Report of the inversion unified algorithm working group

I Workshop on Lidar Inversion Algorithms-LALINET

10 to 14 of March, 2014

The main goal of the I Workshop on Lidar Inversion Algorithms of LALINET, held in the University of Concepción, Chile, from 10 to 14 of March 2014, was to compare the inversion algorithms for elastic backscatter lidars from each Latin-American Lidar group in order to develop a uniform, unified and improved algorithm. The following scientists attended the event, which was supported by the Center for Optics and Photonics (CEFOP) of the Universidad de Concepción:

1. Henrique Barbosa, USP – Sao Paulo, Brazil
2. Fabio Lopes, IAG-USP/IPEN – Sao Paulo, Brazil
3. Elena Montilla, CEFOP – Concepción, Chile (organizer)
4. Daniel Nisperuza, UNAL – Medellin, Colombia
5. Pablo Ristori, CEILAP – Buenos Aires, Argentina
6. Antonieta Silva, CEFOP – Concepción, Chile

The algorithm evaluation and improvement was based on the analysis of three simulated lidar datasets and comparison with the expected results. The first dataset was provided by Dr. Holger Baars, from the Institute for Tropospheric Research (iFT), Leipzig – Germany, and corresponds to the data used in the EARLINET paper Boeckmann et al., Appl. Opt. (2004). The second dataset is a modification of the first one, by Henrique Barbosa, from USP, Sao Paulo – Brazil, to include different levels of Poisson noise. These two datasets provided the elastic channels: 355, 532 and 1064 nm. The third dataset was provided by Pablo Ristori, from CEILAP, Buenos Aires – Argentina. This is based on the same input T, P profiles but include boundary layer aerosols and a cloud. All these input and output files are stored and available from our website: www.lalinet.org/.

The first day (Monday, March 10th) was dedicated to the presentation and acquaintance with the first case. Units and file structure were defined for both input and output. The first exercise was to process the signal without knowing the answer. The molecular region was defined between 12 and 15km and the LR was prescribed for each wavelength, 28 sr, 39 sr and 77 sr for, 355nm, 532 nm and 1064 nm, respectively. Results for the particle backscatter are shown in the figure 1. Results from SP were displaced downward due to a bug in the lidar altitude. Results from CO failed for 532nm due to a wrong LR used.

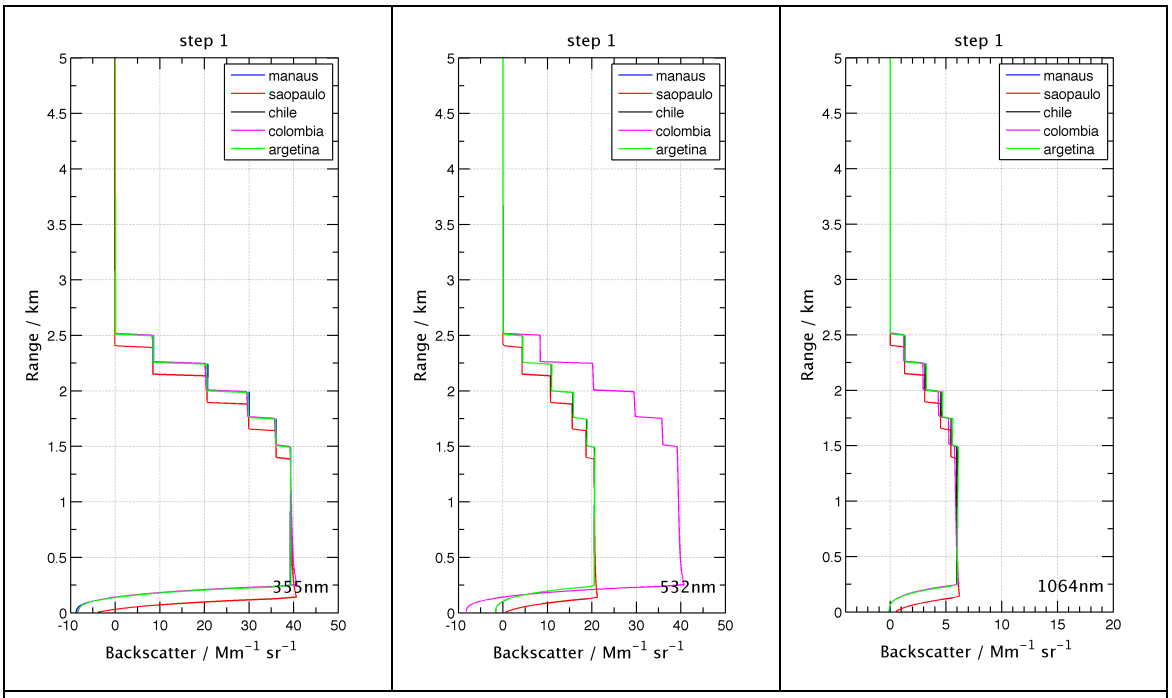

Figure 1 – Particle backscatter (Mm-1 sr-1) after step 1.

