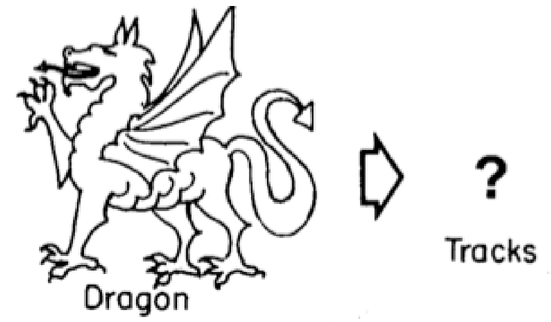
Retrieval problems in Remote Sensing



Maarten Krol
August 2006
m.c.krol@phys.uu.nl

1. Goal

To show that remote sensing is an indirect measurement, i.e. that we do not measure directly, but by means of (for instance) electromagnetic radiation as an information carrier. This means that we have to reconstruct from the measurements the unknown quantity that we are interested in. This reconstruction is called 'retrieval or inversion'. The goal of this exercise is to highlight problems that are often encountered in the retrieval process. We will use the relative simple case of temperature retrieval to highlight non-unique solutions and ill-posed inversions.

2. Relevance

Satellite remote sensing is playing an increasingly important role in Earth oriented research. Satellites have enabled a global view of the Earth. Clouds, forests fires, ocean currents, and ice sheets can all be observed from space. Remote sensing can even observe objects that are invisible to the human eye. Examples are stratospheric ozone, temperature profiles, the Earth gravity field, and ocean temperature. All these measurements are based on sensors that are sensitive to some aspect of the unknown quantity that we are interested in. Sometimes, however, one tends to forget that we only sense a fingerprint of these objects. The ocean surface is emitting radiation that is sensed to reconstruct its temperature; solar radiation is sensed that is influenced by absorption of stratospheric ozone, etc. This indirect observation has important consequences for the reconstruction of the unknown. Sometimes we have to conclude that we cannot retrieve the whole object, but only parts of it. Knowledge of the basic limitations of remote sensing techniques is vital for a sound application of remote sensing data.

3. Theory

3.1 The Nature of Inverse problems

The reconstruction of an object that is remotely sensed, is called an inverse problem. Figure 1 shows the problem that we face in reconstructing an object from which we have only traces of information. Can we reconstruct the dragon from its tracks? (how does its head look like? Does it have wings?)

Figure 1: The nature of inverse problems (from Stephens).

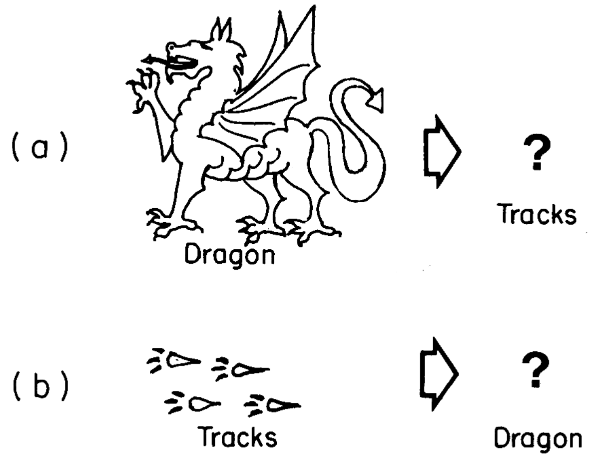


Figure 1.2 (a) The direct problem: Describe the tracks of a dragon. (b) The inverse problem: Describe a dragon from its tracks (from Bohren and Huffman, 1983.)

Clearly, by knowing only the tracks of the dragon, it is hard to tell the exact shape of the animal that created the tracks. The direct or forward problem is much easier! Knowing the animal (weight, size of feet, number of toes) it can be assessed how the tracks will look like.

Now we consider a physical analog. In this exercise we will study the sensing of the temperature profile of the atmosphere. In order to sense this temperature profile, satellite instruments measure the infrared radiation that is emitted by absorbing gases in the atmosphere (e.g. CO_2). The radiation that is emitted by these gases depends on the temperature. Suppose now for simplicity that the wavelength of the emitted radiation is related to a single level in the atmosphere (which is not true in reality). Then the relation between wavelength and temperature can be sketched as in figure 2.

