MSR, a multi-spectrum retrieval technique for spatially-temporally correlated or common Venus surface and atmosphere parameters

David Kappel*

German Aerospace Center (DLR), Institute of Planetary Research, Rutherfordstrasse 2, 12489 Berlin, Germany

Abstract

A common way to regularize mathematical ill-posed retrieval problems in atmospheric remote sensing is the incorporation of single-spectrum Bayesian *a priori* mean values and standard deviations for the parameters to be retrieved, along with measurement and simulation error information. This decreases the probability to obtain unlikely parameter values. For a reliable evaluation of measurements with sparse spectral information content, like Venus' nightside emissions in the infrared as acquired by the VIRTIS-M-IR instrument aboard ESA's Venus Express spacecraft, it can help to consider further *a priori* knowledge.

A new multi-spectrum retrieval technique (MSR) is presented that allows to incorporate expected correlation lengths and times for the retrieval parameters used to describe several spectra. It is demonstrated by examples that this decreases the probability to retrieve spatial-temporal state vector distributions that are incompatible with these *a priori* spatial-temporal correlations. Also, *a priori* correlations between the parameters used to describe a single spectrum and exhibiting similar *a priori* spatial-temporal behavior, act to rule out unlikely single-spectrum state vectors. Parameters with infinite correlation length or time and identic single-spectrum *a priori* data are spatially or temporally constant and can be retrieved as parameters that are common to a certain selection of measurements. This is shown to be especially useful to retrieve surface emissivity in the infrared as parameter that is common to several measurements that repeatedly cover the same target, and to determine deep atmospheric CO_2 opacity corrections, which are common to all Venus nightside spectra. Also this way, all considered measurements can be parameterized by a fully consistent set of atmospheric, surface, and instrumental parameters that respects all available *a priori* data as well as the measurement and simulation error distributions and that does not neglect the context between adjacent measurements. MSR is demonstrated to enhance the retrieval reliability and accuracy and pushes the VIRTIS-M-IR data evaluation to its limits.

Keywords: Remote sensing, Infrared spectroscopy, Venus, Multi-spectrum regularization, Correlation length, Common parameter

1. Introduction

The geology and composition of Venus' surface are topics of active research. There are only a few *in situ* measurements, performed by the VENERA probes [1]. Most areas appear to consist of basaltic material, but it is poorly classified. Knowledge of surface emissivity in the infrared can provide constraints for surface composition and weathering affected texture. Global surface emissivity maps can only be acquired by applying remote sensing techniques. A global topography, reflectivity, and emissivity data base, referred to 2.385 GHz, has been obtained by the Magellan mission [2]. However, a detailed analysis of Venus' geology requires input from spectral ranges that are more diagnostic to surface composition, like the infrared.

Venus' surface can not be directly observed in the visible and infrared. The hot surface (735 K at 0 km altitude when assumed to be in thermodynamic equilibrium with the bottom of the atmosphere according to the Venus International Reference Atmosphere VIRA [3, 4]) emits altitude dependent thermal radiation, as it does the hot deep atmosphere. This radiation is absorbed and multiply scattered by the gaseous and particulate components of the atmosphere, thereby strongly affecting the signals of the surface emissions that carry information about surface temperature and emissivity. The atmosphere is opaque with the exception of a few transparency windows between 0.8 and 1.3 µm that probe down to the surface. Additional windows between 1.3 and 2.6 µm are affected by the deep atmospheric temperature field and composition [5, 6]. Reflected sunlight strongly outweighs these emissions, thereby limiting the data usable for surface emissivity extraction in the infrared to nightside measurements.

ESA's planetary probe Venus Express (VEX) orbits the planet since 2006. The Mapping channel in the InfraRed of the Visible and InfraRed Thermal Imaging Spectrometer (VIRTIS-M-IR) aboard VEX acquires spectrally resolved (432 spectral bands uniformly dividing the range 1.0–5.2 µm) two-dimensional images of targets on Venus [7–9]. The carefully calibrated and preprocessed measurements [10, 11] provide the data base where the surface information shall be extracted from.

Hyper-spectral data can be quantitatively evaluated by

*Tel.: +49 30 67055 414

Email address: David.Kappel@dlr.de (David Kappel)

using a retrieval algorithm in conjunction with a detailed radiative transfer simulation model ('forward model'). The radiance spectrum that is detected by the space-borne measuring instrument, can be simulated by the forward model that numerically solves the radiative transfer equation. The result depends on the state parameters of atmosphere (altitude profiles of temperature and of gaseous and particulate constituents, absorption and scattering properties of the constituents), surface (elevation, temperature, emissivity), instrument (band-to-wavelength-mapping, full width at half maximum (FWHM) of the instrumental response function), and others (observational and illuminational geometry, O_2 nightglow, etc.). The parameters that shall be retrieved are compiled into the so-called 'state vector'. The retrieval algorithm compares the state-vector dependent simulation to the measurement and iteratively varies the state vector until the simulation well fits the measurement. The corresponding state vector then adequately parameterizes the measurement and is interpreted to represent the physical states of the atmosphere, surface, instrument, and others, that led to the measured spectrum. Herefore, a forward model is used, similar to that described by Haus and Arnold [12]. It is a plane-parallel, non-LTE, line-by-line code taking into account thermal emissions by surface and atmosphere, and absorption and multiple scattering by gases and clouds. Some additional noteworthy details are presented in Section 4.1.

But depending on spectral resolution and information content of the measurement and on complexity of the forward model, different state vectors can parameterize the same measurement equally well. Thus, this inversion of the radiative transfer equation is mathematically an ill-posed problem. The usual way to treat such problems is a regularization [13], for instance by defining *a priori* probability distributions, which the state vectors are assumed to follow. A convenient distribution is a Gaussian with certain mean value vector and covariance matrix. The *a priori* mean vector is defined to be the physically expected value of the state vector and the *a priori* covariance matrix its expected covariance matrix. The utilized *a priori* data for VIRTIS-M-IR retrievals is based on former observational *in situ* results gathered during the VENERA missions [1] and on the analysis of earlier ground based high-resolution data [5, 6, 14, 15] as well as on other space-borne experiments (limb observations [16], radio science [17]). The spectra also suffer from measurement and calibration errors on the one hand and from simulation errors on the other hand. This information can also be incorporated into the retrieval algorithm in order to arrive at a Bayesian interpretation [18] of the regularization. Here, the measurements and simulations, along with the *a priori* and error information, lead to an *a posteriori* probability distribution. The location of its maximum represents the best estimate of the state vector and has to be iteratively determined by the retrieval algorithm. The standard deviations of, and the correlations between, the retrieved parameters can be estimated through an approximation of the Hessian of this *a posteriori* probability distribution at the best estimate of the state vector. They are measures for the retrieval uncertainties and interferences. This single-spectrum regularization decreases the probability to obtain unlikely parameter values.

However, influence of noise, the presence of parameters with very similar impacts on spectra, as well as the possible existence of subsidiary solutions due to the complex non-linear dependence of radiance on the state vector may cause unexpected discontinuities in the spatial and temporal distribution of the retrieved parameters. Especially for measurements with sparse spectral information content, like the VIRTIS-M-IR measurements of Venus' nightside emissions, this may seriously degrade the reliability of retrieved single-spectrum parameters.

To overcome this problem, a multi-spectrum retrieval algorithm (MSR) is presented (Section 2) that allows for the utilization of additional *a priori* knowledge such as *a priori* spatial-temporal correlations. These are usually neglected but nevertheless always present, since contiguous measurements are unlikely to originate from completely unrelated state vectors. One pivotal aspect is the design of a suitable *a priori* covariance matrix (Section 3) that is positive definite by construction and allows for the encoding of the single-spectrum parameters' *a priori* standard deviations, spatial-temporal correlations, and local correlations (for instance as required for single-spectrum temperature profile regularization). In the limit of infinite correlation lengths or times for certain parameters, this covariance matrix will degenerate. Then, a special-case treatment can be derived that corresponds to the concept of retrieving parameters common to a selection of measurements (Section 3.5). Details of the forward model as well as retrieval parameters relevant for VIRTIS-M-IR measurements of Venus' nightside are presented in Section 4. Section 5 demonstrates by corresponding examples that MSR decreases the probability to retrieve unlikely parameter distributions, helps to avoid subsidiary solutions and to disentangle parameters with strongly differing spatial-temporal *a priori* correlations, compensates for noise effects, and allows to combine the information content of several spectra to determine hard-to-retrieve parameters that are common to a selection of measurements. A selection of mathematical details is presented in Appendix A, as well as some notes on the implementation of MSR. First results have been presented by Kappel et al. [11, 19, 20, 21], and the present paper develops the corresponding mathematical background.

2. Multi-spectrum retrieval algorithm (MSR)

This section recites the basics of Bayesian regularization as presented by Rodgers [18, Sections 2.3 and 5.2], but already in a generalized multi-spectrum retrieval formulation, and the involved quantities are defined. The required *a priori* covariance matrix will be constructed in Section 3.

Let $\mathbf{y}_i \in \mathbb{R}^{m_i}$, $i \in \{1, \cdots, r\}$ be the column vector representing the measured spectrum number i out of r measurements. The entry number $j \in \{1, \cdots, m_i\}$ of $\mathbf{y}_i$ shall be denoted as $(\mathbf{y}_i)_j$ and is the radiance acquired by measurement i at a certain wavelength or wavenumber. The dimensions m_i of the $\mathbf{y}_i$ may differ, to allow for the utilization of spectra from different data sources at varying spectral resolution and different spectral intervals. This also allows for the proper treatment of NaN ('Not a Number') values at certain wavelengths by simply ignoring these data points. Then let the 'extended measurement vector' $\mathbf{Y}$ be the concatenation $\mathbf{Y} = (\mathbf{y}_1^T, \cdots, \mathbf{y}_r^T)^T \in \mathbb{R}^M$, forming an element of the 'extended measurement space' of dimension $M = \sum_{i=1}^r m_i$, which is the direct sum of single-spectrum measurement spaces. For notational convenience, column vectors like $(\mathbf{y}_1^T, \cdots, \mathbf{y}_r^T)^T$ will be abbreviated as $(\mathbf{y}_1, \cdots, \mathbf{y}_r)$ in the following.

$\mathbf{X}$ shall denote the 'extended state vector' used to compute the 'extended simulation outcome vector' $\mathbf{F}(\mathbf{X})$. The iterative algorithm has to fit $\mathbf{F}(\mathbf{X})$ to $\mathbf{Y}$ by varying $\mathbf{X}$. $\mathbf{X}$ is the concatenation of r single-spectrum state vectors $(\mathbf{x}_1, \cdots, \mathbf{x}_r)$. In anticipation of Section 3.5 that introduces the concept of the retrieval of a parameter vector $\mathbf{x}_C \in \mathbb{R}^c$ common to r spectra, the extended state vector may also assume the form $\mathbf{X} = (\mathbf{x}_C, \mathbf{x}_1, \cdots, \mathbf{x}_r) \in \mathbb{R}^N$. All single-spectrum state vectors $\mathbf{x}_i$ are assumed to have the same dimension, $\mathbf{x}_i \in \mathbb{R}^n$, and thus $N = c + rn$, but generalization is straightforward. When the m_i-dependent single-spectrum simulation outcome for spectrum i is denoted as the column vector $\mathbf{f}_i \in \mathbb{R}^{m_i}$, then $\mathbf{F}(\mathbf{X}) \in \mathbb{R}^M$ can be written as $\big(\mathbf{f}_1(\mathbf{x}_C, \mathbf{x}_1), \cdots, \mathbf{f}_r(\mathbf{x}_C, \mathbf{x}_r)\big)$. In the following, for convenience, the adjective 'extended' will not be explicitly written anymore, except when it is not clear from context, whether a single-spectrum or a multi-spectrum object is referred to.