Figure 2 shows the second interaction, i.e. after the comparison with other groups, which was a good reference of the expected results. A close look at the results that seem coincident, on figure 2, revealed that groups were having systematic difference. Molecular quantities were then compared and the routines were revised in order to find the differences, summarized in table 1.

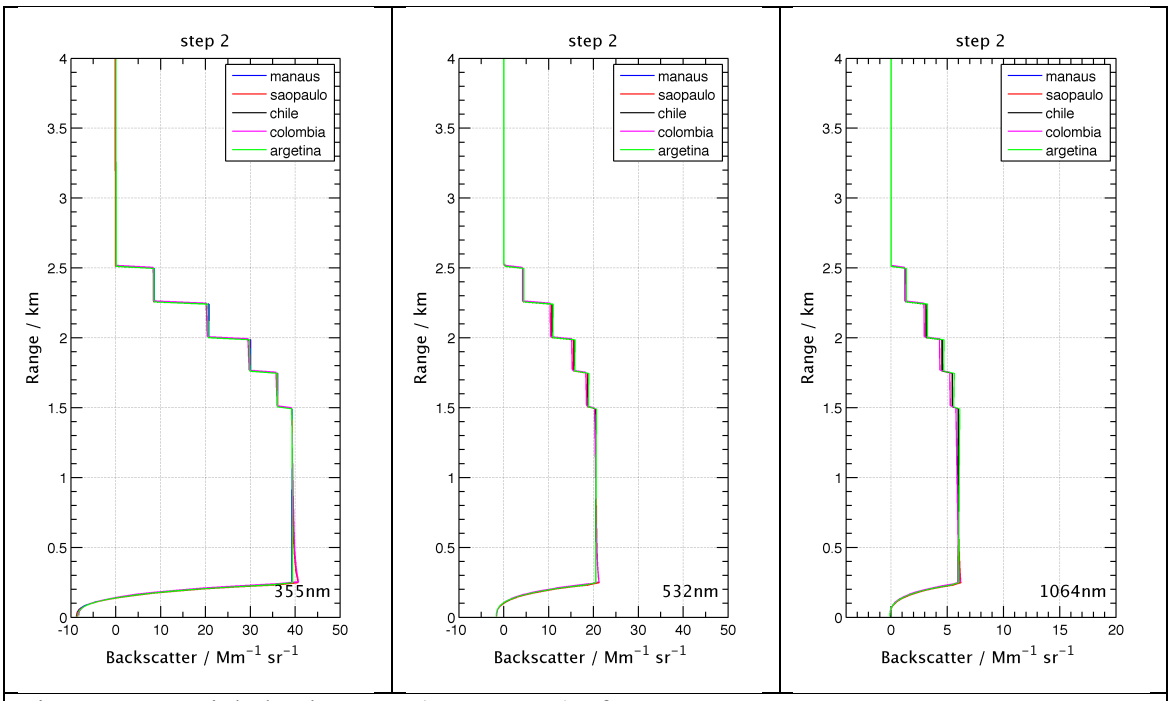

Figure 2 – Particle backscatter (Mm-1 sr-1) after step 2.

The differences in the molecular backscatter at the reference height lead to further investigation of each calculation inside the molecular routines. The cross section for the Rayleigh scattering varied about 7% among different groups (2.63 – 2.77 x 10^{-30} m^2) because two used Bucholtz (1995) and Bodhaine (1999) while three groups used Nicoleti's simplified equations. The final molecular lidar ratio varied between 8.3776 and 8.50411 sr. It was decided during the meeting that doing the calculation based on first principles was the best approach so AR and CH changed their algorithms according to Matlab code from AM. Moreover, CO_2 concentration was modified to match the value used by SP and CO, 392 ppmv, and wavelengths were set precisely to 355.0, 532.0 and 1064.0 nm. The last change was that SP and CO started doing the Rayleigh fit for determining the background. The values shown in red in table 1 are those after this first set of changes to the algorithms. A much better agreement was found for the molecular scattering cross section (<0.32%), and molecular lidar ratio (<0.01%) and molecular backscatter at the reference height (<0.3%).

