Radio Science, Volume 35, Number 1, Pages 45–56, January–February 2000

An inversion algorithm for nonlinear retrieval problems extending Bayesian optimal estimation

Markus J. Rieder and Gottfried Kirchengast

Institute for Meteorology and Geophysics, University of Graz, Graz, Austria

Abstract. This paper proposes effective extensions to the well-known Bayesian optimal estimation, allowing one to cope not only with the ill-posedness but also with the intrinsic nonlinearity of many geophysical inversion problems. We developed a physical-statistical retrieval algorithm, which combines nonlinear optimal estimation with further optimization techniques. Profiling of water vapor based on (synthetic) downlooking microwave sounder data as an example for a typical geophysical nonlinear optimization problem is used to demonstrate the skills of the algorithm. Starting with a nonlinear scalar penalty function derived from a Bayesian approach, the sensible guess of a priori information, the selection of useful probability density functions, the advantages of simulated annealing, and the utility of Monte Carlo methods are discussed. These techniques together furnish capability for retrieving state vectors, which depend on the data in a (highly) nonlinear manner. The sensible combination as implemented in the introduced hybrid algorithm can provide solutions to problems that could not be tackled with standard (linearized) inversion methods properly.

1. Introduction

Estimating parameters about the state of a system from remotely measured data can be viewed as an inverse problem. Geophysical inverse problems in many cases cannot be solved directly because of their inherent ill-posedness and nonlinearity. Physical retrievals seem accurate when the physics of the problem can be formulated clearly and the forward model is not too nonlinear to find a global minimum of the penalty function. Retrievals of this kind are intended to invert the forward function describing the physics of the problem by constructing a mapping from measurement space into state space. On the contrary, statistical retrievals exploit the capacity of statistical interrelations between state and measurement but suffer from the lack of physical reasoning, constituting serious troubles in ill-posed environments. Retrievals of this kind are not based on an understanding of the physical principles linking state and measurement, so knowledge of the nature of the measurement process is not included in the retrieval process.

We propose a combination of physical and statistical tools to give a sensible hybrid algorithm, capable of providing stable and accurate retrievals despite ill-posedness and nonlinearity. After introducing the Bayesian approach to inversion in our context, we present a typical geophysical example of a nonlinear inverse problem, namely water vapor profiling from spaceborne downlooking sounder data. The algorithm provides a solution to the problem of retrieving water vapor profiles from vertical sounder measurements under clear conditions when the temperature profile is known.

We then address the shortcomings of straightforward solutions which exhibit the need for a more advanced hybrid algorithm, which is introduced in detail in sections 5 – 8. Finally, an exemplary result is discussed, and conclusions are given.

Copyright 2000 by the American Geophysical Union.

Paper number 1998RS002145.
0048-6604/00/1998RS002145$11.00

2. Bayesian Inversion

A remote measurement is one in which the quantity measured is a function of the parameters actually required. This function, in general termed the "forward model" is assigned to predict the results of measurements on the basis of the physical processes relevant to the problem under investigation. Inverse theory refers to the inversion of this function; thus inverse theory deals with mathematical techniques for retrieving useful information about the state of a physical system from (remotely) measured data. In other words, inversion addresses the reverse problem of the forward model: Starting with data obtained by an experiment, it is intended to estimate parameters of the physical state based on the knowledge of the model. In this work we implicitly assume to know the forward model correctly, so we do not consider problems related to forward modeling here.

The measurements are considered as a vector $\mathbf{y} = (y_1,y_2,...,y_m)^T$, the measurement vector, and the unknowns (often referred to as the model parameters, which are to be retrieved) are assembled into the state vector $\mathbf{x} = (x_1,x_2,...,x_n)^T$, describing the state of the system. Random error or "measurement noise" will be denoted by the vector $\boldsymbol{\varepsilon}$.

The physics of the problem (or, at least, some model of the physical processes) can be formally described as a forward model $\mathbf{f}$, determining for each state vector $\mathbf{x}$ a corresponding measurement vector $\mathbf{y}$. Therefore the relationship between the measurement vector and the state vector reads

$$\mathbf{y} = \mathbf{f}(\mathbf{x}) + \boldsymbol{\varepsilon} . \quad (1)$$

The Bayesian approach provides a framework to inspect the inverse problem. Given a measurement together with a description of its error statistics, a forward model describing the relation between the measurement and the unknown state, and any kind of a priori knowledge available, it allows one to assign a probability density function (pdf) to the possible states, consistent with the measurement and the a priori information.

Bayes' theorem reads, involving a pdf of the a priori state, $P(\mathbf{x})$, a pdf of the measurement, $P(\mathbf{y})$, and a conditional pdf of $\mathbf{y}$ given $\mathbf{x}$, $P(\mathbf{y}|\mathbf{x})$, respectively,

$$P(\mathbf{x}|\mathbf{y}) = P(\mathbf{y}|\mathbf{x})\, P(\mathbf{x}) / P(\mathbf{y}) . \quad (2)$$

The Bayesian approach yields a posterior conditional pdf of $\mathbf{x}$ given $\mathbf{y}$, $P(\mathbf{x}|\mathbf{y})$, for which one wants to find a maximum likelihood solution in order to obtain a "retrieval formula." In general, taking (-2 times) the logarithm of Bayes' theorem yields

$$-2\ln P(\mathbf{x}|\mathbf{y}) = -2\ln P(\mathbf{y}|\mathbf{x}) - 2\ln P(\mathbf{x}) + 2\ln P(\mathbf{y}) . \quad (3)$$

Inserting a Gaussian-distributed measurement error (denoting the corresponding covariance matrix as $\mathbf{S}_\varepsilon$) but keeping the pdf of the a priori state general and regarding $2P(\mathbf{y})$ as a constant c gives

$$-2\ln P(\mathbf{x}|\mathbf{y}) = [\mathbf{y}-\mathbf{f}(\mathbf{x})]^T \mathbf{S}_\varepsilon^{-1} [\mathbf{y}-\mathbf{f}(\mathbf{x})] - 2\ln P(\mathbf{x}) + c . \quad (4)$$

