–2016. The estimated value of the error variances in normalised height and $log_{10}(1/\tau)$, $s_{\varepsilon'}$ and $s_{\delta'}$, are given along with the estimated value of λ from SCT calibration (λ^{OLS}_{eff}) and EIV method (λ^{EIV}_{eff}). The literature values are from [1](Hocking2004), [2](holdsworth2006), and [3](kim2012).

–	Mean	STD	Literature Values
$\triangle(log_{10}(1/\tau))/s^{-1}$	0.11	0.02	$0.14^{1,2}$
$\triangle(height)/km$	2.13	0.32	3.25^{1}, 1.1^{2}, 1.0^{3}
$s_{\varepsilon'}$	0.36	–	–
$s_{\delta'}$	0.29	–	–
λ^{OLS}_{eff}	1.24	–	–
λ^{EIV}_{eff}	1.39	0.02	–

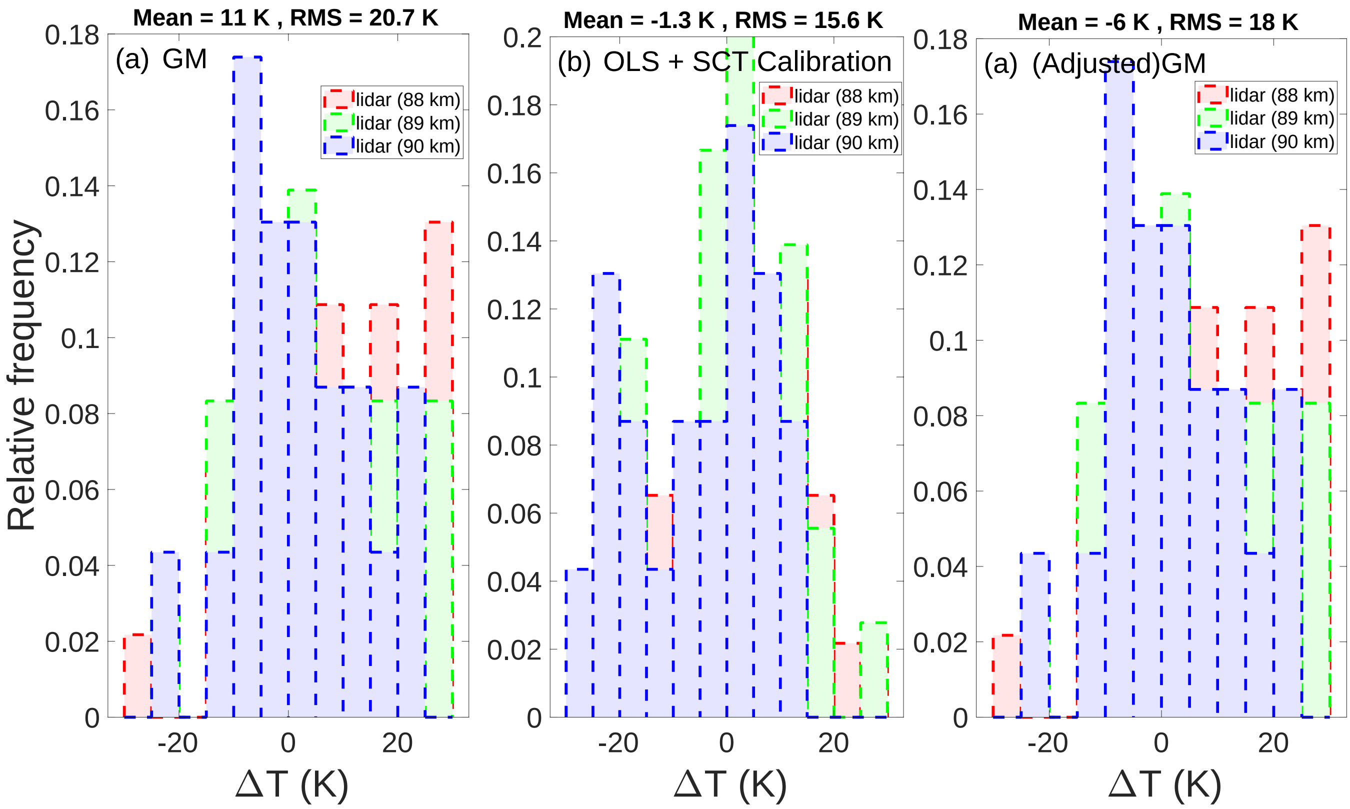

Figure A.7: Difference between MR temperatures and lidar data for (a) GM solution and (b) SCT calibration applied to OLS estimate of β^{d}_{OLS}, and (c) adjusted GM solution . MR temperatures are shown in Fig. A.6.