Table 1 – Differences before the first step (black), after the comparison of molecular quantities and first changes to the inversion algorithms (red) and after the comparison with known answer (green) are shown.

	$\beta_{par}(z_0)$ Mm^{-1}	**$\beta_{mol}(z_0)$ Mm^{-1}**	**Scale x10^{14}**	**BG**	**Bin z_0**	**σstd 10^{-30} m^{-2}**	**LR sr^{-1}**
AM	0	1.6157 1.6054	4.8967 4.89763	1.2294 10^{-3} 7.6318 10^{-4}	899 900	2.7694	8.5058
SP	0	1.60404	1.748 1.75087 4.8952	- 8.16868 10^{-7} 6.045 10^{-4}	900	2.7606	8.50411
CH	0	1.5598 1.6065	2.08 2.034 4.8859	0 8.33 10^{-4}	900	2.6381 2.7694	8.3776 8.5058
CO	0	1.51 1.60404	1.8489 1.75087 4.8929	- 8.16868 10^{-7} 5.5 10^{-4}	900	 2.7606	8.50411
AR	10^{-3}	1.5597 1.609	4.83 4.89	4.81 10^{-12} 1.16 10^{-3} 7.34 10^{-4}	900	2.639 2.7694	8.3776 8.5058

The second day (Tuesday, March 11th) started with the comparison of the molecular profiles obtained in the previous day. This is shown in figure 3 for molecular backscatter coefficient, the synthetic molecular signal. The results from all groups are coincident at this resolution. After this step, time was dedicated to the so-called molecular or Rayleigh fitting. In this procedure, one performs a linear regression between the lidar signal, P(z), and the molecular signal, $P_{mol}(z)$, between a range of altitudes, z, for which it is believed that the atmosphere is pure molecular. The equation used is $P(z) = A * P_{mol}(z) + B$, where A represents the molecular scale (shown in column 4 of table 1) and B is the signal background. The evaluation of the fitting

methods was necessary because the scale and BG were still very different after the improved molecular calculation (red values in table 1).

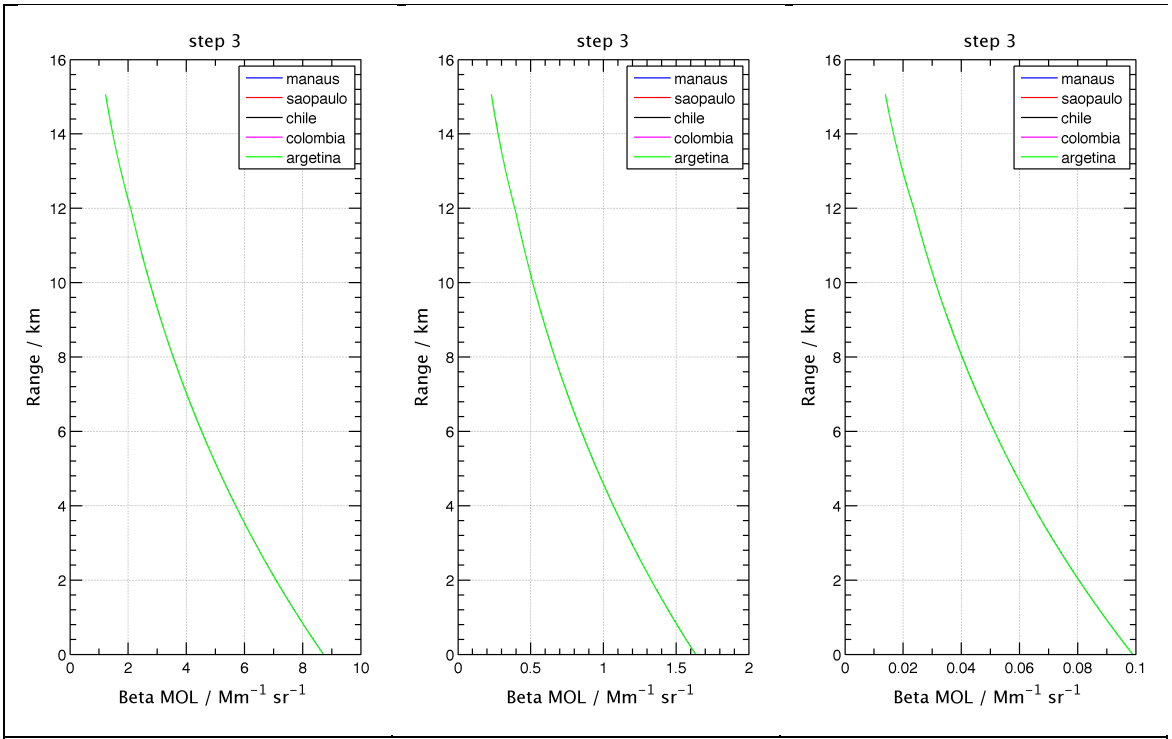

Figure 3 – Molecular backscatter (Mm-1 sr-1) at begin of step 3.