Similar to Rodgers [18], it may be assumed that the measurement, calibration, and simulation error distribution can be characterized by a Gaussian distribution in the residual $\mathbf{Y} - \mathbf{F}(\mathbf{X})$ between measurement $\mathbf{Y}$ and simulation $\mathbf{F}(\mathbf{X})$ of the measurement given the state vector $\mathbf{X}$. The corresponding error covariance matrix $\mathbf{S}_E$ is of dimension $M \times M$. A Gaussian distribution is usually a good approximation for real world errors (central limit theorem) and simple enough to derive useful formulas. Also, the probability distribution function with the least information content (meaning without implying further knowledge) that is consistent with the parameterization by a mean value vector and a covariance matrix, is the corresponding Gaussian [18]. Then the conditional probability distribution function for measurement $\mathbf{Y}$, provided that the state vector is $\mathbf{X}$, is given by

$$P_m(\mathbf{Y}|\mathbf{X}) = \frac{1}{N_1} \exp\left(-\frac{1}{2}\big(\mathbf{Y} - \mathbf{F}(\mathbf{X})\big)^T \mathbf{S}_E^{-1} \big(\mathbf{Y} - \mathbf{F}(\mathbf{X})\big)\right),$$

with normalization factor N_1. Due to the only limited knowledge on measurement errors and their relations across several measurements, $\mathbf{S}_E$ is assumed to be of block diagonal shape, such that the errors of different measurements are treated as independent. The individual blocks on the diagonal will be denoted as $\mathbf{S}_{\epsilon i}$ for the corresponding spectrum number i. Also, the $\mathbf{S}_{\epsilon i}$ are all assumed to be diagonal, such that the errors at different wavelengths are treated as independent of each other. The values on the diagonal are the variances of the errors (squares of the standard deviations) and shall be independent of $\mathbf{X}$. However, the $\mathbf{S}_{\epsilon i}$ may depend on i and the wavelengths of the corresponding measurements $\mathbf{y}_i$. For VIRTIS measurements, twice the spectrally resolved standard deviation of deep space observations is thereby a first estimation of the random error. Further errors may also enter the error covariance matrix, like those due to data calibration and preprocessing [11] as well as simulation.

The *a priori* probability distribution function the state vector $\mathbf{X}$ is assumed to follow, i.e. the probability density of $\mathbf{X}$ before knowledge of the outcome of the measurement, is analogously written as

$$P_P(\mathbf{X}) = \frac{1}{N_2} \exp\left(-\frac{1}{2}(\mathbf{X} - \mathbf{A})^T \mathbf{S}_A^{-1} (\mathbf{X} - \mathbf{A})\right).$$

N_2 is the normalization factor and $\mathbf{S}_A$ the extended *a priori* covariance matrix of dimension $N \times N$. If $\mathbf{S}_A$ is assumed to be a diagonal matrix, neither coupling between entries corresponding to the same single-spectrum state vector, nor between different single-spectrum state vectors, nor between common parameters is expected. In this case, the diagonal entries denote the individual *a priori* variances of the retrieval parameters. The better the knowledge of a certain parameter is, the smaller should the corresponding variance be set. For a detailed discussion of non-trivial *a priori* covariance matrices, see Section 3. $\mathbf{A} = (\mathbf{a}_C, \mathbf{a}_1, \cdots, \mathbf{a}_r) \in \mathbb{R}^N$ is the extended *a priori* mean value vector of the extended state vector $\mathbf{X} \in \mathbb{R}^N$. The *a priori* data is not very well known for retrieval problems related to Venus. For all practical purposes concerning the VIRTIS measurements, it is therefore not possible to set *a priori* data dependent on the measurement situation. Thus, $\mathbf{S}_A$ will not depend on $\mathbf{X}$, nor will $\mathbf{A}$, and $\mathbf{a}_i \in \mathbb{R}^n$ will not depend on i and will simply be denoted by $\mathbf{a}$.

Let P_M denote the probability distribution function of the extended measurement before it is made. It is not dependent on $\mathbf{X}$, but on $\mathbf{Y}$ only.

The *a posteriori* probability distribution function of $\mathbf{X}$ is the conditional probability distribution function $P_p(\mathbf{X}|\mathbf{Y})$ of the extended state vector $\mathbf{X}$, provided that the extended measurement vector is found to be $\mathbf{Y}$.

It follows from Bayes' theorem [18] that $P_p(\mathbf{X}|\mathbf{Y}) = P_m(\mathbf{Y}|\mathbf{X}) P_P(\mathbf{X}) / P_M(\mathbf{Y})$. With the abbreviation $F_c(\mathbf{X}) := -2 \log P_p(\mathbf{X}|\mathbf{Y}) - 2 \log N_3$, the term $F_c(\mathbf{X})$ reads

$$(\mathbf{X} - \mathbf{A})^T\, \mathbf{S}_A^{-1}\, (\mathbf{X} - \mathbf{A}) + \big(\mathbf{Y} - \mathbf{F}(\mathbf{X})\big)^T \mathbf{S}_E^{-1} \big(\mathbf{Y} - \mathbf{F}(\mathbf{X})\big), \tag{1}$$

where the normalization terms and $\mathbf{X}$-independent terms are all absorbed into N_3 and are unimportant for the retrieval.

Note that both $\mathbf{S}_E$ and $\mathbf{S}_A$ as covariance matrices have to be real, symmetric, and positive semi-definite. They even have to be positive definite to allow for the proper definition of the various probability distribution functions. Also note that $0 \leq F_c(\mathbf{X}) \in \mathbb{R}$, since $\mathbf{S}_E^{-1}$ and $\mathbf{S}_A^{-1}$ are consequently also positive definite.

The information on $\mathbf{X}$ is improved from $P_P(\mathbf{X})$ before measurement to $P_p(\mathbf{X}|\mathbf{Y})$ after measurement. Knowledge of $P_p(\mathbf{X}|\mathbf{Y})$ corresponds to the solution of the retrieval problem. It is useful to approximate $P_p(\mathbf{X}|\mathbf{Y})$ as function of $\mathbf{X}$ by a function parameterizable by a few characteristic parameters. Again, this suggests a Gaussian, the least-information-content function with mean value vector and covariance matrix, now to be derived from $P_p(\mathbf{X}|\mathbf{Y})$.

The mean value vector is approximated by the location of the (global) maximum $\widehat{\mathbf{X}}$ of $P_p(\mathbf{X}|\mathbf{Y})$. $\mathbf{F}$ can be assumed to be a continuously differentiable function of $\mathbf{X}$, provided that the basic input of the radiative transfer equation solver (altitude profiles of optical depth, single scattering albedo, Legendre moments of the scattering phase function, and temperature; some boundary conditions) depends differentiably on the retrieval parameters [22]. Also, $\mathbf{F}$ should be bounded for finite arguments to be physically reasonable. Thus, a necessary condition for $\widehat{\mathbf{X}}$ to be the location of a local maximum of $P_p(\mathbf{X}|\mathbf{Y})$ is its zeroing of the derivative of $P_p(\mathbf{X}|\mathbf{Y})$, or equivalently of F_c, with respect to $\mathbf{X}$. Since a local maximum of $P_p(\mathbf{X}|\mathbf{Y})$ corresponds to a local minimum of F_c, F_c is called the cost function of the retrieval problem, the function to be minimized. It must be kept in mind that due to the possibly complex nature of $\mathbf{F}$ as non-linear function on a high-dimensional state space, there may be more than one local maximum of $P_p(\mathbf{X}|\mathbf{Y})$, but the global maximum is the best estimate of the state vector. In $P_p(\mathbf{X}|\mathbf{Y})$, Gaussian damping by the *a priori* information $P_P(\mathbf{X})$ mathematically somewhat improves the identification of the best estimate. If the *a priori* covariances are small enough, this leads to the elimination of subsidiary solutions, at least in the extreme case where the standard deviations tend to 0. In this case, the solution is forced to attain the vector of the *a priori* mean values. But in practice, it must be kept in mind that a determined local maximum could be just a subsidiary maximum. In absence of a fast and reliable global minimizer, the retrieval algorithm will determine local minima of $F_c(\mathbf{X})$, see Appendix A.1, and as a rough test, it should be checked whether different initial guesses and *a priori* data lead to the same or a similar solution.

An expression for the width of $P_p(\mathbf{X}|\mathbf{Y})$ can be obtained by using the quadratic term in the Taylor expansion of F_c at $\widehat{\mathbf{X}}$. By neglecting the derivative of the Jacobian in comparison to the other terms, one arrives at

$$\widehat{\mathbf{S}}^{-1} = \mathbf{S}_A^{-1} + \mathbf{K}(\widehat{\mathbf{X}})^T \mathbf{S}_E^{-1} \mathbf{K}(\widehat{\mathbf{X}}) \tag{2}$$

of dimension $N \times N$, where $\nabla \mathbf{F}(\mathbf{X}) =: \mathbf{K}(\mathbf{X})$ is the Jacobian of dimension $M \times N$ of the forward model at $\mathbf{X}$. $\widehat{\mathbf{S}}$ will be interpreted as the covariance matrix of the *a posteriori* probability distribution function of the state vector $\mathbf{X}$ at the retrieved solution $\widehat{\mathbf{X}}$ [18]. The diagonal entries provide a measure of the uncertainty of the retrieved parameters, and the off-diagonal entries bear information on the interdependence of the parameters. However, it is still necessary to perform a detailed retrieval error characterization [23], as will be presented in a subsequent paper.

Appendix A discusses some details of the implementation of MSR.

3. *A priori* covariance matrix

In this section, an *a priori* covariance matrix S_A is constructed that is suitable for use in MSR.

First, the retrieval of common parameters shall not be considered. Let there be n single-spectrum retrieval parameters $(\mathbf{x}_i)_k$ for each of the r measurements ($i \in \{1, \cdots, r\}$ and $k \in \{1, \cdots, n\}$).

Let the number i of the spectrum be fixed. Then for the measured spectrum $\mathbf{y}_i$, $\mathbf{S}_A$ shall encode (diagonal entries) the *a priori* variance of the corresponding single-spectrum state vector $\mathbf{x}_i$ that parameterizes the corresponding forward model simulation $\mathbf{f}_i(\mathbf{x}_i)$. Also, $\mathbf{S}_A$ shall encode expected correlations between the n entries $(\mathbf{x}_i)_k$ of single-spectrum state vector $\mathbf{x}_i$ (*a priori* 'local correlations'). Such correlations can be accomplished by a non-diagonal covariance matrix for the single-spectrum retrieval problem.

Now let k be fixed and i vary. Then $\mathbf{S}_A$ shall encode expected correlations between the r parameters $(\mathbf{x}_i)_k$. These will be called *a priori* 'spatial-temporal correlations', as they describe the correlation behavior of the entries number k of the single-spectrum state vectors in space and time. These exist due to a certain continuity of the physical state of the atmosphere, caused by the inertia of matter and the drive to compensate thermodynamic disequilibria.

Finally, $\mathbf{S}_A$ shall allow to treat parameters with infinite spatial or temporal *a priori* correlation. As will be seen (Appendix A.3), such parameters are spatially or temporally constant, i.e. they are common to certain sets of spectra.

$\mathbf{S}_A$ shall be constructed as an easily parameterizable covariance matrix that can transparently encode both *a priori* local and spatial-temporal correlations as special cases. Also, it shall encode the standard deviations of all parameters of all single-spectrum state vectors, and shall allow to retrieve common parameters. It must be positive definite and ideally so already by construction. In the following, the terms '*a priori* correlation' and 'coupling' will be used synonymously.