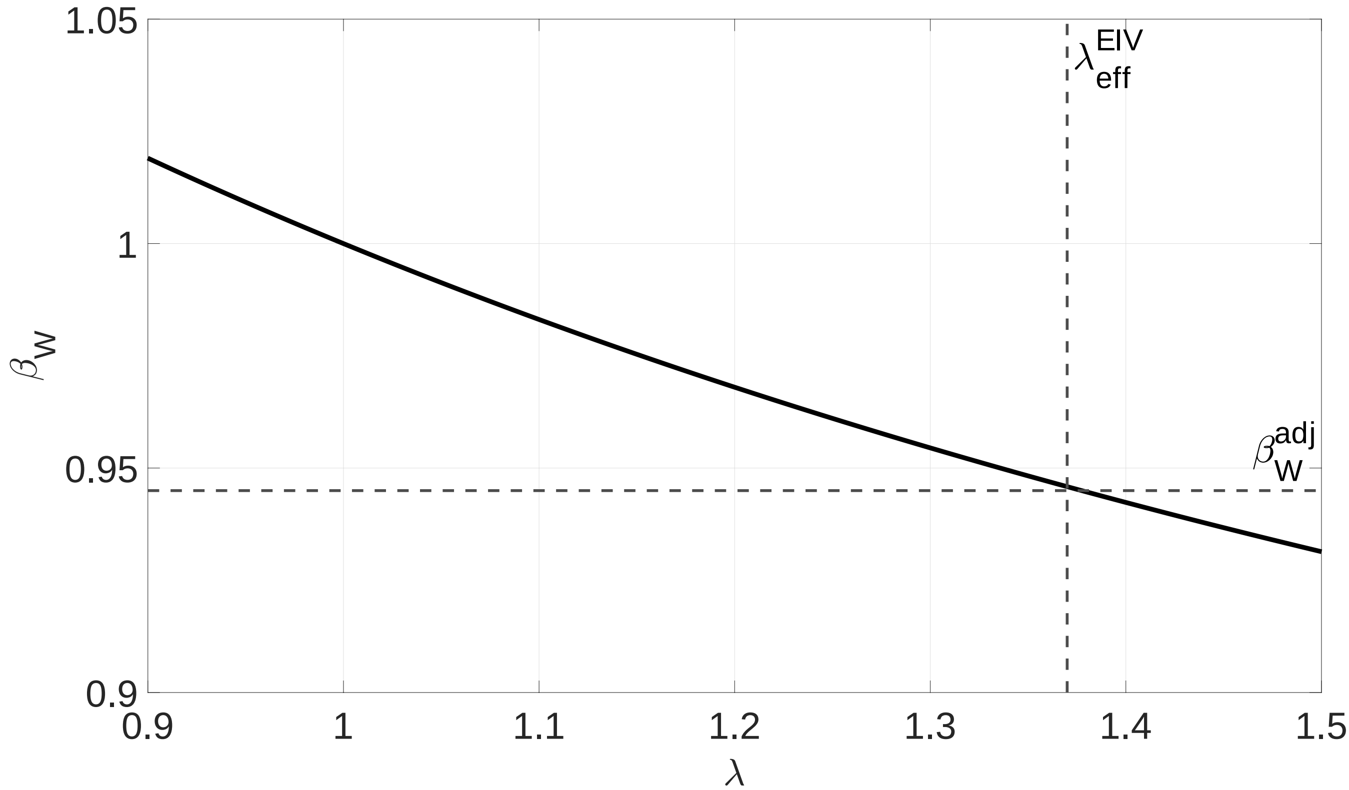

Figure A.8: Two independent estimates of slope using EIV method for the date 14 Nov 2015. The solid line corresponds to the slope estimated using Eq. (26) with λ as free parameter. β^{adj}_{W} is the slope estimated from Eq. (37). Comparison of these two methods shows that the value of $\beta^{adj}_{W} = 0.945$ is equivalent to using $\lambda = 1.38$ in Eq. (26).