During this comparison, many bugs were found in all codes. First, the calculation of the molecular signal by AM was missing a 2 in the attenuation term exp[-2τ]. SP and CO were using $\beta_{mol}(z)$ instead of $P_{mol}(z)$ for the fitting. CH was removing the background from the signal but not from the range corrected signal, and finally AR was modifying the lidar signal instead of the synthetic molecular signal, which leads to fitting errors when the lidar signal is very noise. After these changes, the values for the molecular scale and background noise changed, shown in green in table 1. The scale's differences are smaller than 0.3% and the background noise, which in those synthetic data were zero, are all of the order of 10^{-4}. The resulting aerosol backscatter coefficients are shown in figure 4. The results from all groups are all superimposed and fit perfectly with the simulated profile, except in the region affected by the overlap. For better accessing the differences, the lower panel of the same figure shows the differences between the reconstructed and simulated aerosol backscatter. Differences for 355, 532 and 1064 nm are less than 0.05, 0.01 and 0.06 Mm^{-1} sr^{-1}, which is about 0.1%, 0.05% and 1% of the backscatter respectively.

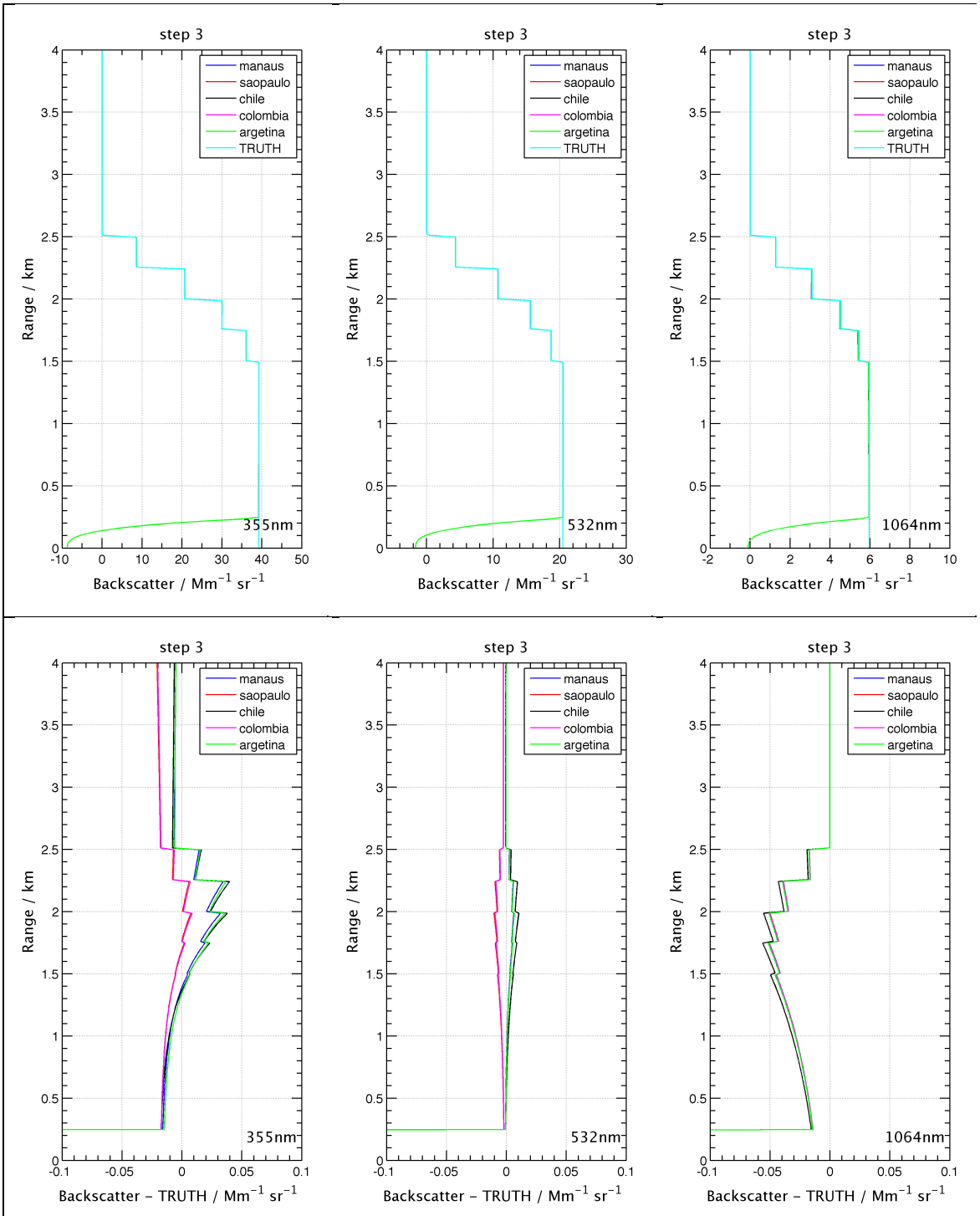

Figure 4 - Particle backscatter (Mm-1 sr-1) after step 3 and comparison with simulation input (truth, cyan) is shown on top. Differences from each group's solution and the truth are shown in the lower panel.

Part of the day was used to try some automatic procedure to identify the molecular region. Two ideas were pursued: wavelets and a running linear fit. They both seem to work for reasonable S/N ratios, but failed the next day, when we tried more noise-data and/or non poisson types of noise. Part of the day was also dedicated to paperwork, as we had to sign the receipts, get the checks and withdraw the perdiems from the bank at the Concepcion University.

The third day (Wednesday, March 12th) started with a visit to the lidar station from the Concepción group at TIGO observatory. After that, we gave interviews to CEFOP's journalist, and to the local newspaper (a piece on Lalinet workshop at Concepcion will appear on Sunday's edition, March 23rd).