Section 3.1 reduces the construction of the covariance matrix $\mathbf{S}_A$ to the construction of a correlation matrix $\mathbf{C}_A$. Also, for multi-spectrum problems without common parameters, the general structure of $\mathbf{C}_A$ is presented. As a first step in the construction of such a $\mathbf{C}_A$, the single-parameter problem for several spectra with arbitrary spatial-temporal data point distribution is discussed in Section 3.2. Next, single-spectrum correlation matrices for several parameters are discussed in Section 3.3. Then, for given single-parameter spatial-temporal correlation data and local single-spectrum correlation data, the Kronecker product is the key to construct correlation matrices for several parameters and several spectra in Section 3.4. The treatment of common parameters is discussed in Section 3.5.

3.1. A priori correlation matrix

A finite dimensional covariance matrix is a matrix containing the covariances between finitely many random variables. It is therefore real, symmetric, and positive semi-definite. Conversely, each finite dimensional, real, symmetric, positive semi-definite matrix is a covariance matrix [24, Theorem 2.3.1].

A positive semi-definite matrix has exclusively non-negative eigenvalues. But for a covariance matrix, a zero eigenvalue corresponds to a parameter which is exactly known (zero uncertainty) or perfectly coupled to a linear combination of other parameters. Such a parameter does not need to be retrieved or must be treated in a different way (Appendix A.3), respectively. Therefore, only real, symmetric, positive definite matrices (positive eigenvalues) will be considered in the following. This also ensures the existence of the inverse, which is needed in Section 2.

To make the construction of a *covariance* matrix $\mathbf{S}_A$ more transparent, a normalized form is defined, the *correlation* matrix $\mathbf{C}_A$ with entries

$$(\mathbf{C}_A)_{ij} := \frac{(\mathbf{S}_A)_{ij}}{\sigma_i \sigma_j}, \quad \text{or} \quad \mathbf{C}_A = \mathbf{B}^T \mathbf{S}_A \mathbf{B}, \tag{3}$$

and with the *a priori* standard deviations $\sigma_i := \sqrt{(\mathbf{S}_A)_{ii}}$ and the diagonal matrix $\mathbf{B}$ that has the entries $B_{ii} = 1/\sigma_i$. $\mathbf{C}_A$ is well defined, since the diagonal entries of a positive definite matrix are positive. All entries of $\mathbf{C}_A$ must be in the (closed) real interval $[-1, 1]$ (a consequence of Cauchy-Schwarz inequality [25, 0.6.3] when applied to vectors $\mathbf{v}_i := \sqrt{\mathbf{S}_A}\mathbf{e}_i$. Here, $\sqrt{\mathbf{S}_A}$ is defined by spectral decomposition [25, 4.1.5], and $\mathbf{e}_i$ is the i-th standard vector). According to Eq. (3), each diagonal entry equals 1 (each parameter is perfectly correlated to itself), $\mathbf{C}_A$ is symmetric, and it is positive definite since $\mathbf{B}$ is [25, 7.1.6].

For an easy construction and parameterization of $\mathbf{S}_A$, first $\mathbf{C}_A$ will be constructed, and then $\mathbf{S}_A$ is obtained by scaling $\mathbf{C}_A$ with the *a priori* standard deviations σ_i according to Eq. (3). For ease of use, not the variances σ_i^2, but 2 times the standard deviations $(2\sigma_i)$ are used as the input variables in the computer implementation, encoding the typical lengths of the expected variation intervals of the parameters.

Not considering the retrieval of common parameters (Section 3.5), in the notation of Section 2, the structure of the nr-dimensional extended parameter vector $\mathbf{X}$ (with $c = 0$) implies the structure of $\mathbf{S}_A$. The correlation matrix $\mathbf{C}_A$ inherits the same structure. As $\mathbf{X}$ consists of r sub-vectors of length n, $\mathbf{C}_A$ has to be a symmetric $nr \times nr$ matrix composed of $r \times r$ blocks of size $n \times n$. For $i, j \in \{1, \cdots, r\}$ and $k, l \in \{1, \cdots, n\}$, $(\mathbf{b}_{ij})_{kl}$, the entry (k, l) of block (i, j), encodes the *a priori* correlation between parameter k corresponding to measurement i and parameter l corresponding to measurement j. It should not be different from the coupling between the l-th parameter of measurement i and the k-th parameter of measurement j, because neither of the measurements i or j shall be distinguished from the other, i.e. $(\mathbf{b}_{ij})^T = \mathbf{b}_{ij} = (\mathbf{b}_{ji})^T$, where the latter equality follows from symmetry of $\mathbf{C}_A$. In particular, the blocks $\mathbf{b}_{ii}$ on the diagonal of $\mathbf{C}_A$ are symmetric, and they are positive definite and all identically, as they are the couplings between the parameters of the same measurement and no measurement shall be distinguished from the other. This single-spectrum correlation matrix $\mathbf{b}_{ii} =: \mathbf{h}$ will be constructed in Section 3.3.

3.2. *Single-parameter problem for several spectra*

As a first step in constructing $\mathbf{C}_A$, a correlated retrieval problem for several measurements with arbitrary distribution of distinct footprints in space and time is considered, where each of the spectra is described by only one single parameter. This could be a total cloud column factor, for instance. Therefore, all blocks $\mathbf{b}_{ij}$ (Section 3.1) of $\mathbf{C}_A$ are just real numbers, with $\mathbf{b}_{ij} =: (g_{ij}) \in \mathbb{R}^1$ and all $g_{ii} := 1$. Assume that correlation between any two measurements only depends on distance between their footprints. Here, distance is initially defined by an abstract metric $d(\cdot, \cdot)$ and will be specified later. It could be Euclidean distance in $\mathbb{R}^2$ (planetary surface treated as plane), $\mathbb{R}^3$ (as surface in $\mathbb{R}^3$), or $\mathbb{R}^4$ (including temporal separation). The 'distance matrix' $\mathbf{d} \in \mathbb{R}^r$ with entries $d_{ij} := d(\mathbf{x}_i, \mathbf{x}_j)$ is defined as the matrix of distances between the measurement footprints at locations $\mathbf{x}_i$, $i \in \{1, \cdots, r\}$ of r measurements.

For an easy and transparent parameterization and interpretation of $\mathbf{C}_A$ for the single-parameter problem, only correlation length and correlation time shall determine the strengths of the *a priori* correlations for a fixed footprint distribution. A long correlation length $\lambda \gg \pi R_{\text{Venus}}$, with the maximum spatial distance πR_{Venus} for footprints on Venus and Venus radius R_{Venus}, corresponds to a strong coupling close to 1, and similar with correlation time τ and reference time scale of four Earth days (time scale of atmospheric super-rotation). The correlation matrix for the case where any coupling is absent is defined as $\mathbf{C}_A := \mathbb{1}$, which should also be the limit of $\mathbf{C}_A$ for $\lambda \downarrow 0$ and $\tau \downarrow 0$.

In the following, it will be discussed, how a proper correlation matrix $\mathbf{C}_A$ can be yielded from $\mathbf{d}$. The difficulty lies in the requirement that for *arbitrary* distinct footprints, $\mathbf{C}_A$ will turn out positive definite by construction. The occurrence of some measurements with coinciding footprint space-time coordinates is not treated here. That case is only possible by using more than one measuring instrument.

3.2.1. *Positive definite functions*

Before a proper treatment of general spatial-temporal separations between measurement footprints will be established in Section 3.2.3, only spatial separations are considered for now.

The key to yield $\mathbf{C}_A$ from $\mathbf{d}$, is the concept of positive definite functions, see Schoenberg [26] for a collection of a series of the original papers from 1938, and Baxter [27] for the notation and definitions adopted for this work.

Let $\mathscr{H}$ be a real (possibly infinite dimensional) separable Hilbert space with norm $\|\cdot\|_2$ and $f : \mathbb{R}_+ \to \mathbb{R}$ be a function for which the quadratic form

$$\sum_{i=1}^{r}\sum_{j=1}^{r} a_i a_j f\left(\|\mathbf{x}_i - \mathbf{x}_j\|_2^2\right) \tag{4}$$

is non-negative for any $r \in \mathbb{N}^*$, any real $\mathbf{a} = (a_1, \cdots, a_r)^T \in \mathbb{R}^r$, and any points $\mathbf{x}_1, \cdots, \mathbf{x}_r \in \mathscr{H}$, then f is called *positive definite on* $\mathscr{H}$. f is positive definite on $\mathscr{H}$, if and only if f is *completely monotonic* [26], i.e. f is continuous in 0 and

$$(-1)^k f^{(k)}(x) \geq 0, \quad \forall k \in \mathbb{N} \cup \{0\}, \text{and for } 0 < x < \infty. \tag{5}$$

This can be applied to $\mathscr{H} = \mathbb{R}^n$ with Euclidean norm $\|\cdot\|_2$, to see that for $\mathbf{x}_1, \cdots, \mathbf{x}_r \in \mathscr{H}$ and for $d(\mathbf{x}_i, \mathbf{x}_j) := \|\mathbf{x}_i - \mathbf{x}_j\|_2 = d_{ij}$, the real symmetric $r \times r$ matrix $\mathbf{C}_A$ with entries $(\mathbf{C}_A)_{ij} := f\left((d_{ij})^2\right)$ is positive semi-definite, and it is even positive definite for non-constant f (note that then $f(0) > 0$) and distinct points in Euclidean $\mathbb{R}^n$ [27]. Hence, $\mathbf{C}_A$ is a valid correlation matrix, when f is non-constant and normalized such that $f(0) = 1$.

According to Miller and Samko [28, Eq. 1.13], the function defined by $f(x) := \exp\left(-\sqrt{x}\right)$ is completely monotonic. Measuring physical distances d_p in terms of a characteristic length scale λ to get rid of physical units, i.e. $x := d_p/\lambda$, it follows that for arbitrary distinct points in Euclidean $\mathbb{R}^n$, the matrix with entries $\exp\left(-d_{ij}/\lambda\right)$ is a correlation matrix.

However, this is not the most suited choice. Define the 'correlation function' $f_d(x) := f(x^2)$. It takes the (normalized) distances as arguments and reads for this choice $f_d^1(x) := \exp\left(-x\right) = \exp(-d_p/\lambda)$ with the physical distance d_p. Its non-zero derivative at $x = 0$ means that observations that are only slightly separated, are modeled to perceive fast changing correlations for varying distances. $f_d^1(x)$ is associated with a first order Gauss-Markov-process, a 'memory-less' system describing an atmosphere where only *location* is of importance [29, Section 4.3] and [30]. For many physical processes, it is more realistic to consider systems which respect certain inertial properties. This includes atmospheric physics with its inert atmospheric molecules and the fast balancing of thermodynamic disequilibria. These systems are more suitably modeled by a second order Gauss-Markov-process where *location* and *momentum* are considered [31, pp. 44–45, and example 3.9-1]. Balgovind et al. [30] derive $f_d^2(x) := (1+x)\exp(-x)$, which has a derivative of 0 at $x = 0$, i.e. observations that are only slightly separated, are modeled to perceive slowly changing correlations. f_d^2 is also given by other investigations on Earth's atmosphere [29, 31, 32]. Note that the derivative of $f_d^2(\sqrt{x})$ is $-\exp(-\sqrt{x})/2$. Since $\exp(-\sqrt{x})$ has already been established as completely monotonic, $f_d^2(\sqrt{x})$ satisfies Eq. (5) and is therefore positive definite.

But widely separated measurements have non-vanishing correlations for both f_d^1 and f_d^2. This is on the one hand not realistic, see discussion by Rood et al. [32, Section 4.4] and references therein on forecast error correlations for Earth's troposphere. On the other hand, the largest structure in MSR, the Jacobian (Eq. (A.2)), would have many small non-zero entries which are unimportant for the retrieval. To not waste computational resources, the Jacobian should possess many zero entries (see sparse matrix formulation in Appendix A.4), and the correlation function should thus have compact support, i.e. it should be zero outside of a compact (closed and bounded in Euclidean space) set. Even for the very well probed terrestrial atmosphere, the empirically substantiated correlation is quite ambiguous and there is no unique 'best' model [29, Fig. 4.5], and Venus correlations are much less known. Thus for simplicity it seems best, to choose a single appropriate correlation function that is compatible with the considerations above and will be utilized with suitable scaling for different retrieval parameters, and to test robustness against reasonable *a priori* data variations.