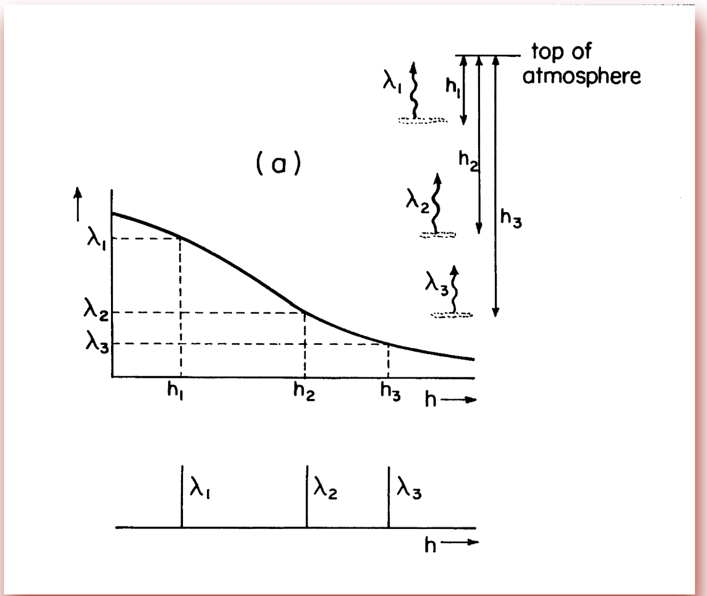


Figure 2: Schematic diagram that shows the hypothetical case in which each level (with a certain temperature) emits radiation at a discrete wavelength (from Stephens, after Twomey, 1977).

If the emitting gas is well-mixed (e.g. CO_2), then the measured intensity at a specific wavelength is directly related to the temperature in a specific layer. This would provide a way to *retrieve* the temperature as a function of height from the measured wavelength dependent intensity at the top of the atmosphere.

In reality, the situation is unfortunately more complex than the ideal case just described. The emitted radiation at a specific wavelength does not come from a single height but is originating from a broad layer in the atmosphere. This results in a blurring from the one-to-one correspondence between wavelength and height that was illustrated in figure 2. Figure 3 shows how this blurring looks like.

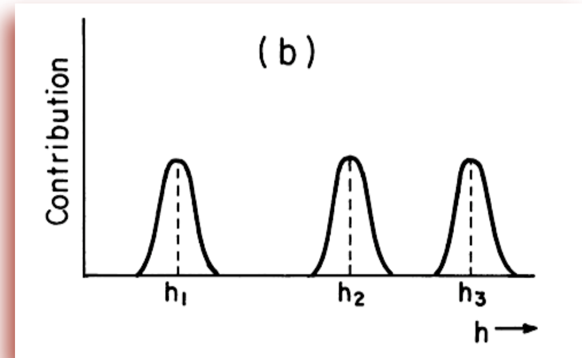


Figure 3: Blurring of the height profile. The information is not coming from a single layer, but from a vertical range of heights (from Stephens).

3.2 Weighting Functions

The example of the previous section is typical of many of the inversion problems in remote sensing where a set of measurements (in this case emitted radiation as a function of wavelength) is influenced by all values of the unknown distribution (in this case the vertical distribution of temperature throughout the layer that contributes to the emission). We represent the blurring effect in the following way. The thought-after temperature profile is represented as $T(z)$. $K_i(z)$ is the relative contribution curve for a wavelength λ_i. The height-interval between z and $z + \Delta z$ contributes to the measurement of the radiation with λ_i the amount $T(z)K_i(z)\Delta z$. The total radiation with wavelength λ_i is then

$$g_i = \int_{z_0}^{z_t} K_i(z)T(z)dz \qquad (1)$$

where the limits of the integral run from the surface to the top of the atmosphere. The function $K_i(z)$ is known as the "kernel function" of "averaging kernel" and determines is this case what part of the vertical temperature profile is sampled. The inversion problem now is to find $T(z)$ given measurements at different wavelengths g_i. Since there is normally not a one-to-one correspondence (figure 3), this inversion is often ill-posed. One of the consequences is that their may be more profiles $T(z)$ that produce the same measurements g_i. This problem is normally overcome by restricting the class of admissible solutions to physically realizable ones. In this way, *prior information* is introduced in the retrieval scheme.