From the science point of view, we worked on the analysis of the same input data but with different levels of noise. Firstly, the noise was added as a Poisson distribution around the input signal + noise, ie. $P(z) \rightarrow Poisson(P(z) + bg)$, where bg varied from 10^0 to 10^6 in step of decades. This, however, lead to problems as the input signal smallest values were about 1.0 with 4 decimal places (either mV or average #photons) and the Poisson function introduced a bunch of zeros. After we realized that, we changed it to $P(z) \rightarrow Poisson(10^3(P(z) + bg))$, which gave a proper simulation of a noise signal. This dataset is also available from Lalinet webpage.

Results are shown in figure 5. All algorithms gave results equal to the truth for a level of noise up to 10^3. For 10^4, 10^5 and 10^6, results are increasingly more noise although not showing any systematic deviation or convergence problem. For 10^7, CO and SP could not do the inversion of the 1064nm channel. Even the reasonable inversions started to show a lot of noise as none of the groups were applying any early smoothing of binning at lidar signal. For this reason, we have identified that implement a smoothing strategy to noise data on the algorithm is an important point to be discussed in the next meeting. For 10^8, only AR (355 and 532) and CH (532 and 1064) were able to invert some of the data. As shown in figure 6, differences for 355, 532 and 1064 nm are less than 0.05, 0.01 and 0.06 Mm^{-1} sr^{-1}, which is about 0.1%, 0.05% and 1% of the backscatter respectively, for level of noise up to 10^4. Our investigation of these failures indicates that it was not possible to correctly remove the BG by fitting the molecular reference signal. This may not be a problem with our own lidar data, however, as we will have much more than 1000 bins to work with, hence, the standard way of averaging over some far region could be used. These, however, need to be treated carefully as some PMT's have an overshooting at the end. Further discussion about these issues needs to take place in the next meeting.