Equation (4) could formally be used to calculate the mean a posteriori state by calculating the first moment of the distribution. The integration over state space is a difficult task and works only for some explicit forms of $P(\mathbf{x})$ analytically, but the maximum likelihood state as an alternative estimator may be inferred. Taking the derivative of (4) with respect to $\mathbf{x}$ and equating the result to zero yields

$$\partial \ln P(\mathbf{x}|\mathbf{y})/\partial \mathbf{x} = \partial\{-\tfrac{1}{2} [\mathbf{y}-\mathbf{f}(\mathbf{x})]^T \mathbf{S}_\varepsilon^{-1} [\mathbf{y}-\mathbf{f}(\mathbf{x})]\}/\partial \mathbf{x} + \partial \ln P(\mathbf{x})/\partial \mathbf{x} = 0 . \quad (5)$$

Denoting the derivative of the forward model $\mathbf{f}$ with respect to the state $\mathbf{x}$, $\partial \mathbf{f}_i/\partial \mathbf{x}_j = \mathbf{K}_{ij}$, and rewriting the second term, (5) gives

$$\partial \ln P(\mathbf{x}|\mathbf{y})/\partial \mathbf{x} = \mathbf{K}^T \mathbf{S}_\varepsilon^{-1} (\mathbf{y}-\mathbf{f}(\mathbf{x})) + (1/P(\mathbf{x}))(\partial P(\mathbf{x})/\partial \mathbf{x}) = 0 \quad (6)$$

which is an implicit equation for the state in question, $\mathbf{x}_{retr}$. If we had some idea of the shape of the a priori state distribution, for example, by evaluating large data sets of observations, we could interpret the measurements more accurately (compare the analytical examples of *Jackson* [1985]). The implicit equation, equation (6), gives a device by which to calculate the a posteriori pdf of the state, given the measurement and the a priori distribution. In fact, the main reason why (6) cannot be readily used as a retrieval formula for a general a priori pdf is the lack of a useful analytical description and of a convincing assumption on the shape of $P(\mathbf{x})$ in the case of nontrivial and non-Gaussian pdfs.

When using Gaussian-shaped pdfs for measurement and a priori error distributions, the theorem of Bayes can be used to derive a scalar cost function χ^2, which enables one to describe the inverse problem as an optimization problem of finding the minimum of this scalar function [cf. *Rodgers*, 1976],

$$\chi^2 = (\mathbf{y}-\mathbf{f}(\mathbf{x}))^T \mathbf{S}_\varepsilon^{-1} (\mathbf{y}-\mathbf{f}(\mathbf{x})) + (\mathbf{x}-\mathbf{x}_{ap})^T \mathbf{S}_{ap}^{-1} (\mathbf{x}-\mathbf{x}_{ap}) \quad (7)$$

The assumption of Gaussian-shaped pdfs is a very reasonable one for measurement errors (covariance

matrix $\mathbf{S}_\varepsilon$), as we nominally can reasonably consider them as unbiased (nonsystematic) and random. For a priori errors (covariance matrix $\mathbf{S}_{ap}$) it is valid only in some particular problems, but many problems are rather poorly approximated by a Gaussian a priori error distribution; this may lead to estimated states that are influenced by the a priori state in an inappropriate manner [cf., e.g., *Jackson*, 1985].

Thus, when being confronted with the optimization problem of (7), two difficulties have to be kept in mind: On the one hand, too severe influences of the a priori statistics on the retrieval should be avoided; on the other hand, techniques to deal with the nonlinearites have to be employed, since it is not possible, in general, to uniquely develop a simple retrieval formula based on (7) in the case of a nonlinear forward model **f**.

3. A Nonlinear Retrieval Problem: Water Vapor Profiling

The term "nonlinearity" denotes that the forward model mapping the state vector from state space into the measurement space is a nonlinear function of the state. The Bayesian approach by no means is restricted to the linearity of the forward model, although some of the widely used specific formulae of optimal estimation are strictly valid only in the case of a linear forward model [cf. *Jackson,* 1985].

Profiling constituents of the atmosphere using spaceborne downlooking sounder data is a typical geophysical example of a nonlinear optimization problem [see, e.g., *Wang et al.,* 1983; *Ulaby et al.,* 1986]. We used the Special Sensor Microwave Humidity Sounder (SSM/T-2) data together with total precipitable water (TPW) data for the purpose of water vapor profiling (brightness temperatures around the 183.31 GHz water vapor absorption line; see *Littlejohn* [1995] for a description of the sensor). As described by *Grody* [1993], the weighting functions of the sensor SSM/T-2 are considerably dependent on the actual state, implying substantial difficulties for straightforward linearized retrieval schemes.

For this water vapor profiling example the state is decribed using eight relative humidity values (RH values) at eight height levels, spaced 2 km in the vertical; the humidities represent average values for the corresponding 2-km layers. Using eight layers to describe a water vapor profile is enough to roughly determine the humidity structure while still being roughly consistent with the sensor's information content. The measurement vector **y** comprises the brightness temperatures of SSM/T-2 (five channels) and the total precipitable water content. Consequently, the forward operator **K** has to consist of the weighting functions corresponding to the five channels and, in the last row, an integration over the actual state values in order to obtain the total precipitable water (TPW). For the convenience of treating the elements of the measurement vector as quantities with homogenous physical dimensions (Kelvin) and because of its intrinsically linear contribution to **K**, we skip the TPW (which occupies the last row of the matrix **K** and the the sixth element vector **y**) contribution in the nonlinearity discussion closing this section.

When trying to establish a forward model operator in matrix form, **K**, this matrix is considerably dependent on the actual state **x**. Nevertheless, linearization about a certain state is possible, and (1) can be written as

$$\mathbf{y} = \mathbf{K}(\mathbf{x})\,\mathbf{x} + \boldsymbol{\varepsilon}\,. \tag{8}$$

One way to investigate the nonlinearity of a problem is, for example, to compare the singular vectors of the forward model operator $\mathbf{K}(\mathbf{x})$ at different states **x**. Figure 1, which serves to indicate the nonlinearity of the problem at hand, shows the singular vectors of the matrix **K** at two different states **x**, one of which corresponds to rather high, the other to rather low, relative humidity (RH) values.