Another difficulty in solving equation 1 concerns the problem of instability, which arises for example from errors in the observations g_i. With a small error ε_i in the observations g_i the equation becomes

$$g_i + \epsilon_i = \int_{z_0}^{z_t} K_i(z)T(z)dz \tag{2}$$

where ε_i produces a large change in $T(z)$. The ultimate success of any retrieval largely depends on the accuracy of the measurement g_i and in the shape of $K_i(z)$. A simple example underlines this important point.

3.3 Example of instabilities

Suppose we want to measure the temperature in two vertical layers $T(z_1)$ and $T(z_2)$. We make two measurements I_1 and I_2 that depend on the temperatures $T(z_1)$ and $T(z_2)$ in the following way:

$$I_1 = K_{1,1}T(z_1) + K_{1,2}T(z_2)$$
$$I_2 = K_{2,1}T(z_1) + K_{2,2}T(z_2)$$

which is just a 2-layer discretization of equation 1 (taking $\Delta z = 1$). Let us suppose for the sake of illustration that the weighting functions K have the numerical values $K_{1,1} = 1$, $K_{1,2} = 1$, $K_{2,1} = 2$, and $K_{2,2} = 2.000001$, and that the measured intensities are $I_1 = 2$ and $I_2 = 4.000001$. As a solution for the unknown temperatures we obtain by solving

$$T(z_1) = 1$$
$$T(z_2) = 1$$

Suppose that there is a small uncertainty in the measured intensity I_2, such that the value $I_2 = 4$ is recorded instead of the value 4.000001. Then the solution changes to:

$$T(z_1) = 2$$
$$T(z_2) = 0$$

This is a dramatic change to the solution and nicely illustrates the problem of instability. In reality, matters may become even worse. Figure 4 provides an example of how random errors propagate in an actual temperature retrieval. The retrieved temperature profile obtained using infrared measurements from a radiometer when a spurious error of 10% is added to one of the measurement points is shown in this diagram. During the exercise, we will explore these effects further.

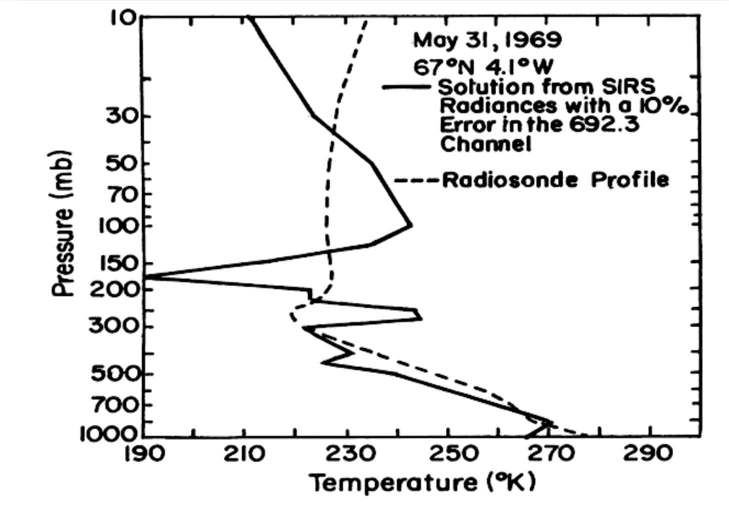


Figure 4: Comparison of radiosonde and a retrieved temperature profile when a spurious error of 10% is added to one measurement (from Stephens (1994), after Smith (1972)).

4. In Practice

4.1 Weighting functions and prior information

We will start the analysis by selecting a known temperature profile. This profile will be multiplied with a set of pre-defined weighting functions to produce radiance values at the top of the atmosphere. These pre-defined weighting functions are sketched in figure 5.