Therefore, a third class of correlation functions with compact support, vanishing derivative at $x = 0$, and satisfying Eq. (4) is used. Rood et al. [32] construct such a function as self-convolution $c := g * g$ of a continuous real function g with compact support. Herefore, observe that $\mu := \mathscr{F}[c] = (2\pi)^{n/2}\mathscr{F}[g]\cdot\mathscr{F}[g] \geq 0$, where $*$ denotes convolution, $\mathscr{F}$ the Fourier transform in n dimensions such that $\mathscr{F}[g](\mathbf{k}) := \int g(\mathbf{x})\exp(-i\langle\mathbf{k},\mathbf{x}\rangle)\text{đ}\mathbf{x}$, $\text{đ}\mathbf{x} = \mathrm{d}\mathbf{x}^n/(2\pi)^{n/2}$, $\langle\cdot,\cdot\rangle$ the Euclidean standard scalar product, and μ the Fourier transform of a correlation function c with $c(\mathbf{x}_i - \mathbf{x}_j) := f(\|\mathbf{x}_i - \mathbf{x}_j\|_2^2)$ as in Eq. (4). Then for any vector $\mathbf{a} = (a_1, \cdots, a_r)^T \in \mathbb{R}^r$, the quadratic form

$$\begin{aligned}\sum_{i,j=1}^{r} a_i a_j c(\mathbf{x}_i - \mathbf{x}_j) &= \sum_{i,j=1}^{r} a_i a_j \mathscr{F}^{-1}[\mu](\mathbf{x}_i - \mathbf{x}_j)\\ &= \sum_{i,j=1}^{r} a_i a_j \int \mu(\mathbf{k})\exp\left(i\langle\mathbf{k},\mathbf{x}_i - \mathbf{x}_j\rangle\right)\text{đ}\mathbf{k}\\ &= \int \Big|\sum_{i=1}^{r} a_i \exp\left(i\langle\mathbf{k},\mathbf{x}_i\rangle\right)\Big|^2 \mu(\mathbf{k})\text{đ}\mathbf{k} \geq 0,\end{aligned}$$

will be non-negative. Hence, the matrix with entries $c(\mathbf{x}_i - \mathbf{x}_j)$ will be positive semi-definite. See Reed and Simon [33, Chapters IX.1 and IX.2] and Rood et al. [32, Theorems 2.10 and 3.a.3] for a more complete discussion.

Rood et al. [32] use the real function $g(\mathbf{z}) := (1 - |\mathbf{z}|/\nu)I_\nu(|\mathbf{z}|)$ on $\mathbb{R}^3 \ni \mathbf{z}$, where $I_\nu(|\mathbf{z}|) = \begin{cases}1 & |\mathbf{z}| \leq \nu\\ 0 & |\mathbf{z}| > \nu\end{cases}$. The radial dependence of the resulting homogeneous and isotropic correlation function on $\mathbb{R}^3$ is a fifth-order piecewise rational function f_d^3 [32, Eq. 4.10]

$$f_d^3(x) = \begin{cases} -\frac{x^5}{4} + \frac{x^4}{2} + \frac{5x^3}{8} - \frac{5x^2}{3} + 1 & 0 \leq x < 1\\ \frac{x^5}{12} - \frac{x^4}{2} + \frac{5x^3}{8} + \frac{5x^2}{3} - 5x + 4 - \frac{2}{3x} & 1 \leq x < 2\\ 0 & x \geq 2\end{cases} \tag{6}$$

with $x = |\mathbf{z}|/\nu = d_p/\nu$. f_d^3 is continuous and twice continuously differentiable on $\mathbb{R}$. It is a proper correlation function by construction [32]. Thus, the non-constant $f_d^3(\sqrt{y})$ is completely monotonic as function of y, and f_d^3 is therefore a valid correlation function in any dimension. f_d^3 has derivative 0 at $x = 0$ and is explicitly of compact support since it is set to 0 for $d_p \geq 2\nu$.

The correlation length λ is defined as the physical distance where the correlation function attains the value e^{-1}. For f_d^3 this is the case at $d_p = \lambda = n_3\nu$ with the normalization factor $n_3 \approx 0.808768$. For comparison, $n_1 = 1$ for f_d^1, and $n_2 \approx 2.14691$ for f_d^2, such that $f_d^i(n_i d_p/\lambda) = e^{-1}$ at $d_p = \lambda$. Fig. 1 compares f_d^3 and f_d^2 with f_d^1. Note that f_d^3 vanishes for $d_p \geq 2\lambda/n_3 \approx 2.47\lambda$. Also note its zero derivative at $d_p = 0$ and its stronger relative weighting of the correlation of nearby measurements ($d_p < \lambda$) which better represents the inertial properties of the physical system.

3.2.2. *Spherical planetary surface*

The measurement footprints can be approximated to be located on a spherical surface $S^2 \subset \mathbb{R}^3$ (planetary surface or top of cloud deck). A positive definite function on S^2 can be induced by restricting a positive definite function on $\mathbb{R}^3$ to S^2. Therefore substitute the physical distance d_p, measured in $\mathbb{R}^3$, of points on the spherical surface, by the chordal distance $d_p = 2R\sin\frac{\vartheta}{2}$ [32, Section 2.3]. For a sphere with fixed radius R, this distance depends only on the geodetic angle of separation $\vartheta \in [0, \pi]$ between the two points on the sphere, and the induced correlation function is invariant under isotropic transformations of S^2 [32, Section 2.2]. ϑ follows directly from the Euclidean standard scalar product between any two considered vectors $\mathbf{v}_i$ and $\mathbf{v}_j$ on the unit sphere: $\cos\vartheta_{ij} = \langle\mathbf{v}_i, \mathbf{v}_j\rangle$. In terms

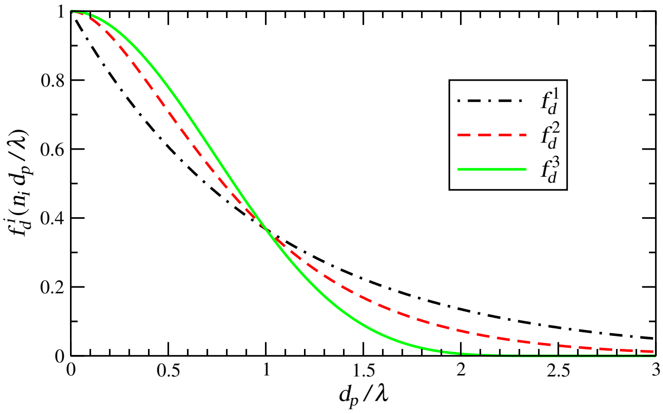

Figure 1: Comparison of the correlation functions f_d^1, f_d^2, and f_d^3.

of geographic longitude ϕ and latitude ψ, such a vector reads $\mathbf{v}_j = (\cos\phi_j \cos\psi_j, \sin\phi_j \cos\psi_j, \sin\psi_j)^T$. For a set of r *distinct* points on the sphere, f_d^3 then leads to the wanted positive definite $r \times r$ *a priori* correlation matrix with entries $(\mathbf{C}_A)_{ij} = f_d^3(n_3 \frac{2R}{\lambda} \sin \frac{\vartheta_{ij}}{2})$. $\mathbf{C}_A$ is symmetric and its diagonal entries are manifestly 1, and thus only the $r(r-1)/2$ entries below the diagonal have to be computed.

3.2.3. Spatial-temporal and other separations

By measuring temporal separation $d_{t,ij}$ between measurements i and j in terms of a correlation time τ, points on the planetary surface (or cloud top) with arbitrary distinct space-time footprints ('observation movies') and with Euclidean distance measured on the space $\mathbb{R}_s^3 \times \mathbb{R}_t^1 \cong \mathbb{R}^4$ comprising the three-dimensional spatial space $\mathbb{R}_s^3$ and the temporal space $\mathbb{R}_t^1$, lead to the correlation matrix with entries

$$(\mathbf{C}_A)_{ij} = f_d^3\left(n_3\sqrt{\left(\frac{2R}{\lambda}\sin\frac{\vartheta_{ij}}{2}\right)^2 + \left(\frac{d_{t,ij}}{\tau}\right)^2}\right). \quad (7)$$

This includes as special cases the purely spatial ($d_t = 0$) version discussed in Section 3.2.2 as well as measurements that were acquired at the same spatial coordinates but at different times, like in time series observations of a certain surface spot ('target tracking mode', $d_p = 0$). $\mathbf{C}_A$ is positive definite, because f_d^3 is a correlation function on any Euclidean $\mathbb{R}^n$, and it is associated to a second order Gauss-Markov-process, reflecting the inertial properties of the system. Measurements with a wide (compared to λ and τ) spatial-temporal separation are not correlated. A hyper-surface of constant positive correlation is an ellipsoid ($|\Delta\mathbf{x}/\lambda|^2 + |\Delta t/\tau|^2 = K^2$).

However, current knowledge is not sufficient to single out the best approximation to reality. It is also conceivable to define the space-time distance as $|d_p/\lambda| + |d_t/\tau|$ or similar. But then the argument with the Euclidean structure does not carry over, and so Eq. (7) will be adopted.

There are also parameters that experience inter-measurement *a priori* correlations, but not with respect to the measurement target coordinates (Section 4). These include detector related parameters like full width at half maximum (FWHM) of the instrumental response function of VIRTIS-M-IR which are related to the location on the detector and to time. Small changes in these coordinates are unlikely to yield abrupt changes in the parameters, and the *a priori* correlations can be defined by

$$(\mathbf{C}_A)_{ij} = f_d^3\left(n_3\sqrt{\left(\frac{|s_i - s_j|}{\sigma}\right)^2 + \left(\frac{d_{t,ij}}{\tau}\right)^2}\right), \quad (8)$$

where s_i is the sample coordinate on the detector associated with the measurement i. σ is the correlation strength in sample direction.

Finally, there are parameters that will be treated as temporally constant, like local surface emissivity when it is retrieved as parameter common to measurements covering the same surface spot, or the deep atmosphere temperature profile as parameter common to measurements associated to a fixed latitude. In these cases, the purely spatial correlation matrix discussed in Section 3.2.2 is used, corresponding to Eq. (7) with $\tau = \infty$ (compare Appendix A.3).

For easier use, all correlation types discussed in this section will be referred to as spatial-temporal correlations, in contrast to local correlations.

3.3. Single-spectrum problem for several parameters

As the next step in constructing $\mathbf{C}_A$, this section discusses, how valid single-spectrum *a priori* correlation matrices $\mathbf{h}$ with entries h_{ij} (Section 3.1) for several parameters can be constructed.

h_{ij} encodes the strength of the coupling between single-spectrum parameters x_i and x_j. Let c_i ($|c_i| < 1$) denote the 'nearest neighbor coupling' between the 'neighboring' parameters x_i and x_{i+1}. $c_i = 0$ translates to vanishing correlation and $c_i < 0$ to anti-correlation. Define coupling between parameters x_i and x_{i+2} as a function of the nearest neighbor couplings c_i and c_{i+1}. The modulus of 'next-to-nearest neighbor coupling' shall be smaller than either of the nearest neighbor couplings moduli. If one of the nearest neighbor couplings is zero, the next-to-nearest neighbor coupling shall be zero. If one of the nearest neighbor couplings is positive while the other is negative, then the next-to-nearest neighbor coupling shall be negative. The simplest way to implement these requirements is, to define the next-to-nearest neighbor coupling as the product of the nearest neighbor couplings. Analogously, this can be done with the coupling to the third neighbors, and so on. Therefore $\mathbf{h}$ shall be defined as

$$\mathbf{h} = \begin{pmatrix} 1 & c_1 & \cdots & c_1 c_2 \cdots c_{n-1} \\ c_1 & 1 & \cdots & c_2 \cdots c_{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ c_1 c_2 \cdots c_{n-1} & c_2 \cdots c_{n-1} & \cdots & 1 \end{pmatrix}, \quad (9)$$

an easy-to-implement matrix, which is parameterized by just the $n-1$ nearest neighbor couplings of n parameters.