4. A Straightforward Solution: Marquardt-Levenberg Iteration

Given the cost function (7), derived from a Bayesian approach to inversion under the assumption of Gaussian statistics for the errors involved, the minimum of this function has to be found. This could (in principle) be done by an iteration algorithm, whereby at each step of iteration the complete forward model $\mathbf{f}(\mathbf{x})$ is replaced by the linear term of its Taylor series expansion, a method that is known as Newtonian iteration. However, if the solution is sufficiently far from the state found for a given iteration step, it is possible that the step taken under the assumption of a linear approximation fails to improve the solution and may even increase rather than decrease the residual. Under these circumstances a so-called steepest-descent approach is safe, though

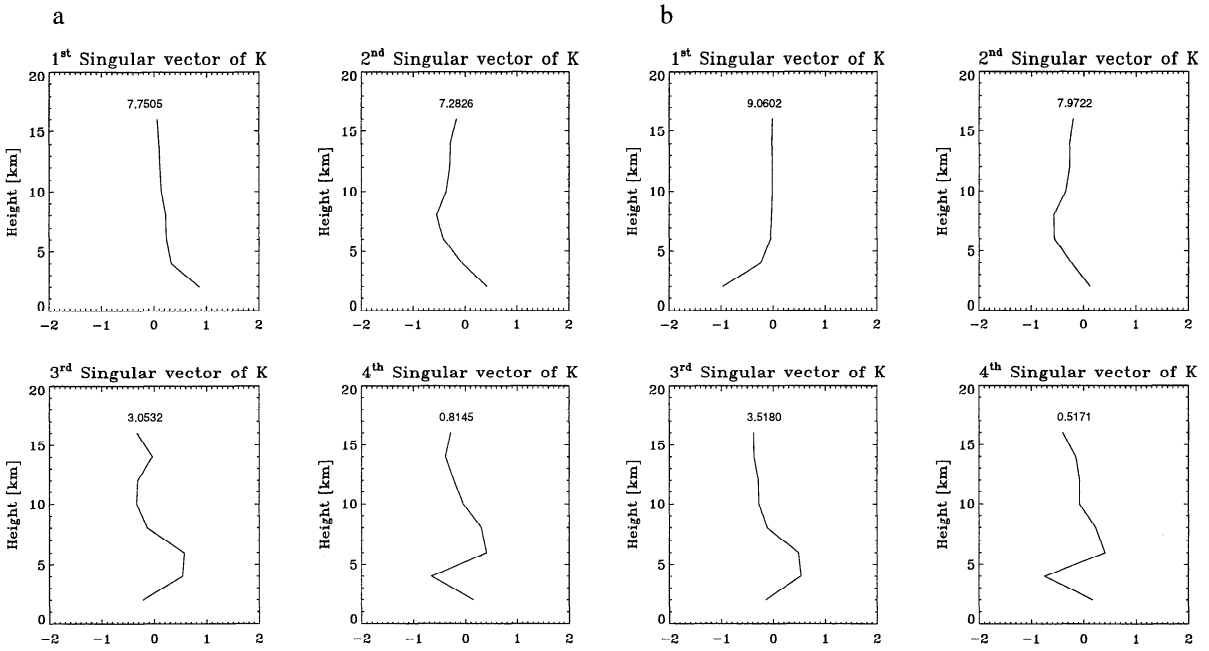

Figure 1. Singular vectors and corresponding singular values for the forward model operator **K** obtained using a matrix **K** for a state with (a) high relative humidities (b) low relative humidities.

comparatively slow. This gradient method simply steps off from the current trial value in the direction of the negative gradient of χ^2.

The Marquardt-Levenberg algorithm [*Marquardt*, 1963] is a way of combining steepest-descent and Newtonian iteration in a very favorable way. The algorithm shares with the Newtonian iteration the ability to close in on a converged value rapidly after the vicinity of the converged value has been reached. The algorithm shares with the gradient methods the ability to converge from an initial guess which may be outside the region of convergence of other methods because of the nonlinearity. Thus the method combines the favorable features of its predecessors while avoiding their most serious limitations [*Marquardt*, 1963].

We employed the Marquardt-Levenberg iteration in order to find the minimum of the cost function in a mathematically rigorous way. The iteration procedure is demonstrated using synthetic data cases, where we know the true profile and want to estimate the retrieved profile based on the corresponding (simulated) measurements. The calculated measurements were perturbed by white Gaussian noise compliant with the variance specified within the measurement error covariance matrix $\mathbf{S}_\varepsilon$.

Figure 2 shows the a priori profile, the true profile, and the retrieved profile at each step of iteration (for three iterations) together with the actual differences between the true and the retrieved profile. Furthermore, the value of the χ^2 cost function is presented in each plot of Figure 2. The true profiles are some typical examples of U.S. Standard Atmosphere profiles [see *Bilitza*, 1992, and references therein], and the a priori profiles are humidity profiles from the TIGR/IASI (Thermodynamic Initial Guess Retrieval / Infrared Atmospheric Sounding Interferometer) data set [*Chédin et al.*, 1985]. The a priori profile was the best estimate after employing a library method of scanning all stored profiles as to how well they match with their corresponding (forward modeled) measurement vector, the observed measurement vector.

The optimization problem (7) can be viewed, in a geometrical interpretation, as the search for the minimum of a multidimensional hypersurface described by the cost function χ^2. In the linear case this surface is of parabolic shape, which clearly has one global minimum that can be readily found applying standard methods. In the nonlinear case this surface might be distorted, and thus it is conceivable that there is more than one minimum, a

situation we refer to as the existence of multiple (local) minima. In such a case, where the nonlinearities are too gross, severe problems may arise when trying to use a straightforward optimal estimation solution like Marquardt-Levenberg iteration.