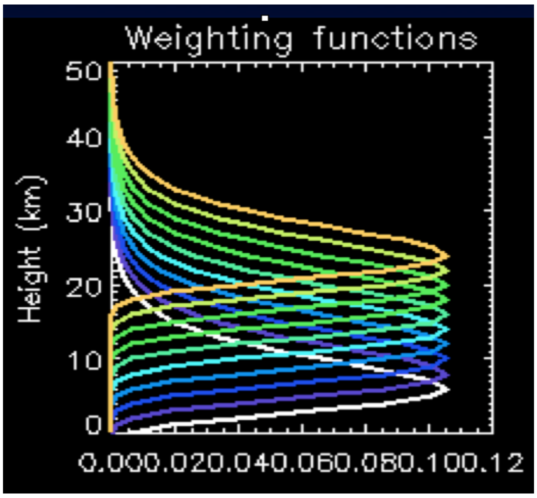


Figure 5: pre-defined weighting functions used to generate measurements from a selected temperature profile.

From equation 1 it is clear that the value of a weighting function at a specific height indicates how sensitive the measurement at the top of the atmosphere is for the temperature at that height level. For instance, the white weighting function has its largest sensitivity at a height around 5 km. In contrast, the orange weighting function is sensitive around a heigh of 25 km.

The task of the inversion is now to reconstruct the pre-defined temperature profile from the generated measurements. Non-uniqueness of this retrieval is clear at the top of the atmosphere. None of the weighting functions is sensitive to these upper layers, which means that the temperature of these layers is not reflected in the generated measurements.

To solve this problem, retrievals often use *prior information*. For instance, we request the temperature at the top layers to remain within a pre-defined temperature range 260±50K. Now the question arises how to combine the prior information and the generated measurements to retrieve the temperature profile. This is done by minimizing a cost function, as will be explained in the next section.

4.2 Cost function

A cost function is a function that sums

- The deviations between the measurements g_i and the radiances calculated from the retrieved temperature profile $T(z)$ using the weighting functions $K_i(z)$
- The deviations between the prior temperature profile $T_p(z)$ and the retrieved profile $T(z)$.

The first part is weighted with the measurement uncertainty. The second part is weighted with the pre-defined uncertainty in the prior temperature profile. First assume that the errors in the prior temperature profile are uncorrelated and that also the errors in the observations are uncorrelated. Then the cost-function J looks like:

$$J = \frac{1}{2}\sum_i \frac{(K_i\mathbf{T} - g_i)^2}{\epsilon_i^2} + \frac{1}{2}\sum_j \frac{(T(z_j) - T_p(z_j))^2}{\sigma_{T_p}(z_j)^2}$$

where the first part represents the differences between the measurements g_i and the modeled measurements ($\mathbf{T}$ represents the temperature profile, K_i the i^{th} weighting function), weighted with the measurement errors ε_i. The second part represents the deviations between the temperature profile $T(z)$ and the prior temperature profile $T_p(z)$, weighted by the estimated error in the prior profile σ_{Tp}. Note that all the terms are squared to sum up positive and negative deviations. The retrieval program that you will use minimizes this cost function by varying the temperature profile $T(z)$.

In practice, the errors in the prior profile as well as in the measurements are correlated. This complicates the cost function, but the principle stays the same. In this case, the cost function is more conveniently written in matrix notation:

$$J = \frac{1}{2}(\mathbf{KT} - \mathbf{g})^{\mathrm{T}}\mathbf{S}^{-1}(\mathbf{KT} - \mathbf{g}) + \frac{1}{2}(\mathbf{T} - \mathbf{T_p})^{\mathrm{T}}\mathbf{S_a}^{-1}(\mathbf{T} - \mathbf{T_p})$$

where $\mathbf{S}$ and $\mathbf{S_a}$ represent the error covariance matrices for the measurements and the prior temperature profile, respectively. $()^{\mathrm{T}}$ means that the transpose is taken. J represents the cost function that is minimized by the computer program that will be used.