To show that $\mathbf{h}$ is a valid correlation matrix, recall that any $n \times n$ matrix $\widetilde{\mathbf{h}}$ with entries $\widetilde{h}_{ij} := |h_{ij}| =: \exp\left(-\|\mathbf{x}_i - \mathbf{x}_j\|_2/\lambda\right)$ is a correlation matrix for distinct vectors $\mathbf{x}_i$ and $\mathbf{x}_j \in \mathbb{R}^m$ and $0 < \lambda \in \mathbb{R}$ (f_d^1 in Section 3.2.1, $\lambda := 1$, $m := 1$, the n vectors $\mathbf{x}_i$ are points on the real line with nearest-neighbor distances $-\log|c_k|$). Define $\mathbf{V}_l$ as a diagonal $n \times n$ matrix with the first l entries on the diagonal -1 and the remaining 1. $\mathbf{V}_l^T\mathbf{h}\mathbf{V}_l$ changes signs of h_{ij} in the blocks ($i \le l$, $j > l$) and ($i > l$, $j \le l$). Then $\mathbf{V}_l^T\widetilde{\mathbf{h}}\mathbf{V}_l$ is positive definite [25, 7.1.6]. Since all of its diagonal entries are still 1, it is a correlation matrix. This can be reformulated by allowing $-1 < c_l < 0$ in addition to $0 \le c_k < 1$, $k \in \{1, \cdots, n-1\}$ for $\mathbf{h}$ in Eq. (9) to still qualify as correlation matrix. Similarly, this independently follows for all $l \in \{1, \cdots, n-1\}$. Hence, $\mathbf{h}$ in Eq. (9) is a correlation matrix for $|c_k| < 1$, $k \in \{1, \cdots, n-1\}$.

$\mathbf{h}$ can describe local *a priori* correlations for a single spectrum, like between column factors (minor gases, cloud modes) that may have correlations or anti-correlations. It can also describe inter-level correlations of atmospheric profiles ($c_k := \exp(-|z_k - z_{k+1}|/\lambda)$ with z_k the altitude of level k and λ the inter-level correlation length, an obvious generalization of Rodgers [18, Eq. 2.83]). Also, *a priori* correlations between common parameters (Sections 3.5 and 4.3) can be modeled this way. Note that, properly taking into account the signs (see $\mathbf{V}_l^T(\cdot)\mathbf{V}_l$), other correlation functions than f_d^1 from Section 3.2.1 still lead to proper single-spectrum correlation matrices.

3.4. Full covariance matrix for correlated retrievals

This section discusses the construction of the full covariance matrix for several parameters and several spectra for given single-parameter spatial-temporal correlation data (Section 3.2) and local single-spectrum correlation data (Section 3.3). While there seems to be no simple solution for arbitrary local couplings, it is possible to provide correlation matrices sufficiently general for VIRTIS data retrieval.

First, a correlation matrix is constructed for a group of several locally coupled parameters that experience identical fixed correlation length and time. For example, temperature altitude profiles have local inter-level-correlations and may be described by the same horizontal spatial-temporal correlation data. The Kronecker product [34, Section 4.2] is well suited to treat this case.

The Kronecker product $\mathbf{A} \otimes \mathbf{B}$ [34, Definition 4.2.1] of an $m \times n$-matrix $\mathbf{A}$ with entries A_{ij} and a $p \times q$-matrix $\mathbf{B}$ is defined to be the $mp \times nq$-dimensional block matrix

$$\mathbf{A} \otimes \mathbf{B} := \begin{pmatrix} A_{11}\mathbf{B} & \cdots & A_{1n}\mathbf{B} \\ \vdots & \ddots & \vdots \\ A_{m1}\mathbf{B} & \cdots & A_{mn}\mathbf{B} \end{pmatrix} \tag{10}$$

and is not commutative in general. According to Horn and Johnson [34, 4.2.4 and 4.2.13], the result of the Kronecker product of symmetric positive definite matrices is symmetric positive definite.

Let there be G groups of parameters with group-specific correlation length λ_g and time τ_g for each group $g \in \{1, \cdots, G\}$. Then let $\mathbf{h}_g$ be the positive definite $n_g \times n_g$ matrix of the single-spectrum correlations between the n_g parameters of group g (constructed as $\mathbf{h}$ in Section 3.3). The total count n of single-spectrum parameters over all groups amounts to $n = \sum_{g=1}^{G} n_g$. The spatial-temporal $r \times r$ correlation matrix for a single parameter from this group g shall be denoted by $\boldsymbol{\varrho}_g$. By using λ_g, τ_g, and f_d^3, it is computed according to Section 3.2.3 from the space-time coordinates of the r measurement footprints. It is the same for all parameters of group g.

The matrix $\boldsymbol{\varrho}_g \otimes \mathbf{h}_g$ is then the correlation matrix $\mathbf{C}_A^g$ for group g (note that $(\boldsymbol{\varrho}_g)_{ii} = 1$).

$$\boldsymbol{\varrho}_g \otimes \mathbf{h}_g = \begin{pmatrix} \mathbf{h}_g & (\boldsymbol{\varrho}_g)_{12}\mathbf{h}_g & \cdots & (\boldsymbol{\varrho}_g)_{1r}\mathbf{h}_g \\ (\boldsymbol{\varrho}_g)_{12}\mathbf{h}_g & \mathbf{h}_g & \cdots & (\boldsymbol{\varrho}_g)_{2r}\mathbf{h}_g \\ \vdots & \vdots & \ddots & \vdots \\ (\boldsymbol{\varrho}_g)_{1r}\mathbf{h}_g & (\boldsymbol{\varrho}_g)_{2r}\mathbf{h}_g & \cdots & (\boldsymbol{\varrho}_g)_{rr}\mathbf{h}_g \end{pmatrix} =: \mathbf{C}_A^g \tag{11}$$

$\mathbf{C}_A^g$ satisfies all necessary conditions for the correlations of parameters belonging to group g. It includes spatial-temporal coupling for single parameters as well as local coupling for single spectra as special cases. $\mathbf{C}_A^g$ is symmetric positive definite with all diagonal entries 1, and the single blocks are symmetric, compare Section 3.1.

Next, groups of parameters with different spatial-temporal correlations are combined. For example, temperature vs. cloud altitude profiles with different inter-level-correlations may have different horizontal spatial-temporal correlations.

Let $(\mathbf{x}_i)_k$ denote the k-th single-spectrum parameter of spectrum i. The concatenated list $\mathbf{X}$ (as in Section 2) of the n parameters for each of the r measurements

$$\mathbf{X} = \Big(\underbrace{(\mathbf{x}_1)_1, \cdots, (\mathbf{x}_1)_n}_{\text{measurement 1}}, \cdots, \underbrace{(\mathbf{x}_r)_1, \cdots, (\mathbf{x}_r)_n}_{\text{measurement r}} \Big)^T \tag{12}$$

is a permutation of the list $\overline{\mathbf{X}}$ of the r measurements for each of the n single-spectrum parameters

$$\overline{\mathbf{X}} = \Big(\underbrace{(\mathbf{x}_1)_1, \cdots, (\mathbf{x}_r)_1}_{\text{parameter 1}}, \cdots, \underbrace{(\mathbf{x}_1)_n, \cdots, (\mathbf{x}_r)_n}_{\text{parameter n}} \Big)^T. \tag{13}$$

The associated permutation $\mathbf{\Pi}_n^r$ with $\overline{\mathbf{X}} = \mathbf{\Pi}_n^r\mathbf{X}$ acts for all $i \in \{1, \cdots, r\}$ and for all $k \in \{1, \cdots, n\}$ as

$$(\mathbf{X})_{n(i-1)+k} = (\mathbf{x}_i)_k = (\overline{\mathbf{X}})_{r(k-1)+i} =: (\overline{\mathbf{x}}_k)_i.$$

The permutation $\mathbf{\Pi}_{n_g}^r$ can be used to permute $\mathbf{C}_A^g$ to obtain $\mathbf{\Pi}_{n_g}^r(\boldsymbol{\varrho}_g \otimes \mathbf{h}_g)(\mathbf{\Pi}_{n_g}^r)^T =: \overline{\mathbf{C}}_A^g$.

$$\overline{\mathbf{C}}_A^g = \begin{pmatrix} \boldsymbol{\varrho}_g & (\mathbf{h}_g)_{12}\boldsymbol{\varrho}_g & \cdots & (\mathbf{h}_g)_{1n_g}\boldsymbol{\varrho}_g \\ (\mathbf{h}_g)_{12}\boldsymbol{\varrho}_g & \boldsymbol{\varrho}_g & \cdots & (\mathbf{h}_g)_{2n_g}\boldsymbol{\varrho}_g \\ \vdots & \vdots & \ddots & \vdots \\ (\mathbf{h}_g)_{1n_g}\boldsymbol{\varrho}_g & (\mathbf{h}_g)_{2n_g}\boldsymbol{\varrho}_g & \cdots & (\mathbf{h}_g)_{n_gn_g}\boldsymbol{\varrho}_g \end{pmatrix} \tag{14}$$

This is the same as $\mathbf{h}_g \otimes \boldsymbol{\varrho}_g$ and hence positive definite.

Now a block diagonal matrix $\overline{\mathbf{C}}_A$ can be formed with all $\overline{\mathbf{C}}_A^g$ as the G positive definite blocks on the diagonal.

$$\overline{\mathbf{C}}_A := \begin{pmatrix} \mathbf{h}_1 \otimes \boldsymbol{\varrho}_1 & 0 & \cdots & 0 \\ 0 & \mathbf{h}_2 \otimes \boldsymbol{\varrho}_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \mathbf{h}_G \otimes \boldsymbol{\varrho}_G \end{pmatrix} \tag{15}$$

As a direct sum of positive definite matrices, $\overline{\mathbf{C}}_A$ is positive definite (for a correspondingly partitioned $\mathbf{v} = (\mathbf{v}_1, \cdots, \mathbf{v}_G)$ $\neq 0$, the quadratic form $\langle \mathbf{v}, \overline{\mathbf{C}}_A\mathbf{v} \rangle = \langle \mathbf{v}_1, \overline{\mathbf{C}}_A^1\mathbf{v}_1 \rangle + \cdots +$

$\langle \mathbf{v}_G, \overline{\mathbf{C}}_A^G \mathbf{v}_G \rangle$ is positive). In general, $\overline{\mathbf{C}}_A$ can not be represented as Kronecker product.

Each of the G blocks on the diagonal corresponds to the permuted correlation matrix of one of the G parameter groups, encoding local as well as spatial-temporal correlations. The correlation lengths and times are allowed to differ from block to block but must be constant within each block. Parameters belonging to a certain group may locally be arbitrarily coupled among each other. However, no local coupling between the different groups can be implemented this way.

In order to recover the sorting associated to $\mathbf{X}$ (Eq. (12)), the inverse permutation has to be applied to $\overline{\mathbf{C}}_A$ to obtain $\mathbf{C}_A := (\mathbf{\Pi}_n^r)^T \overline{\mathbf{C}}_A \mathbf{\Pi}_n^r$. Although the dimensions $n_g r \times n_g r$ of $\overline{\mathbf{C}}_A^g$ depend on g, this is the correct permutation, since it is associated to r measurements and n parameters, which is the underlying structure of $\overline{\mathbf{C}}_A$. As $\overline{\mathbf{C}}_A$ is positive definite and a permutation matrix has full rank, $\mathbf{C}_A$ is positive definite [25, 7.1.6]. It is symmetric with all diagonal entries 1, because $\overline{\mathbf{C}}_A$ is.