In the case of (strong) dependence of the retrieved state on the a priori estimate, as is true for ill-posed, underdetermined inversion problems, the a priori estimate is still well reflected in the retrieval after applying a physical algorithm. That is, many inversion algorithms provide retrieved states that resemble closely in shape the a priori states, which is an implication of the second term in the cost function (7), which gives considerable contributions to the cost function value the farther the estimated state $\mathbf{x}_{retr}$ is from an a priori state $\mathbf{x}_{ap}$. This clearly can be seen in Figure 2, where the a

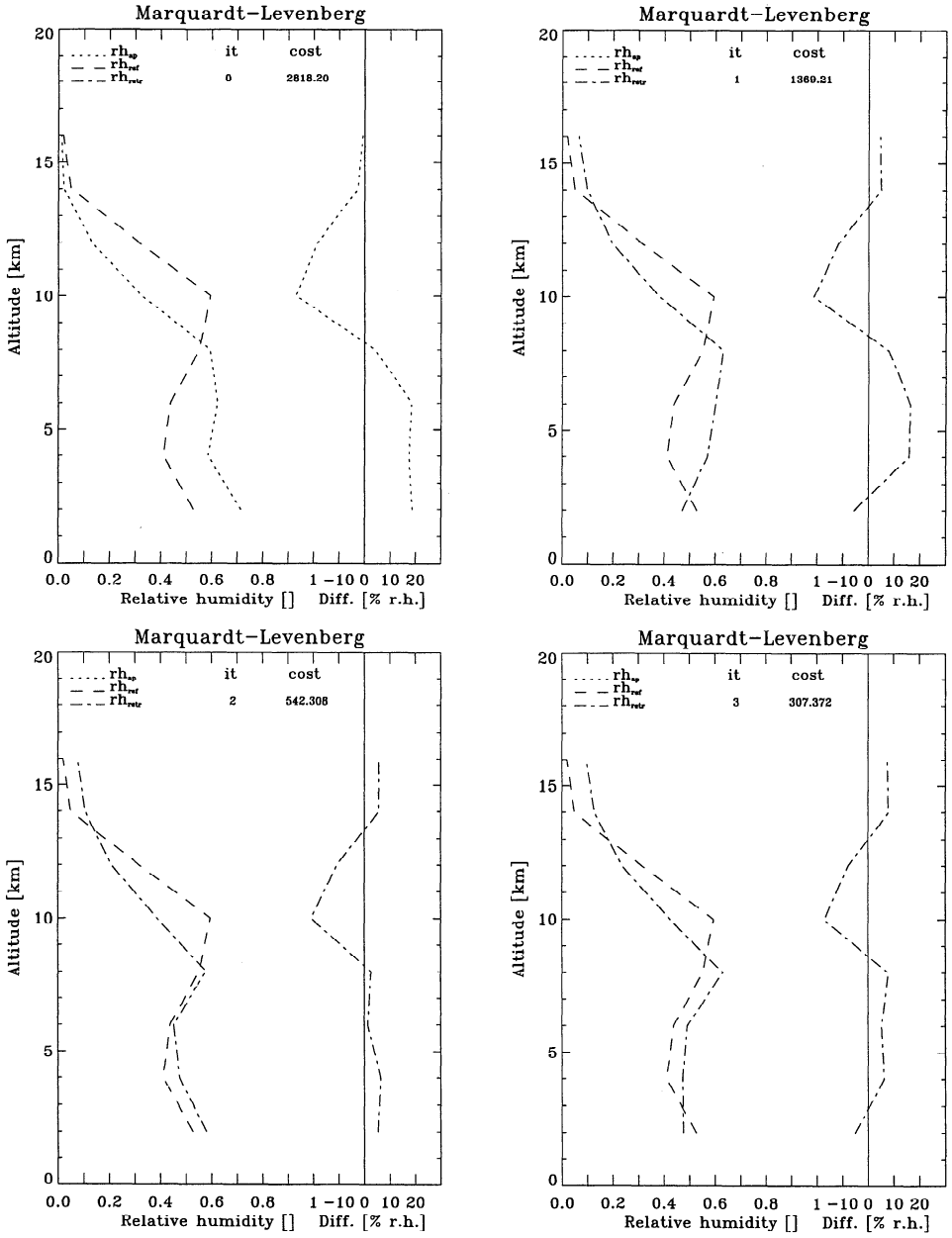

Figure 2. Example of the Marquardt-Levenberg iteration. For description, see text.

priori state has been used to start the algorithm and the retrieved states roughly show the same shape as the a priori state.

Another drawback of a physical solution like the Marquardt-Levenberg iteration is the need for calculating the weighting function matrix **K** explicitly. In a flat hypersurface χ^2 the gradients (and thus the changes in **K** from one iteration step to the next) are numerically very small, so that a gradient (iteration) method would get stalled at some point. The calculation of **K**, usually implemented by a numerical perturbation of the forward model, could induce severe inaccuracies in the course of iteration, followed by errors in the solution.

Statistical algorithms, like simulated annealing and Monte Carlo iterative a priori lightweighting described in sections 7 and 8, need neither any theoretical input of inversion theory nor a computation of the gradient of the cost function [cf. *Marchuk et al.,* 1980]. They are not an inversion method in the sense of constructing a kind of inverse mapping from measurement space into state space at all but are essentially an iterative application of the forward model with the goal of finding a minimum of a scalar measure of the fit, while being blind to a "physically preferrable direction" of minimization of the cost function.

5. A Hybrid Algorithm

Even when employing a subtle minimization technique like the Marquardt-Levenberg iteration outlined in section 4, every inversion is faced with three difficulties at the same time: First, in cases where the inherent nonlinearities are too high, multiple minima may arise; second, the ill-posedness of some problems may lead to solutions too near to the a priori estimate; and third, calculation of the weighting functions is numerically unfavorable. As a promising and suitable solution to all of these problems we introduce the concept of a hybrid algorithm, which combines physical and statistical retrieval tools in a sensible manner.

We established a hybrid algorithm with the following subsequent properties: (1) Finding a proper a priori estimate; this is best accomplished by a search in a catalogue of representative states. (2) Assuring that one goes in the correct direction when starting at some a priori state; this is best accomplished by a "physical" part in the algorithm. Utilization of the Marquardt-Levenberg iteration well serves this need. (3) Further improving the physical solution by a technique which furnishes capability of escaping some local minima of the cost function; this is best accomplished by temporarily allowing increased cost values during an iteration. Utilization of (multiple) simulated annealing well serves this need. (4) Improving the solution by relaxing its strong resemblance to the a priori profile; this is best accomplished by modifying the cost function in terms of changing the weight of the a priori estimate. Utilization of a Monte Carlo method, which we term iterative a priori light weighting, well serves this need.

The tools described above all together are organized as follows into a hybrid algorithm. The algorithm starts with the search for a suitable a priori state out of a catalogue, containing independent information on the system, i.e., comprising a set of well-founded representative states. The best states, judged in terms of their cost function value, are stored. Five different a priori profiles were found to be enough to ensure that a correctly retrieved state can be found, finally. The following steps are done for each of the five different a priori states individually.