Only the result of the minimization is given here. The minimum is found by setting the derivatives of J with respect to $\mathbf{T}$ to zero. Without going in to more detail, the results looks like:

$$\mathbf{T} = \mathbf{T_p} + \mathbf{A}(\mathbf{g} - \mathbf{K}\mathbf{T_p}) \qquad (3)$$

with $\mathbf{A} = \mathbf{S_a}\mathbf{K}^\mathbf{T}(\mathbf{K}\mathbf{S_a}\mathbf{K}^\mathbf{T} + \mathbf{S})^{-1}$

Thus, the temperature profile **T** that minimizes *J* can be written as a correction to the prior temperature profile $\mathbf{T}_p$. Note that this correction is zero for those measurements g_i that are equal to $K_i\mathbf{T}_p$. For the other measurements, the prior profile $\mathbf{T}_p$ is adapted with an amount that depends on **A** and the difference (g_i - K_i**T**). **A** depends on the error covariance matrices S and S_a.

For completeness, the program also returns the posterior error covariance matrix S_T.

$$\mathbf{S_T} = (\mathbf{K}^\mathbf{T}\mathbf{S}^{-1}\mathbf{K} + \mathbf{S_a}^{-1})^{-1} \qquad (4)$$

Posterior errors are always smaller than the errors in the prior and indicate where the measurements reduced the estimated error S_a in the prior temperature profile $\mathbf{T}_p$.

4.3 Research Questions

During the exercise, you will be able to investigate how the retrieved temperature profile depens on

- the prior temperature profile $\mathbf{T}_p$
- the errors in the prior S_a
- the assumed error S in the measurements g
- the amount of noise **ε** added to the measurements

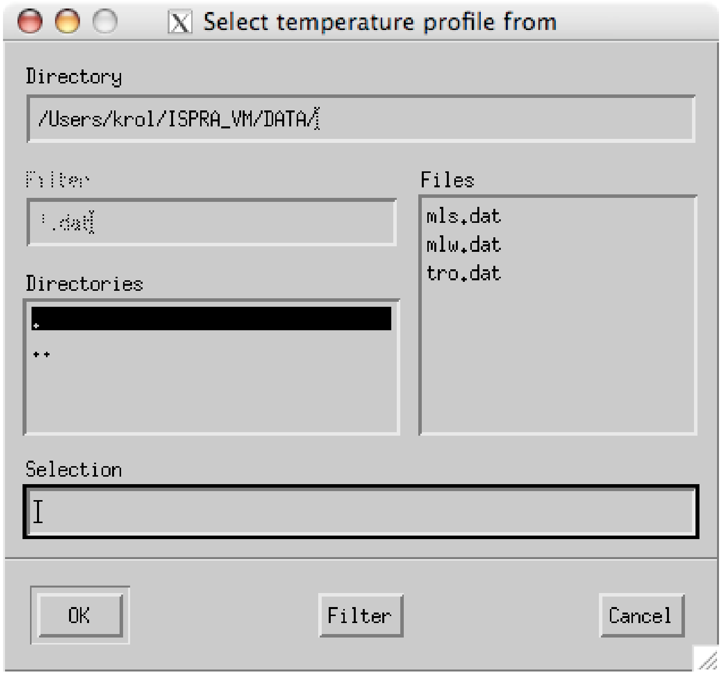
Specific questions that are addressed are:

1. *How does the retrieved profile depend on the prior information?*
2. *When does a retrieval becomes unstable (figure 4)?*
3. *How and where does the uncertainty reduce as a result of the measurements?*

5 Analysis

5.1 Input and Output

First, select "Temperature sounding" from the main menu. This will allow you to select a prior temperature profile. As explained in section 4.1, this profile will be used to generate measurements by multiplying the profile with the averaging kernels. You can choose from three different profiles