The full covariance matrix $\mathbf{S}_A$ is computed according to Eq. (3), by scaling $\mathbf{C}_A$ with the single-spectrum standard deviations σ_k of the single-spectrum parameters k.

$$(\mathbf{S}_A)_{n(i-1)+k\,,\,n(j-1)+l} = \sigma_k \sigma_l \left(\mathbf{C}_A\right)_{n(i-1)+k\,,\,n(j-1)+l} \tag{16}$$

This is done for all measurements $i, j \in \{1, \cdots, r\}$ and for all parameters $k, l \in \{1, \cdots, n\}$.

The $nr \times nr$ matrix $\mathbf{S}_A$ satisfies all requirements, except for a possible inter-group coupling. As can explicitly be shown for the case of two measurements, each described by two locally coupled single-spectrum parameters that have different *a priori* correlation lengths, $\mathbf{C}_A$ can fail to be positive definite for too strong inter-group coupling. The underlying reason is that strong local coupling between the parameters forces the retrieved values to have strongly correlated spatial-temporal behavior, which is not consistent when strongly differing correlation lengths are chosen. This manifests itself by the failure to construct a proper correlation matrix, unless the local coupling is relaxed or the spatial-temporal correlations are equalized. This explains, why all correlation data can be chosen freely and independently as long as inter-group coupling is disregarded.

Thus, in order to allow arbitrary space-time distributions of the measurement footprints for problems with many measurements and parameters, either arbitrary spatial-temporal correlation data can be set for the different parameters, or arbitrary local coupling may be required. For the first case, these parameters have to be assigned to different parameter groups and no local coupling is allowed between them. For the second case, these parameters have to be assigned to the same parameter group and spatial-temporal correlation data must coincide.

In general, however, a correlation matrix is not unique for given local and spatial-temporal correlation data. For instance, the distance matrix approach (Section 3.2) could be applied to a suitable point distribution on the Cartesian product space of the spatial-temporal and the local parameter dimensions. But then, local and spatial-temporal correlations for the full multi-spectrum problem would not be independent anymore in their impacts, and the advantages of the Kronecker product construction would get lost: It is ideally suited to the computation of the scaled residual (Appendix A.2) and the sparse matrix formulation of the retrieval algorithm (Appendix A.4). It also enables a clean derivation of the retrieval of common parameter vectors (Appendix A.3). In addition, inter-group coupling is not necessary for Venus retrieval problems, as is demonstrated by Section 4.2, which presents a basic categorization of the relevant parameter groups herefore.

3.5. Retrieval of common parameters

This section discusses the *a priori* covariance matrix for retrieval problems involving parameters with perfect spatial or temporal coupling (spatial or temporal *a priori* correlation equal to 1, corresponding to infinite correlation length or time, leading to parameters being affinely linear functions of each other) and identic single-spectrum *a priori* data (affinely linear functions are then the identity function, Appendix A.3). Such parameters can not spatially or temporally vary for a certain set of r considered measurements. To avoid a degenerated *a priori* covariance matrix, such a parameter will not be treated as r individual parameters that are perfectly coupled, but as a single parameter that is common to the involved measurements. Also, this helps to save computer resources. To verbally distinguish common parameters from spatially-temporally varying parameters, the latter are cited as 'local parameters'. This should not be confused with 'local coupling between parameters' as opposed to 'spatial-temporal coupling'.

The retrieved value of a common parameter can be defined as the limit of the retrieved values of the corresponding r local parameters for ever stronger spatial or temporal coupling. In Appendix A.3 it is shown that computing this limit is equivalent to retrieving a common parameter in the sense of Section 2 while uniquely defining the corresponding *a priori* covariance matrix. A statistical $\sqrt{r}$-like relative weighting factor between common and local parameters is conceivable that reflects the influence of a common parameter on r spectra. Appendix A.3 clarifies that the weighting is exactly equal.

The *a priori* covariance matrix $\mathbf{S}_A$ follows as block diagonal with $\mathbf{S}_C$ and $\mathbf{S}_L$ as the blocks on its diagonal. $\mathbf{S}_C$ encodes the *a priori* correlations between the various common parameters and their *a priori* standard deviations. It can be constructed by defining a suitable correlation matrix $\mathbf{C}_C$ by nearest-neighbor-coupling (Section 3.3) or by a distance matrix (Section 3.2.3), and by scaling $\mathbf{C}_C$ with the *a priori* standard deviations σ_k of the common parameters (Eq. (3)). $\mathbf{S}_L$ can be constructed according to Section 3.4 and encodes the *a priori* local and spatial-temporal correlations as well as standard deviations of the local parameters. $\mathbf{S}_A$ is a proper covariance matrix when $\mathbf{S}_C$ and $\mathbf{S}_L$ are.

For VIRTIS-M-IR measurements of Venus, some relevant common parameters along with a suitable $\mathbf{C}_C$ are discussed in Section 4.3.

4. Parameters for Venus retrieval problems

This section presents a basic categorization of relevant parameters for Venus retrieval problems. Compare also

[11, 12] for a discussion of these parameters in context of the radiative transfer forward model. Section 4.1 summarizes the most important properties of the forward model. The categorization of the parameters also demonstrates that it suffices here to construct correlation matrices without considering inter-group coupling. Local parameters are treated by an *a priori* correlation matrix according to Section 3.4, common parameters according to Section 3.5.

4.1. Forward model

A radiative transfer forward model is utilized to simulate the observable radiances. It is a plane-parallel, non-LTE, line-by-line code taking into account thermal emissions by surface and atmosphere, and absorption and multiple scattering by gases and clouds. It is similar to the forward model described by Haus and Arnold [12], but the underlying radiative transfer equation solver DISORT [35] is replaced by LIDORT [22, 36]. This way, the forward model is capable of providing analytic derivatives of the simulated radiances with respect to a number of atmospheric, surface, and instrumental parameters. Jacobians with respect to the remaining parameters (mainly temperature variables) can be evaluated perturbatively (slower and possibly affected by numerical noise). To increase numerical efficiency, uninteresting wavelength ranges can be blacked out automatically (initial radiances or Jacobians below certain thresholds) or manually.

From the VIRTIS-M-IR spectral range (1.0–5.2 µm), only 1.0–2.5 µm shall be utilized for this study. Venus' nightside emissions in this range mainly originate from altitudes below 40 km [12, Fig. 4] where temperature is quite stable with time, and they are thus nearly unaffected by the strong mesospheric temperature variations above 59 km [17]. Also, details of the cloud altitude distribution have almost no impact here, since the main cloud deck (≥48 km [37]) resides above the line forming altitude region. In contrast, spectral signatures longward of 3 µm are strongly influenced by variations of temperature and cloud altitude distributions above 48 km.

Temperature and pressure altitude profiles are taken from the Venus International Reference Atmosphere (VIRA [3, 4]) at the equator (midnight). The surface is assumed to be in thermodynamic equilibrium with the bottom of the atmosphere, and therefore, the surface temperature equals the VIRA temperature at the respective surface elevation. Surface emissivity must lie in the interval $[0, 1]$.

The main constituent of Venus' atmosphere is CO_2 (96.5% by volume). Considered minor gaseous constituents are H_2O, CO, SO_2, OCS, HCl, and HF. Altitude profiles of their volume mixing ratios are given by Haus and Arnold [12] and are based on the profiles by Pollack et al. [5]. Quasi-monochromatic absorption cross sections due to their allowed molecular transitions are computed from the spectral line databases CDSD (CO_2 [38]), HITEMP (CO_2 [5], CO, H_2O isotopes 1–3 [39]; to be in line with [12], the more recent HITEMP2010 [40] is not yet considered here), and HITRAN08 (H_2O isotopes 4–6, SO_2, OCS, HCl, HF [41]) by using spectral line shapes listed by Haus and Arnold [12]. Molecular Rayleigh scattering is treated according to Hansen and Travis [42]. Non-LTE O_2 emissions ('O_2 nightglow') at 1.27 µm from an altitude region around 100 km [43] are not considered, and therefore, the 1.28 µm window will be blacked out in practice due to its contamination by O_2 nightglow.

The high pressure and high temperature environment of Venus' deep atmosphere makes it difficult to characterize the absorption properties of its main constituent CO_2. Neither the line shapes of the allowed transitions, nor other effects contributing to the absorption cross-section (continuum, collisional induced absorption, line mixing) are sufficiently well constrained through laboratory or theory in order to satisfactorily reproduce observed spectra in the infrared. Also, the line data bases utilized for computing the absorption cross-sections of the allowed transitions are not perfect [12, 44]. They are based on theoretical models and numerical computations, and not on laboratory measurements [38]. Good knowledge of the CO_2 opacity is important for a reliable retrieval of parameters like surface emissivity. Wavelength dependent corrections to the CO_2 opacity as given by the allowed transitions, are in the following shortly referred to as 'continuum'. The continuum depends on the utilized line databases and line shapes but is independent of the measurement. For this study, continuum is treated as spectrally constant throughout the range of an atmospheric transparency window, but it can depend on the window. These window-specific scalars for the spectral windows at 1.02, 1.10, 1.18, 1.28, 1.31, 1.74, 2.3 µm are set to 0, 2, 0.35, 3, 4, 15, 120 in units of 10^{-29} cm^2, respectively. These values are inspired by preliminary results from application of MSR to actual VIRTIS-M-IR nightside spectra.

The clouds of Venus are modeled to comprise the four modes 1, 2, 2', and 3. Each mode consists of spherical droplets of 75% sulfuric acid (refractive indices taken from [45, 46]). Cloud particle radii are log-normally distributed with modal radii of 0.3, 1.0, 1.4, 3.65 µm and unitless dispersions of 1.56, 1.29, 1.23, 1.28 for the four modes, respectively [5]. Mode specific initial altitude profiles of particle number densities are taken from [47]. Actual cloud modal abundances are defined by 'cloud mode factors' that scale the number densities of these four initial altitude profiles. The mode factors may strongly vary, and no detailed *a priori* knowledge is available. Wavelength dependent scattering and absorption properties of the clouds are computed by using Mie theory [48].

Fig. 2 (offset 0.0) displays a typical synthetic spectrum ('reference spectrum'). The Jacobians of several retrieval parameters that shall be considered in the following and are relevant for actual measurements, are shown with various offsets. The values that led to Fig. 2 are: FWHM of instrumental response function 17 nm, surface emissivity 0.65, surface elevation 0 km, no O_2 nightglow, cloud mode factors all set to 1, nadir-looking observational geometry, no noise. The figure illustrates that surface emissivity is observable in the spectral windows at 1.02, 1.10, and 1.18 µm ('surface windows'), continua affect all considered windows, the abundances of the different cloud modes affect the spectrum in a very similar way and in all windows, and the impact of the FWHM is quite distinct. Minor gases do not affect the short-wavelength flank of the 2.3 µm window (2.15–2.30 µm), but strongly affect the long-wavelength flank which thus shall not be used for retrievals of cloud parameters. Even so, minor gas variations shall not be considered in this study, despite their (moderate) impact on the 1.74 µm peak and the range

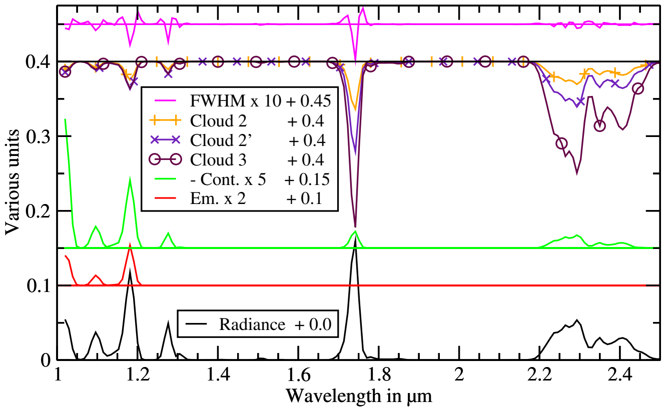

Figure 2: [From bottom to top] Offset 0: synthetic radiance spectrum in W/(m^2 sr µm), various positive offsets: scaled Jacobians with respect to surface emissivity ('Em.', unit W/(m^2 sr µm)), continuum ('Cont.', between 2.1 and 2.5 µm scaled by factor of 10 relative to remaining range for better representation, unit W/(m^2 sr µm) $\cdot$ 10^{29} cm^{-2}), cloud mode column factors ('Cloud', unit W/(m^2 sr µm)), and FWHM of VIRTIS-M-IR instrumental response function ('FWHM', unit W/(m^2 sr µm nm)).