After the a priori state has been found, the Marquardt-Levenberg iteration algorithm begins. Once the convergence criterion for this algorithm is met, a simulated annealing technique (in a multiple version) is applied, starting with the Marquardt-Levenberg solution. When the simulated annealing algorithm ends, a modified cost function, equation (9), discussed in section 8, is established, and an iterative a priori light weighting Monte Carlo method begins. After examining the neighboring states for an improved cost (judged by the modified cost function value), a χ^2 test for correct convergence is done. Hereafter, the algorithm is run starting with the next a priori state.

Finally, all (correctly converged) retrieved states of the five alternative runs of the algorithm are compared, and the best of all correct states, indicated by minimum cost, is selected as the best estimate $\mathbf{x}_{retr}$. While we have discussed above the classical physical part of the retrieval, the Marquardt-Levenberg iteration, the sections 6 – 9 describe the statistical part of the hybrid algorithm in some detail.

6. Choice of an A Priori State

A careful choice of the a priori state is of major importance for the power of the algorithm for two

reasons. First, the a priori state serves as the initial state of the Marquardt-Levenberg iteration; thus it is a single selected state that determines the final state of the iteration, insofar as local minima may occur that are accepted by the iteration as solutions. Second, the a priori state contributes a considerable fraction to the value of the cost function itself; thus we have to expect that every retrieval reflects more or less the overall features of an a priori state, which can be investigated in detail by a characterization of the retrieval method following *Rodgers* [1990].

Regarding the examples of water vapor profiling demonstrated in section 9, a subset of the TIGR data set, the "TIGR/IASI data set" [*Chédin et al.*, 1985], which assembles profiles assigned to some geographical and temporal conditions, was taken to provide the a priori state $\mathbf{x}_{ap}$. Thus by use of a sample of representative profiles for the specific case at hand we made use of some measurement-independent a priori information that is climatologically well founded.

We used a simple library search method to find an a priori state which is statistically reasonable in terms of a moderate cost value; that is, we searched the TIGR/IASI catalogue of representative profiles for profiles which give brightness temperatures that are close to the measured brightness temperatures. The search through a library of probable states is a useful tool to ensure that some typical states are identified through typical features in the measurement vector. In other words, a library search is a possibility to exclude physically very improbable states which could also closely correspond to the measurement because of the ill-posedness of the problem (within a certain range of uncertainty because of the measurement error).

The disadvantage of this search method as a stand-alone retrieval technique, as it is sometimes applied, is the missing physical reasoning when relating the measurement to a state. There is no need to explicitly invert the forward model; thus the pure statistical nature of such a library method could actually lead to ambiguities in cases where the state is not clearly related to the measurements (which is an inherent property of an ill-posed, underdetermined retrieval problem). In the case of a nonlinear problem this ambiguity is even more substantial; local, spurious minima can occur, and therefore completely different states may have the same value of the cost function χ^2.

Nevertheless, for our hybrid algorithm the lowest χ^2 values found by such a search correspond to reasonable profiles suitable to be used as a priori states. It is sensible to start with a small ensemble of a priori states. As noted above we used the five ones that have the lowest cost function value. This is a means to avoid spurious minima and will appear successful in some cases (where such false minimum is circumvented). For our profiling problem we found in most cases the different starting states yielding the same solution, a redundancy well indicating the robustness of the "choice of an a priori state" part of the hybrid algorithm.

7. Simulated Annealing

The method of simulated annealing is a tool to be applied in cases of (highly) nonlinear optimization problems, since it provides an explicit possibility to escape from local minima. The simulated annealing algorithm is based on a simulation of the process of finding the ground state of a solid [*Laarhoven*, 1996]; annealing denotes a physical process by which, if carried out sufficiently slowly, the ground state of a solid can be found.

Given the current state, a small perturbation is applied. The size of the perturbation is randomly chosen to be compliant with the magnitude of the a priori variances. If the perturbation results in a lower χ^2 value, the process is continued with the new state. Even if the χ^2 value after the state transition is greater than the value before the transition, the probability of acceptance of the perturbed state is nonzero: This rule for accepting new states establishes a possibility to escape from local minima by temporarily also allowing, contrary to strict minimization, for cost increases in the course of iterations.

We applied a multiple annealing technique by carrying out the simulated annealing process more than one time [cf., e.g., *Liu et al.*, 1995]: Starting with a state, the simulated annealing technique was applied, and the cost function value of the retrieved state was stored. Next, a "heat up" of the state was simulated by applying some perturbation to the best estimate of all previous steps. Subsequently, the annealing algorithm started again. The best χ^2 value of all these annealings was taken to be the one assigning the best estimated state.

We tested the simulated annealing algorithm in order to demonstrate its usefulness within a cost function topology in state space which is rather flat. One example of the performance of the simulated annealing algorithm is given in Figure 3. Figure 3 demonstrates that the