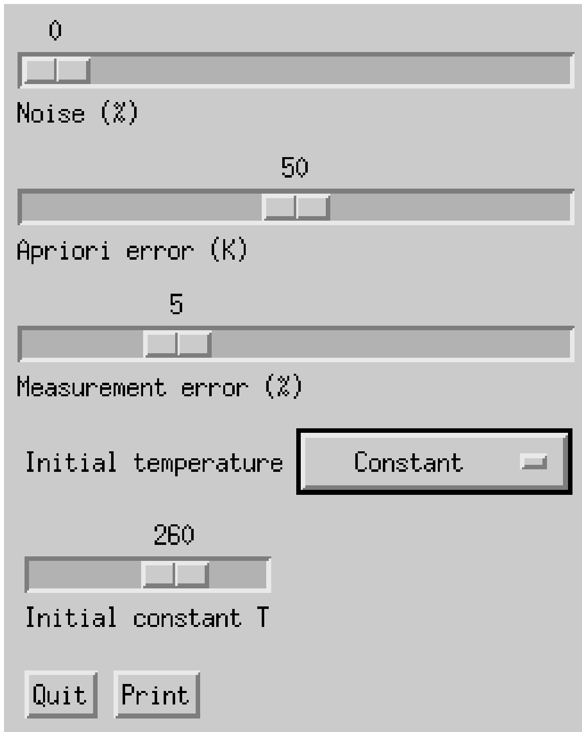
•mls: mid-latitude summer

•mlw: mid-latitude winter

•tro: tropical

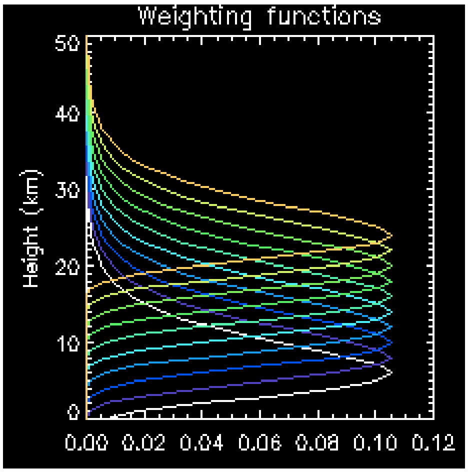
Once the profile is selected, the measurements are directly performed and the results will be displayed. The menu lets you change a number of things. Specifically:

- noise (%): drag the drawer to select an amount of noise **ε** to be added to to the measurements g.
- Apriori error (K): This is the error that is assumed on the prior temperature profile. It is the diagonal of matrix S_a.
- Measurement error (%): This is the error that is assumed to be present in the measurements g. It is the diagonal of the matrix S.
- Initial temperature profile. Here you can select from a drop-down menu:

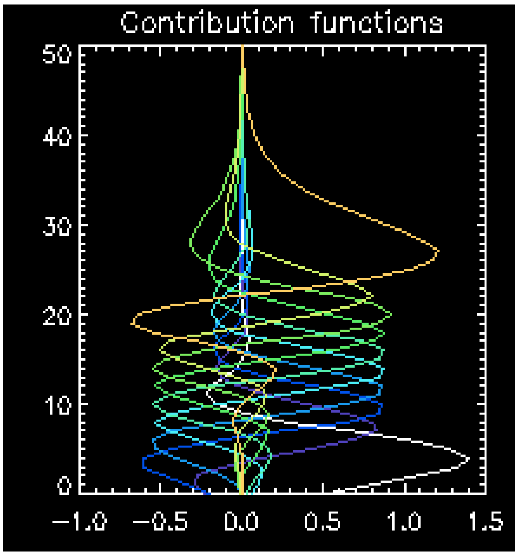


When "constant" is selected, you can modify the "constant" temperature with the "Initial constant T" slider.

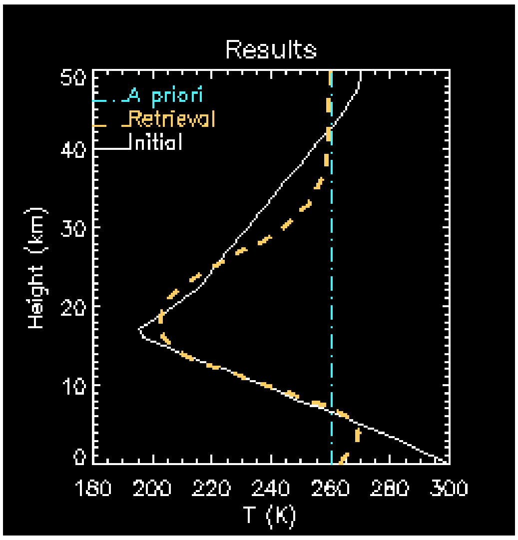
Initially, a prior temperature profile is selected that is constant with height. This constant temperature is initially 260K, and the assumed error is 50K. Given the values that you provide, the program will immediately calculate the retrieved temperature profile. The "output" window shows the following four output windows.