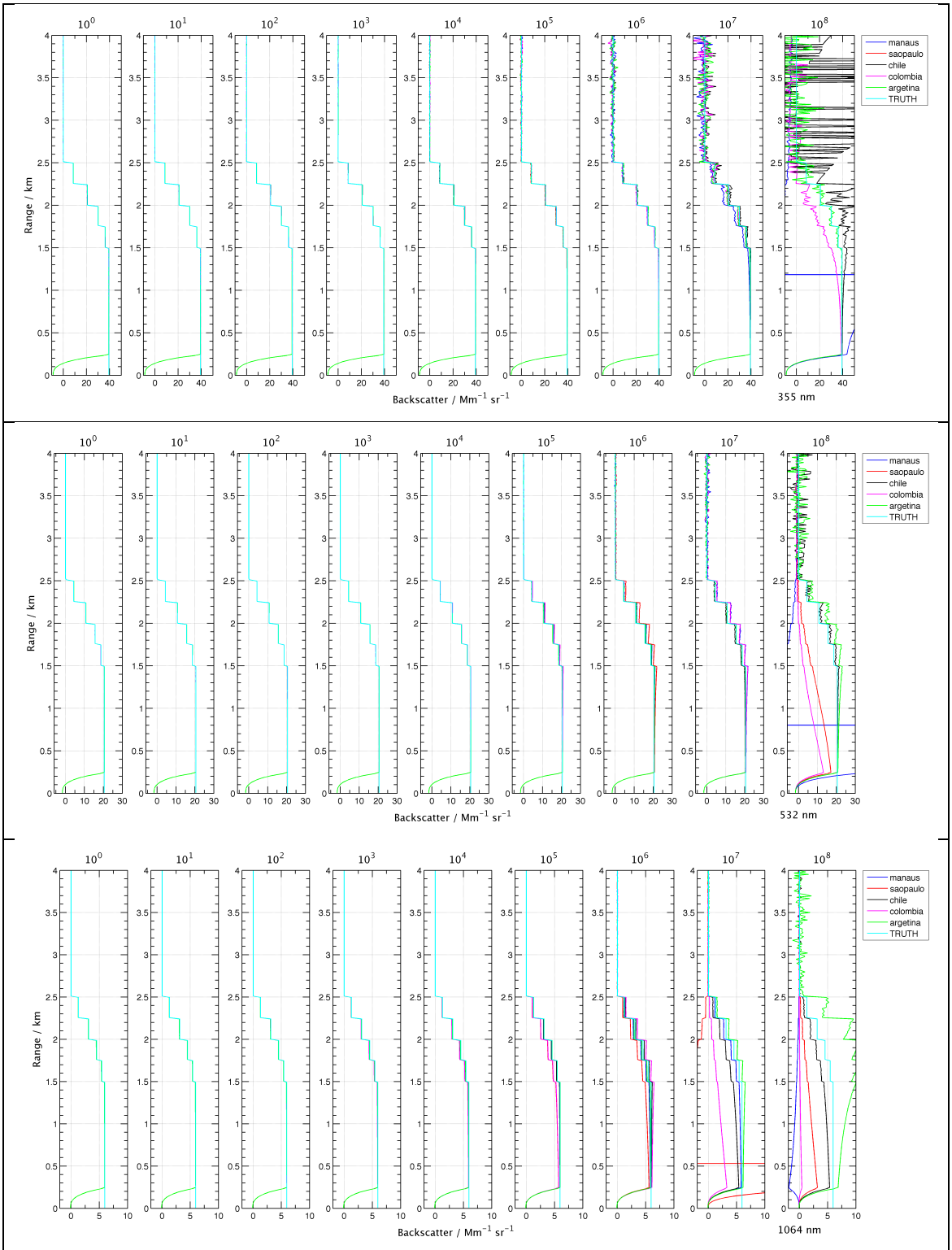

Figure 5 - Particle backscatter (Mm-1 sr-1) for the cases with different levels of noise for 355 (top), 532 (middle) and 1064 nm (bottom).