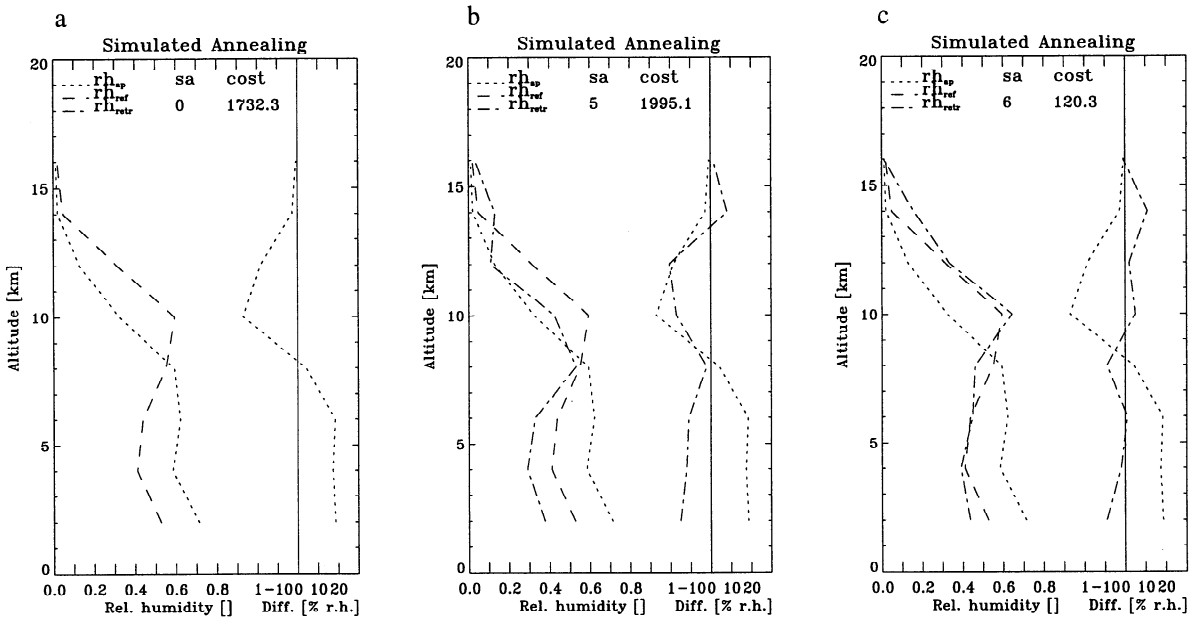

Figure 3. Escape from a spurious minimum. All plots show the a priori profile (rh_{ap}), the true profile (rh_{ref}), and the difference between the a priori profile and the true profile. The solution after a certain step (steps 5 and 6) is the dash-dotted profile rh_{retr}.

simulated annealing algorithm is capable of escaping from a local minimum. The simulated annealing algorithm takes a step leading to an increased cost value, followed by a step of an improved fit, which significantly surpasses the fit of the previous steps. Figure 3b shows the case where a profile clearly is less well fitted to the measurements after some iteration step (step 5), as the cost function value indicates. Nevertheless, the transition was accepted, and we end finally at a better state (after step 6). This is exactly what is meant by "escape from a local minimum."

A transition like the one to the final state above could take place as well when applying some other Monte Carlo technique, like, e.g., the iterative improvement technique described in section 8; clearly, there could occur a random perturbation ready to change the profile from the state at an earlier iteration step directly to the state of iteration step 6. In practice, however, the probability of such an occurrence is very low. It has been our experience that the property of also accepting steps of increasing cost function values very much increases the probability of finding final states at lower costs (like the one in Figure 3c) more efficiently.

8. Monte Carlo Iterative A Priori Light Weighting

Trying to access the inversion problem using only the Marquardt-Levenberg iteration was experienced to be insufficient in most cases, as demonstrated in Figure 2. On the one hand, the problem of running into a local minimum has to be faced, which is avoided well by use of a simulated annealing algorithm. On the other hand, a solution which is not influenced too much by the a priori estimates is desirable. A proper way to achieve this goal, while at the same time solving the inverse problem, is the attempt to minimize the cost function (defined by equation (9)) by a random or quasi-random examination of state space while at the same time performing a priori light weighting, i.e., putting increasingly less weight on the a priori information.

The Monto Carlo iteration works basically as follows: Starting off at a given state, a sequence of trials is generated. In each trial a state is selected from the neighborhood of the current state. If this neighboring state has lower cost, the current state is replaced by this

neighbor, otherwise another neighbor is selected and compared for its cost value. The algorithm terminates when a state is obtained whose cost is not worse than that of any of its neighbors. If this criterion is computationally not effective, the algorithm terminates when a preset number of iterations is exceeded. The run time of such an iterative improvement is proportional to the number of trials, and thus it is important to employ a sensible definition of "neighborhood" and "state transition."

As the infinite members of state space are theoretically all equal before calculating the cost, one needs to employ some prior knowledge of the structure of the probability density of state space. A neighborhood region in the form of a (Gaussian-weighted) volume in n-dimensional state space is practicable, where the previously achieved retrieval state of the simulated annealing algorithm serves as a starting state (center of region) and the covariance matrix of the Marquard-Levenberg algorithm is used to estimate the covariance of the Gaussian pdf.

In order to circumvent the problem of arriving at states that resemble the a priori state rather closely (compare Figure 2) despite involving different measurements, a modified cost function is defined, in terms of which the improvement of states is examined. The modified cost function is a χ^2 function that gives less weight to the term that measures the difference between the estimated state and the a priori state: We thus term this method "a priori light weighting." Mathematically, this modified cost function reads

$$\chi^2 = [\mathbf{y} - \mathbf{f}(\mathbf{x}_{retr})]^T \mathbf{S}_\varepsilon^{-1} [\mathbf{y} - \mathbf{f}(\mathbf{x}_{retr})] + c_r (\mathbf{x}_{retr} - \mathbf{x}_{ap})^T \mathbf{S}_{ap}^{-1} (\mathbf{x}_{retr} - \mathbf{x}_{ap}) . \quad (9)$$

The modified cost function value is a linear combination of two quadratic measures of deviations; as we want to fit the measurement in the first place, we set the here-introduced regularization parameter $c_r < 1$. This "a priori light weighting factor" c_r determines the relative weighting between the measurement fit and the a priori fit; decreasing it puts comparatively less weight on the a priori state.

The factor c_r in (9) is continuously decreased during the Monte Carlo iteration process, which corresponds to enhancing the variances of the a priori state covariance matrix $\mathbf{S}_{ap}$ by c_r^{-1} (at the same time, equation (9) is renormalized to keep costs for different c_r values comparable). The iteration was started with a light weighting factor close to unity, $c_r = 0.9$, in order to initially match (7) closely. Subsequently, the factor was decreased to approach $c_r = 0.3$ at the end of the iteration. This minimal c_r value, corresponding to an enhancement of the variances in $\mathbf{S}_{ap}$ by about a factor of 3, was fixed empirically on the basis of retrieval tests for our profiling problem (other problems may require a different minimal c_r value). The search for the minimum was limited to a maximum of 100 iterations for the sake of computational efficiency. One hundred trials resulted in typically 8 to 10 improving steps, the actual number depending on the fit or misfit of the initial profile. This was experienced to be generally enough for a well-converged fit.