1.10–1.18 µm. As cloud parameters should not be retrieved from the surface windows, and the 1.28 µm window shall be blacked out due to its O_2 nightglow contamination, the range 1.295–2.300 µm is left to retrieve cloud parameters from.

4.2. Local parameters

First, the atmospheric parameters may be divided into the cloud group, the minor gases group, the atmospheric temperature group, and O_2 nightglow.

Horizontally, the cloud opacity exhibits a rather short correlation length of the order of several hundred kilometers and correlation times of a few hours. This can be checked by analyzing the auto-correlation function of nightside radiance observation movies in the short wavelength flank of the 2.3 µm window as proxy. This correlation data will be carried over to the column densities of the individual cloud modes 1, 2, 2', and 3.

Minor gas column density variations seem to be better represented by longer correlation lengths of the order of more than thousand kilometers. This can be estimated for CO by observing the spatial variation in the results by Tsang et al. [49] and is also applied to the other minor gases as first estimate. Similarly, the correlation time is likely larger than the cloud correlation time. The spatial-temporal scales of atmospheric super-rotation and convection can serve as a motivation herefore.

For cloud or minor gases altitude density profiles, it is reasonable to assign the spatial-temporal properties of the column densities to the horizontal variability of the profiles as well and to treat the vertical variability as locally coupled parameters.

Thus, the cloud parameter group comprises the total cloud column factor, the column factors of the individual cloud modes, and the parameters describing the corresponding cloud mode profiles. Similarly, the minor gases parameter group includes the column factors and profiles of all minor gases. *A priori* couplings between gases and clouds are neglected, but the parameters of either group can be coupled to other parameters within their group. This includes the vertical variability of the respective altitude profiles. Also, for instance, cloud modes 2' and 3 may well be slightly anti-correlated, as might be OCS and SO_2, but this is not considered in practice.

Atmospheric temperatures in the mesosphere are treated to have correlation properties similar to the minor gases and thus could well be assigned to the minor gases group. But couplings with the minor gases are not expected, and so they will be regarded as a distinct group. O_2 nightglow is also treated as one separate parameter group.

Auxiliary instrumental parameters like the FWHM of the instrumental response function and the slope and intercept of the band-to-wavelength mapping are not sufficiently well predictable by the calibration pipeline at the moment and thus have to be retrieved as auxiliary parameters needed to adequately simulate the observed spectra [11, Section 4.4]. They are not coupled to atmospheric or surface parameters, but may be coupled among themselves. Therefore, they are assigned to one separate parameter group and treated according to Eq. (8).

All these different groups can safely be treated as independent from each other, and inter-group coupling is not necessary to describe the parameter correlations as long as *a priori* knowledge is as limited as presently.

4.3. Common parameters

As discussed in Section 4.1, the CO_2 continuum is not sufficiently well known. However, the observed VIRTIS-M-IR spectra themselves can be regarded as measurements thereof, but now with the locally varying parameters as interfering factors. Although for this study modeled as window-specific and spectrally constant throughout the range of an atmospheric transparency window, the continuum may freely vary in wavelength direction and is treated as a parameter vector. It is clearly common to all measurements of Venus' atmosphere and will be retrieved as parameter that is common to a selection of as many as possible spectra under as many as possible environmental and observational conditions. This way, it shall be ensured that it is compatible with all measurements and as reliable as possible. The continuum has to be determined only once and will thereafter be used as fixed value for subsequent retrievals of atmospheric and surface parameters. It can be regularized by only allowing for limited variation in wavelength direction. This can be achieved by a nearest-neighbor coupling in wavelength direction (Section 3.2) or also by a distance matrix with respect to distances in wavelength direction (Section 3.3), but both ways are purely heuristically, since not much is known about the wavelength dependence of the continuum. First results that are based on MSR have been presented by Kappel et al. [11].

Surface properties should be quite unrelated to variations in the atmosphere, except for a possible coupling of surface temperature and atmospheric temperature at the surface, which will be neglected. When the occurrence of volcanic activity (as observable by VIRTIS-M-IR in the nightside NIR surface windows at 1.02, 1.10, and 1.18 µm) is neglected [50], the spectral surface emissivity can be regarded as common to all measurements that repeatedly cover the same target bin on the surface. Or in other words, the entire surface emissivity map of the planet is common to all measurements targeting the planet. Note that a measurement is only sensitive (i.e. non-zero entry in the Jacobian) to surface emissivity at surface bins that are actually covered by the measurement. Retrieving surface emissivity as parameter common to measurements repeatedly covering a target, yields a fully consistent parameter set describing all considered measurements [21]. In contrast, the corresponding single-spectrum retrievals yield different emissivity values for each of these measurements in most cases. This implies an inconsistent parameter set describing the full set of utilized measurements, implicitly also allowing for inconsistent atmospheric parameter values. The emissivities at nearby spots are possibly correlated. This purely spatial situation can be treated by using the correlation matrix from Eq. (7) with $\tau = \infty$. The typical correlation length is 100 km, i.e. the expected surface resolution as it is limited by atmospheric blurring [51].

The deep atmospheric temperature field is not yet sufficiently well known for a reliable surface emissivity retrieval. The corresponding Jacobians are quite similar to cloud Jacobians, and disentanglement is not feasible for single-spectrum retrievals. A general circulation model [52, personal communication] and measurements [1, 17] are compatible with a temporally rather constant tropospheric temperature field that is altitude and latitude dependent, probably a consequence of high thermal inertia and thermodynamic stable layering in the deep atmosphere. This is not the case for the highly variable mesosphere. Thus, a deep atmospheric temperature altitude profile is essentially a parameter vector common to measurements covering a given latitude. The latitudinal dependencies of the temperature profiles can be coupled by using a correlation matrix similar to that for the surface emissivity map, but spatial separation is only measured in latitude direction, i.e. $(\mathbf{C}_C)_{ij} = f_d^3\,(n_3|\vartheta_i - \vartheta_j|/\Theta)$, compare Section 3.2.3. The latitude corresponding to measurement i is hereby ϑ_i, and the correlation strength Θ in latitude direction can be set to 45 °.

Finally, while it is possible to couple the different mentioned common parameter types, this seems not to be useful. Any coupling to local parameters is also not needed.

5. Examples and discussion

Since MSR determines locations of local minima of the cost function F_c, a non-negative real function on a possibly high dimensional parameter space (Eq. (1), also applicable for single-spectrum retrievals by setting inter-measurement couplings to zero), it is easy to find local subsidiary minima, which are potentially far from the global minimum, especially in presence of measurement noise. This section discusses by two examples, how this situation can be improved by using MSR (Section 2, with *a priori* covariance matrix $\mathbf{S}_A$ from Section 3) by not only incorporating information on expected parameter values, but also expected relations between parameters. The more actually available *a priori* knowledge is utilized, the better certain solutions being incompatible with the *a priori* data can be ruled out from the outset. Some resulting improvements in the data analysis of Venus nightside spectra acquired by VIRTIS-M-IR have already been published by [11, Section 5.1]. The retrieval algorithm has been sketched there only shortly, and the full description has been announced to be published in a subsequent (the present) paper.

To test MSR, a set of synthetic VIRTIS-M-IR radiance spectra of Venus' nightside emissions in the range 1.0–2.5 µm is generated by using the radiative transfer forward model. The utilized 'true' atmospheric, surface, and instrumental parameters underlying these spectra are thus exactly known by definition. Gaussian noise with a certain standard deviation σ is added to the simulated spectra to emulate the loss of the spectra's information content caused by random measurement imperfections. Systematic measurement and calibration errors will be considered in a subsequent paper (set parameters that are not retrieved to values different from their assumed values, distort shapes of the radiance peaks). By using different regularization schemes, the relevant parameters are then retrieved from the synthetic spectra and compared to their 'true' values.

Example (A) studies a swath of 30 concurrent synthetic spectra covering the equator at longitudes from 1–30 °E. Fig. 3 depicts the 'true' cloud column factors 2, 2', and 3. The remaining parameters are as for the reference spectrum (Fig. 2), and Gaussian noise with the unrealistically small double standard deviation $2\sigma = 2 \cdot 10^{-4}$ W/(m^2 sr µm) is added to the synthetic spectra. MSR is used to retrieve the cloud mode factors from the radiance spectra for three different regularization models from the spectral range 1.295–2.300 µm (1.31, 1.74, and 2.3 µm peaks). *A priori* mean values and double standard deviations are set to 1.0 and 2.0, respectively, to allow for a sufficiently wide range the retrieved cloud mode factors may vary in. Note that 1 °longitude at the equator corresponds to about 107 km referred to the cloud top level.

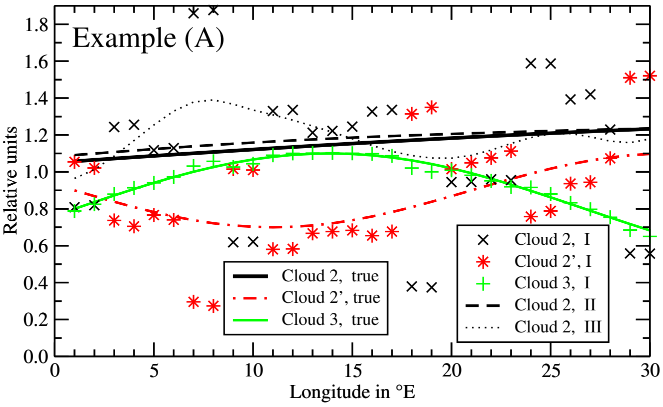

Figure 3: Comparison of 'true' cloud mode column factors (left key) of longitudinal synthetic measurement swath with corresponding retrieved values (right key) for different models of *a priori* correlation lengths λ_i for cloud mode i. I: $\lambda_2 = 0\,\mathrm{km} = \lambda_{2'} = \lambda_3$, II: $\lambda_2 = 20000\,\mathrm{km}$, $\lambda_{2'} = \lambda_3 = 500\,\mathrm{km}$, III: $\lambda_2 = 2000\,\mathrm{km} = \lambda_{2'} = \lambda_3$.

Model I sets all *a priori* correlation lengths to zero and corresponds to single-spectrum retrievals. While cloud mode 3 abundance can be retrieved quite reliably, modes 2 and 2' are difficult to disentangle and strongly deviate from their true values. This is a consequence of the smallness of one of the three available peaks (1.31 µm-peak radiance

$\approx 10^{-2}\,\mathrm{W/(m^2\ sr\ \mu m)}$) that are used to determine the three unknown cloud mode factors, the presence of noise, and the similarity of the Jacobians especially of the cloud mode factors 2 and 2'. Note that for zero noise, the true values can be retrieved exactly.

Model II sets *a priori* correlation lengths that approximate the true parameter distributions. The dependence on correlation length modifications by factors of 0.5 or 2 has been verified to be relatively small. Mode 3 factors are not depicted as they almost exactly coincide with their true values. Mode 2' factors are not shown because deviations from their true values are opposite and of a similar magnitude (but smaller) as deviations for mode 2 (compare model I results in the figure). Retrieved mode 2 factors deviate less than 10% from their true values.