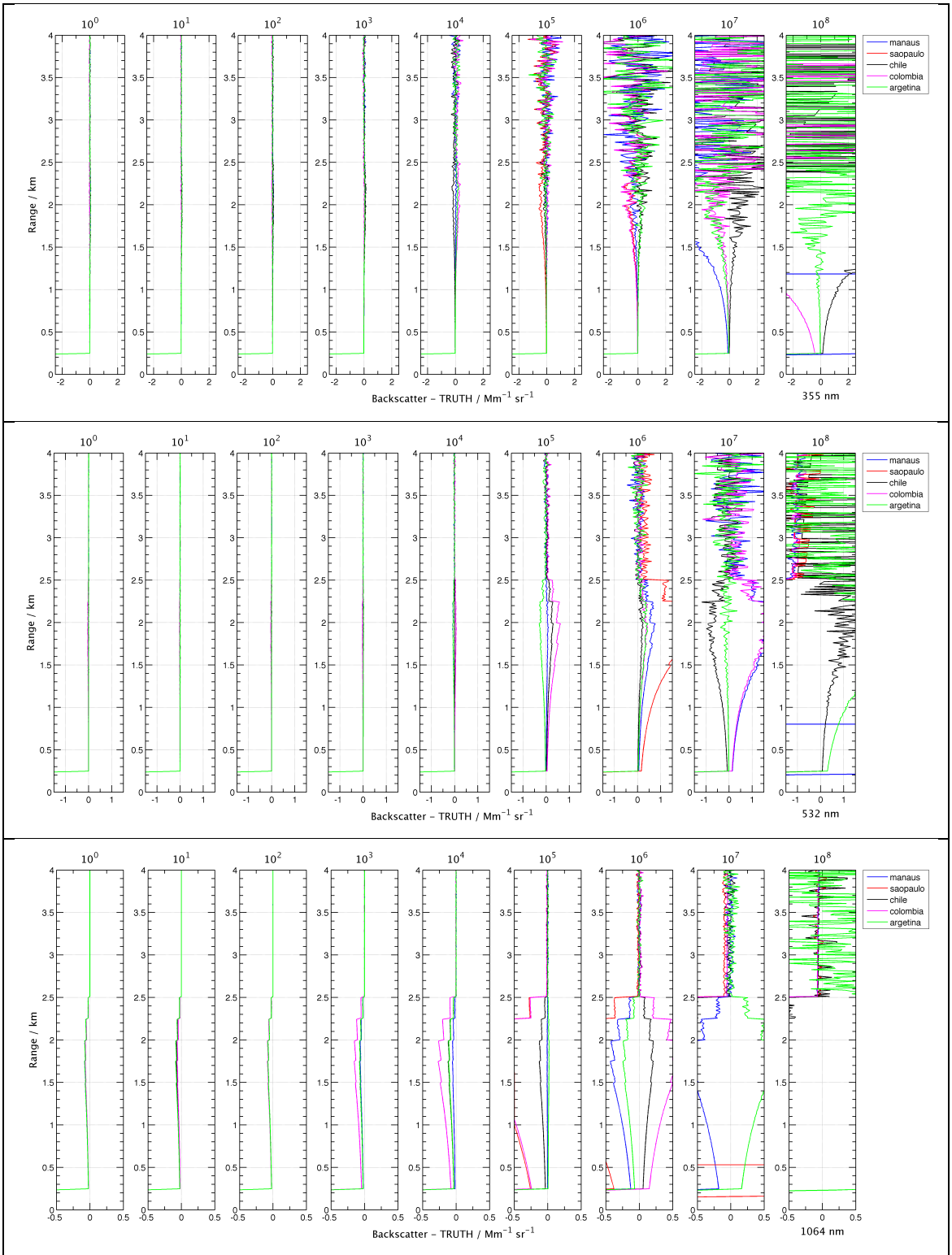

Figure 6 – Difference from inverted particle backscatter and truth (Mm-1 sr-1) for the cases with different levels of noise for 355 (top), 532 (middle) and 1064 nm (bottom).

Proposed recipe for Elastic lidar retrieval for Lalinet – what is ready and what needs further discussion

The team gathered at Concepción agreed on a standard for the retrieval of the particle backscatter from elastic lidar data. Matlab and Mathematica routines for doing these calculation have being prepared, with input from all the groups, and are available at Lalinet website. The official matlab routines are now being used by AM, AR and CH, while the Mathematica official routines are now being used by CO and SP. The steps defined so far are detailed below.

1. The molecular cross-section, molecular lidar ratio and molecular backscattering coefficient should be calculated based on Bucholtz (1995) and Bodhaine (1999).
2. With these and the pressure and temperature profiles interpolated from a nearby radiosonde, the expected lidar signal from a pure molecular atmosphere should be constructed.
3. A linear regression between the lidar signal, P(z), and the molecular signal, $P_{mol}(z)$, should be performed over a range of altitudes for which it is believed that the atmosphere is pure molecular. The angular coefficient should be used to scale the molecular model towards the lidar data, while the linear coefficient should be used as the background noise.
4. In the integration of the Fernald-Klett equation, the reference height z_0 should be chosen within the molecular range and any point is equally good. Within this range, it should be assumed that the aerosol loading in negligible. It should be taken into account the difference between the measured signal (with noise) and the scaled molecular signal (without noise) at the reference height. This should be done by normalizing the backscatter, as $\beta_{mol}(z_0)\ P(z_0) \ / \ P_{mol}(z_0)$.

Although the four days of hard work at Concepcion were very productive it was not possible to finish all we had to discuss regarding the elastic lidar inversion algorithm for Lalinet. The important points that were identified as major topics for the next meeting were:

1. Does the linear fit identify the correct BG in all cases? Or do we need an alternative approach when there is too much noise?
2. How to automatically identify the molecular region? Two ideas came up: wavelets or a running linear fit (maybe use A, B or RMS).
3. How to automatically identify clouds? Boris/Diego have an algorithm working for high clouds in the AM lidar data. Maybe adaptable for other lidars? And lower clouds?
4. How to smooth and/or change binning of the data? Do it before or after analysis? Is it possible to make an automatic decision based on the level of noise of each dataset?
5. How to interpolate from radiosonde T,P to lidar levels? How to extrapolate? Possible to use US standard atmosphere + ground data?