Figure 4 demonstrates the use of this Monte Carlo iterative a priori state light weighting technique. The top left plot shows the profile at the beginning of the algorithm ($\mathbf{x}_{ap}$); then, from top left to bottom right, the plots show subsequently improved profiles (improved in terms of the cost function) together with the corresponding χ^2 values of the cost function. The χ^2 values are multiplied by a factor of 100 for convenience.

9. The Hybrid Algorithm: An Exemplary Result

The hybrid algorithm has been implemented in a way that it is able to provide, in our example inversion problem, humidity profiles from vertical soundings in a numerically stable and computationally efficient manner. It should work well also on different nonlinear inverse problems because of its general setup allowing for any nonlinear function. Employing physical as well as statistical tools together to optimize a Bayesian cost function in a hybrid manner overcomes standard methods of inversion based on a single fixed retrieval tool only, since the hybrid approach is capable of accessing problems that were too nonlinear for standard optimal estimation.

As an example, Figure 5 shows a typical retrieval result as it occurred during the routine application of the hybrid algorithm. The pair of plots illustrates a retrieval of the water vapor relative humidity profile over sea, where we used all five channels of the SSM/T-2 sensor for the retrieval procedure.

We note that we tested the hybrid algorithm for water vapor profiling based on SSM/T-2 sounder data also by a statistical comparison of SSM/T-2 data with European Centre for Medium-Range Weather Forecasts (ECMWF) analyses [*Rieder and Kirchengast*, 1999]. This statistical validation indicated fairly satisfying water vapor profile

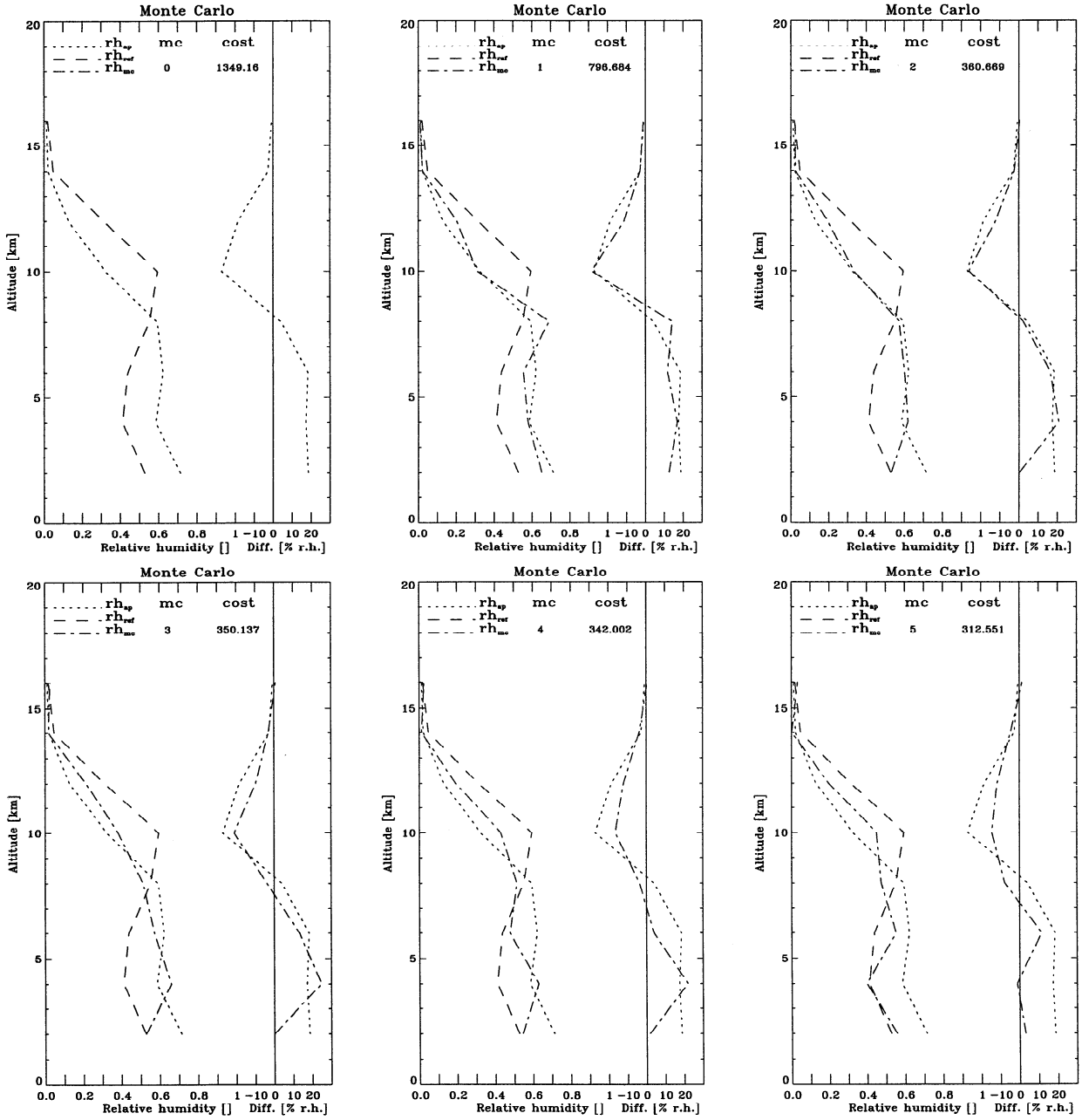

Figure 4. Example for the Monte Carlo iterative a priori light weighting technique. For description, see text.