- On the left top, there are the predefined weighting functions **K** that are used to calculate the measurements **g** from the selected prior temperature profile **T**. This part of the output remains fixed.

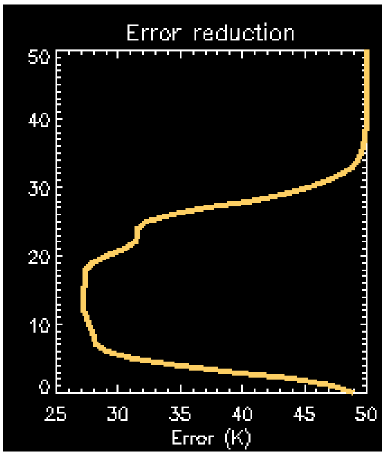


- On the upper right, the "contribution functions" are shown. In fact, they represent the matrix **A** in equation 3. This matrix depends on the prior errors and will be adapted when the prior errors are modified.



- On the lower left, the results are plotted. The window shows in white the initial profile that you selected. This is the profile you want to reproduce, and is often referred to as the "truth". In blue, the window shows the selected prior profile. In this example a constant 260K. In orange, the results **T** from the retrieval are shown.

- Finally, the lower right plot shows the "error reduction". It represents the diagonal of the matrix S_a from equation 4. The upper limit is governed by the value that is set at the "A priori error" slider.



Exercise 1: Influence of the prior profile

First, Modify the value of the "initial constant T" slider. Observe what is the effect on the retrieved temperature.

1. *Can you explain the height dependence of the response?*
2. *When happens to the "Error reduction"?*
3. *Where do the measurements provide the most information?*

Next, select a different prior profiles from the "Initial temperature" drop-down menu.

4. *What happens if the prior profile is the same as the initial profile?*
5. *Does the prior profile influence the "error reduction"? Can you explain why?*

Finally, play around with the prior error slider for different prior profiles.

6. *What happens if you gradually change the slider value from 100K to 0K?*
7. *Can you explain from the formulas in the previous section?*

Exercise 2: Measurements errors vs. Prior errors

The relative weighting of the various terms in the cost function J is important. The weighting is governed by the error matrices S and S_a. In this exercise we investigate how the retrieval results depend on the relative errors in the prior and the measurements.

First, set the "A priori error" slider to 50K. Then gradually increase the measurement error from 5% to 20%. Repeat this for different prior profiles and errors.

8. *What happens to the retrieved profile?*
9. *Is it possible to retrieve the 'true' profile, if you start with a wrong prior?*
10. *Which error settings give the 'best' retrievals?*

Exercise 3: noise on the measurements

Up to now, we assumed perfect measurements. In the retrieval we only provide the 'estimated' error. Now, we will perturb the measurements **g** that are calculated from the prior temperature profile. This is done by adding random noise **ε** to the simulated **g**. The noise is added with the 'noise' slider on top of the window.

First, set the 'measurement error' to 20% and the prior error to 50K. Then, gradually increase the noise to the measurements from 0% to 20%.

11. *What happens to the retrieved profile?*
12. *Can you explain?*

Next, the measurement error is lowered gradually to 0%, while the noise remains high (e.g. 20%)

11. *Compare the retrieved profile to the retrieval error in the 'error reduction output'. Can you explain?*
12. *Does it help to start with the 'true' profile as prior temperature profile?*
13. *What do you conclude about the relative magnitude of the noise compared to the measurement error?*

Finally, set the error on the prior value to a small value (e.g. 1K) and repeat.

14. *What can you conclude about a 'stabilizing' effect of the prior on the retrieval?*
15. *With a small prior error and a large noise, what is the information you gain from the measurements?*

References

Rodgers, C. D. (1990), Characterization and error analysis of profiles retrieved from remote measurements, J. Geophys. Res., 95, 5587-5595.

Smith, W. L. (1972), Satellite techniques for observing the temperature structure of the atmosphere, Bull. Am. Meteor. Soc., 53, 1074-1082.

Stephens, G. L. (1994), Remote Sensing of the Lower Atmosphere, An Introduction, 523 pp., Oxford University Press, New York.