Lalinet papers x Earlinet papers

We looked at the three papers from the EARLINET community and we understand that a similar set of publications for Lalinet would be desirable. From our discussions so far, these papers would be:

1. Instrumental, probably lead by Juan Luis
2. Comparisons and presentation of official lalinet unified algorithm, probably lead by Henrique Barbosa
3. Raman algorithm, a leadership on that has still to come up

Particularly for the paper #2, we have discussed a lot about it. A tentative structure that we propose is shown below. What stands out when comparing with Earlinet is the presentation of an official unified algorithm for the network, evaluation with different noise levels and automatic cloud screening.

- Introduction
- Presentation of cases and tests
- Comparison of different algorithms for 355, 532 and 1064, based on dataset we've worked on during this week.
    - First with know LR, T and P but not knowing the answer
    - Second, fixing calibration height and after looking other group's results
    - Third, after knowing the answer but with increasing levels of extra noise
- Presentation of the unified algorithm, equations, constrains, etc.
- Application of the unified algorithm for a profile simulated with clouds, aerosols and noise.
- Application of the unified algorithm for sample data from each group
- Conclusions

Changes made to each algorithm

Changes made do Henrique's algorithm, São Paulo – Brazil, based on Matlab

1. Normalization of the reference molecular region before the Rayleigh-fit calculation to allow large differences between the synthetic and measured signals.
2. Exclude cloud and aerosols layers automatically by computing the RMS of a running linear fit between the molecular and lidar signals and comparing it to the expected sqrt(signal).
3. Removed bug in the calculation of the molecular signal. The optical depth was not multiplied by two, ie exp(-tau), hence it was wrong.
4. Removed bug in integration of molecular transmission in Fernald's equation to start from ground level instead of first bin.
5. Conversion of all units to standard S.I.

Changes made to Pablos's algorithm, Buenos Aires – Argentina, based on Matlab

1. Signal fitting to synthetic pure molecular profile to get background and backscatter value at reference height. In presence of intense noise this option is the most suitable.
2. Normalize reference region of both profiles by their mean value instead of scaling by a constant to reduce numerical instability.
3. Use Bodhaine et al to retrieve the precise atmospheric extinction and backscatter profiles.

Changes made to Antonieta's algorithm: ALICE (Algorithm of Lidar Cefop), Concepcion – Chile, based on Matlab:

1. Molecular backscatter now calculated from Bodhaine et al 1999 instead of Nicolet's equations. With this new function, the molecular lidar ratio is calculated instead of fixed.
2. Noise calculation was calculated using the mean value of the tail of the data. Now, the noise is calculated using the synthetic molecular signal (obtained from meteorological data and the new molecular function), and fitting this signal with the Lidar signal using a linear fit in the so-called molecular range.
3. There was a bug in the equation to obtain the total backscatter coefficient in Fernald method. The range corrected signal was not background subtracted. The reference backscatter coefficient is now fixed from the molecular backscatter coefficient at the reference height (assuming zero aerosols at the reference height).

Changes to Fabio's algorithm: LEBEAR (Lidar Elastic Backscatter and Extinction Analysis Routine), Sao Paulo - Brazil, based on Mathematica

1. The calculation of molecular simulated signal was included in order to compare with the measured Lidar signal.
2. It was included also a linear fitting calculation ($f(x)=ax+b$) of molecular simulated signal and measured Lidar signal in a reference range free of aerosols. If the interception point b is at least 2 times bigger than its standard error, the b value is used as the background noise value.
3. Changes on the integral calculation of the aerosol backscatter retrieval equation using Trapezoidal rule. It has been re-written as a Riemann sum equation instead to use the Trapezoidal rules package from Mathematica. The calculation time was reduced 100 times.

Changes to Daniel's algorithm: ROCEBSELI (Rutina para Obtener los Coeficientes de Extinción y Backscatter a partir de Señal Elástica LIdar), Medellin – Colombia, based on Mathematica.

1. Correction on routine reading pressure and temperature altitudes from radio-sounding data.
2. Changes in the molecular atmospheric model. Rayleigh extinction cross section is now computed according to Bodhaine et al 1999.

3. A synthetic molecular lidar signal from radio-sounding was included in routine that allows for the so called Rayleigh-fit.
4. Background signal is now considered as intercept value in the lineal fit on synthetic vs. real signal plot in free aerosols altitude range.
5. Klett integrals method was changed to trapezoidal numerical sum, improving the computational cost by a factor of 100.