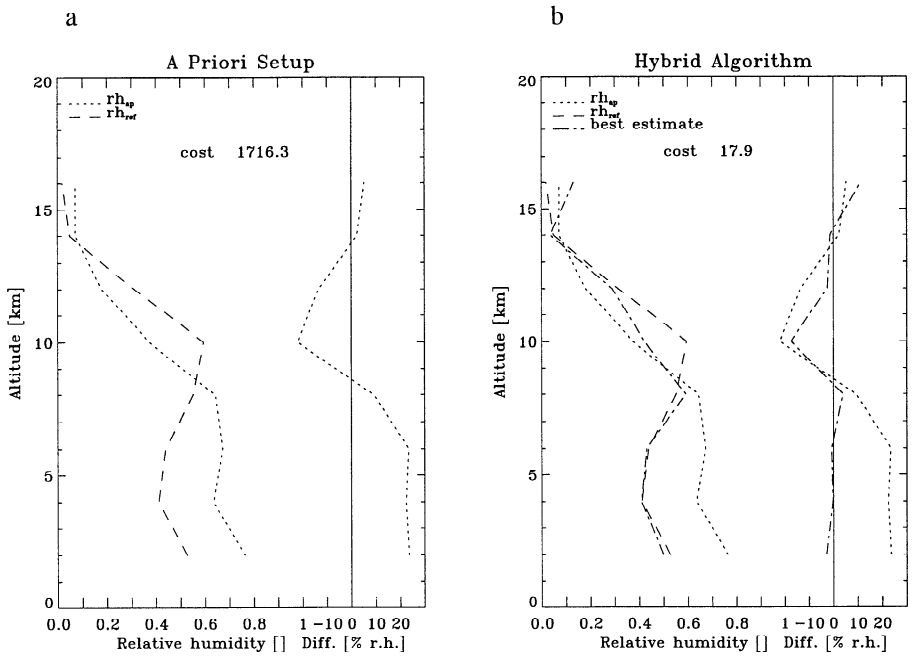

Figure 5. Example of a retrieval employing the hybrid algorithm. (a) The setup of the problem after the library search (a priori profile selected), and (b) The finally achieved solution, the retrieved state (best estimate). The corresponding cost value of the χ^2 cost function is denoted in each plot.

accuracy (~15% for retrievals over sea), confirming the high utility of the algorithm.

10. Conclusions

An advanced physical-statistical retrieval algorithm for nonlinear ill-posed inverse problems has been introduced and implemented using the problem of water vapor profiling based on passive downlooking sounder data as an example. This hybrid algorithm involves a Bayesian penalty function approach, including a sensible guess of a priori information by library search and iterative minimization by the Marquardt-Levenberg method, which is complemented and enhanced by statistical optimization techniques including simulated annealing and Monte Carlo iterative a priori light weighting. The latter methods are capable of coping with nonlinearities and, at the same time, of significantly reducing dependence on a priori information. The hybrid algorithm also allows the retrieval of states in a robust and efficient manner for cases where standard inversion methods fail and appears to be capable of yielding accurate results under such conditions.

Acknowledgments. We acknowledge the support of N. Scott, LMD, Francc, and M. Matricardi, ECMWF, United Kingdom, for providing the TIGR/IASI atmospheric data set. Thanks are due to C. Simmer, Institut für Meereskunde, Kiel, who provided his radiative transfer model MWMOD utilized in this study. Furthermore, we want to express our thanks for the satellite data assembled by the sensor SSM/T-2, kindly provided by the National Oceanic and Atmospheric Administration (NOAA). The first author (M. J. Rieder) received financial support for this work from IMG/UoG's "ATFERN" discretionary funds and from the START Programm Y103-CHE (START-ATCHANGE) research award of the Bundesministerium für Wissenschaft und Verkehr, Vienna, Austria. Additionally, support by the Dr.-Heinrich-Jörg-Stiftung, Karl-Franzens-Universität Graz, Austria, is acknowledged.

References

Bilitza, D., Solar-terrestrial models and application software, *Planet. Space Sci.*, *40*, 541-579, 1992.

Chédin, A., N. A. Scott, C. Wahiche, and P. Moulinier, The improved initialization inversion method: A high resolution physical method for temperature retrievals from satellites of the TIROS-N series, *J. Clim. Appl. Meteorol.*, *24*, 128-143, 1985.

Grody, N. C., Remote sensing of the atmosphere from satellites using microwave radiometry, in *Atmospheric Remote Sensing by Microwave Radiometry*, edited by M. A. Janssen, *Wiley Ser. Remote Sens.*, 572, 1993.

Jackson, D. D., and M. Matsu'ura, A Bayesian approach to nonlinear inversion, *J. Geophys. Res.*, *90* (B1), 581-591, 1985.

Littlejohn, K., System description document for the DMSP SSM/T-2 processing system, Natl. Oceanic and Atmos. Admin. Natl. Environ. Satell. Data and Inf. Serv., Silver Spring, Md., 1995.

Liu, P., S. Hartzell, and W. Stephenson, Non-linear multiparameter inversion using a hybrid global search algorithm: Applications in reflection seismology, *Geophys. J. Int.*, *122*, 991-1000, 1995.

Marchuk, G. I., G. A. Mikhailov, M. A. Nazareliev, R. A. Darbinjan, B. A. Kargin, and B. S. Elepov, *The Monte Carlo Methods in Atmospheric Optics*, Springer-Verlag, NewYork, 1980.

Marquardt, D. W., An algorithm for least-squares estimation of nonlinear parameters, *J. Soc. Ind. Appl. Math.*, *11* (2), 1963.

Rieder, M. J., and G. Kirchengast, Physical-statistical retrieval of water vapor profiles using SSM/T-2 sounder data, *Geophys. Res. Lett.*, *26* (10), 1397-1400, 1999.

Rodgers, C. D., Retrieval of atmospheric temperature and composition from remote measurements of thermal radiation, *Rev. Geophys.*, *14* (4), 609-624, 1976.

Rodgers, C. D., Characterization and error analysis of profiles retrieved from remote sounding measurements, *J. Geophys. Res.*, *95* (D5), 5587-5595, 1990.

Ulaby, F. T., R. K. Moore, and A. F. Fung, *Microwave Remote Sensing, Active and Passive*, 3, Book-Mart Press, North Bergen, N. J., 1986.

van Laarhoven, P. J. M., Theoretical and computational aspects of simulated annealing, CWI Tract, Centr. voor Wiskunde en Inf., Amsterdam, 1996.

Wang, J. R., J. L. King, T. T. Wilheit, G. Szeijwach, L. H. Gesell, R. A. Niemann, D. S. Niver, B. M. Krupp, and J. A. Gagliano, Profiling atmospheric water vapor by microwave radiometry, *J. Clim. Appl. Meteorol.*, *22*, 779-788, 1983.

G. Kirchengast and M. J. Rieder, Institute for Meteorology and Geophysics, Halbaerthgasse 1, A-8010 Graz, Austria. (gottfried.kirchengast@kfunigraz.ac.at; markus.rieder@kfunigraz.ac.at)

(Received December 17, 1998; revised June 2, 1999; accepted October 11, 1999.)