Model III sets for all cloud modes identical *a priori* correlation lengths that approximate the geometric mean of the three correlation lengths from model II. Retrieved mode 3 factors match the true values well. Modes 2 and 3 can not be disentangled (only mode 2 shown). Setting all *a priori* correlation lengths to 20000 km, or to 500 km, respectively, leads to worse results.

Averaged least-squares norms of residuals between synthetic and fitted radiances are least for model II and largest for model I. Within model III, the geometric-mean-case leads to the smallest residuals.

In Example (A), as in many real-world atmospheric remote sensing problems, measurements are not isolated soundings in space and time, but each measurement is accompanied by adjacent measurements. If they are nevertheless treated as independent from other soundings, then spatial or temporal continuity in the measurements may not translate to a certain expected continuity in retrieved parameters like cloud column densities. This is due to the ill-posed nature of the retrieval problem and the existence of subsidiary minima of the cost function. Actually, contiguous real-world measurements are unlikely to originate from completely unrelated state vectors, since the physical state of the atmosphere should obey a certain continuity. This follows from the inertia of matter and the drive to compensate thermodynamic disequilibria and results in a certain continuity of the observable radiance. Therefore, it can help to reduce the effective size of the parameter space by not only incorporating *a priori* mean values and standard deviations as usual, but by also taking *a priori* spatial-temporal correlations into account. This decreases the number of potential solutions and could be expected to produce larger residuals between measurements and fits. But for Example (A), as for actual retrievals of VIRTIS spectra [11], the residuals in fact decrease on average. This indicates avoided subsidiary solutions of F_c, justifying this correlated retrieval. However, when the imposed correlations are too strong, translating to an overly reduction of the effective size of the parameter space, residuals turn out to become larger again, since the global minimum itself is overly affected.

In addition, Example (A) illustrates that parameters with strongly differing correlation lengths can be better disentangled by using multi- than by using single-spectrum regularization. But the more similar any two parameters' Jacobians are (smaller Euclidean angle), the worse they can be disentangled (modes 2 and 2'). Note that for parameters with equal Jacobians, the distribution of the retrieved parameters is determined by *a priori* correlation data only, and disentanglement is not based on measurements anymore. This case must be avoided and underlines the necessity to check robustness of the retrieved results against reasonable *a priori* data and initial guess modifications. In Example (A), noise is sufficiently small to just disentangle cloud modes 2 and 2' by using MSR, but stronger noise destroys more information. In practice (noise with $2\sigma = 2\cdot 10^{-3}\,\mathrm{W/(m^2\ sr\ \mu m)}$ and systematic errors), modes 2 and 2' can not be disentangled from the considered spectral range, and only mode 2' and 3 variations are considered in the following. Resulting real-world retrieval errors will be studied elsewhere.

Note that, when a smooth behavior of the retrieved result is expected, it is far better to increase the probability to find a smooth parameter set as retrieval solution, than to smooth results from uncorrelated retrievals. Due to the non-linear nature of the forward model, smoothing of parameters will in general not lead to a consistent parameter set describing the measurements, especially when jumps between different subsidiary minima are involved, or parameters close to their domain boundaries.

In conclusion of Example (A), *a priori* spatial-temporal correlations can be regarded as 'elastic bands' forcing the parameters to stay close to self-establishing general spatial-temporal trends. This increases the probability of convergence to spatial-temporal parameter distributions compatible with the expectations on continuity and distribution of the true atmospheric state, attenuates the impact of noise, and decreases the probability of running into subsidiary minima of the cost function.

As discussed, parameters with strongly differing spatial-temporal correlations can be better disentangled. The extreme case herefore is that of infinite spatial or temporal coupling, translating to spatial or temporal constancy of the respective parameters when *a priori* mean values and standard deviations are identic. But as that causes $\mathbf{S}_A$ to degenerate, this case must be treated in a different way (Appendix A.3), by retrieving parameters that are common to certain sets of spectra (Section 3.5).

To illustrate this case, **Example (B)** studies a synthetic observation movie with 30 repetitions (1 h intervals) of 30 surface bins that evenly cover the equator from 1–30 °E (Figs. 4 and 5). The spectra are generated as being acquired by the VIRTIS-M-IR detector that has 256 spatial samples in a row, by the samples with numbers $s = 4 + 8\cdot\mathrm{Longitude}/^\circ\mathrm{E}$. Cloud modes 2' and 3 are independently varied according to a pseudo-random spatial-temporal pattern with 1000 km correlation length, 10 h correlation time, mean value ≈ 1.0, and double standard deviation ≈ 0.6. Along the sample direction, the FWHM of the instrumental response function is varied according to $(17 + (s - 124)/48)$ nm which is inspired by test retrievals from actual measurements. Surface emissivities in the spectral transparency windows at 1.02, 1.10, and 1.18 µm are common to all spectra covering the same respective surface bin. 1.02 µm-surface-emissivity is modeled to span the whole range 0–1 and to have three four-bin plateaus at emissivity levels 0.2, 0.65, and 0.98, as well as an abrupt anomaly around 21.5 °E. Emissivities in the other two surface windows are all set to 0.65. The remaining parameters are as for the reference spectrum (Fig. 2). In particular, the continuum is common to all

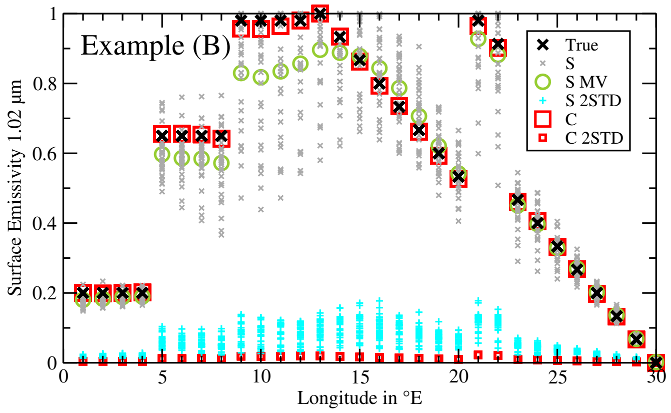

Figure 4: Comparison of 1.02 µm-surface emissivities of synthetic cube ('True') with corresponding retrieved values for different regularization models. S: single-spectrum results, S MV: their mean values for the respective longitude bins, S 2STD: twice their *a posteriori* standard deviations. C: surface emissivities as parameters common to their respective surface bins, C 2STD: twice their *a posteriori* standard deviations.

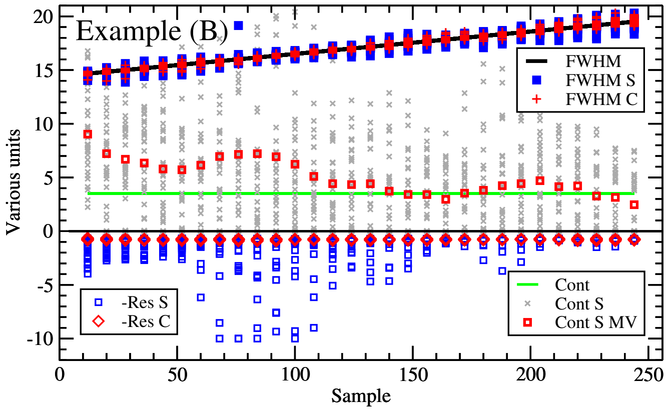

Figure 5: Detector sample dependence of various parameters. FWHM: 'true' FWHM of instrumental response function in nm, FWHM S: retrieved single-spectrum results, FWHM C: results of correlated retrievals. Cont: 'true' continuum parameter ($3.5 \cdot 10^{-30}$ cm^2) for 1.18 µm window, Cont S: retrieved continuum parameters (single-spectrum results, same unit), Cont S MV: their mean values for the respective detector sample bins. Retrieved continuum as parameter common to all spectra is 3.32 ± 0.24 in the same unit. Res S: least-squares norms in 10^{-4} W/(m^2 sr µm) of residuals between synthetic and fitted radiances, divided by number of utilized spectral bands (single-spectrum results), Res C: the same for multi-spectrum results.

spectra. Gaussian noise with double standard deviation $2\sigma = 2 \cdot 10^{-3}$ W/(m^2 sr µm) is added to the synthetic spectra.

In a first stage of the retrieval pipeline, cloud mode factors 2' and 3 (1 ± 20, 1000 km, 10 h), FWHM ((17 ± 30) nm, 75 samples, 5 h), and continua at 1.31, 1.74, and 2.3 µm ((1 ± 10^3)$\cdot 10^{-29}$ cm^2, 0 km, 0 h) are retrieved from 1.295–2.300 µm, where the values in parentheses list *a priori* mean value ± double standard deviation, correlation length and time of the respective parameter. The second stage retrieves 1.02 µm-surface-emissivity (0.5 ± 20, 0 km, 0 h), and 1.10 µm- and 1.18 µm-surface-emissivity and -continuum from their respective peaks. The *a priori* standard deviations are set very wide to largely exclude their and the mean values' impact. Results from single- (no common parameters, no *a priori* correlations) and multi-spectrum regularization (continuum common to all spectra, surface emissivities common to all spectra covering the same surface bin, *a priori* correlation lengths and times as given except for common parameters) are compared to true values in Figs. 4 and 5.

Retrieved single-spectrum continua (only depicted for 1.18 µm) have a large scatter. Their mean values ± their double standard deviations are $(1.8 \pm 1.2, 0.50 \pm 0.90, 4.8 \pm 7.6, 15.8 \pm 4.2, 124 \pm 50) \cdot 10^{-29}$ cm^2 for the spectral windows at 1.10, 1.18, 1.31, 1.74, 2.3 µm (note missing 1.02 and 1.28 µm windows). Continua as common parameters ± their double *a posteriori* standard deviations are $(2.00 \pm 0.09, 0.33 \pm 0.03, 4.2 \pm 0.3, 15.02 \pm 0.07, 119.5 \pm 0.5) \cdot 10^{-29}$ cm^2 (same windows). The single-spectrum mean values and the multi-spectrum values agree with the true values within the given margins, but the margins for the single-spectrum results are considerably wider. Note that the 95%-confidence intervals for the 900 single-spectrum results (multiply double standard deviations by $1.96/(2 \cdot \sqrt{900})$ according to Student's t-distribution) are in the order of magnitude of the multi-spectrum *a posteriori* double standard deviations, meaning that single-spectrum retrieval has statistically failed to retrieve the continuum.

Retrieved cloud mode 2' factors (mode 3 factors) have a root-mean-square deviation (RMSD) of 0.28 and 0.053 (0.15 and 0.027) to their true values for single- and multi-spectrum results, respectively. As the true modes 2' and 3 have the same mean value and standard deviation, this shows that cloud mode 3 can be retrieved more reliably than mode 2'. MSR results are more reliable than single-spectrum results.

The FWHM can be retrieved quite reliably already for single-spectrum regularization (RMSD 0.31 nm), although MSR results (RMSD 0.19 nm) are more reliable.

Retrieved single-spectrum surface emissivities (only depicted for 1.02 µm, RMSD 0.11) have a large scatter and large *a posteriori* double standard deviations. Their bin-wise mean values do not match very well to the true values. Especially the larger emissivities are difficult to retrieve, since the radiance response to emissivity changes near 1.0 is small compared with changes in the lower emissivity range. This effect amplifies noise impacts at higher emissivities. An additional contribution to the degree of the emissivity underestimation is the presence of the upper emissivity domain boundary. This can be seen in the extreme case where true emissivity is 1 and scattered retrieved values are only allowed to extend to values ≤ 1 such that the mean value must be strictly < 1 for non-zero scatter. Still, the spatial fine structure (plateaus, jumps, anomaly) is resolved to a certain degree. MSR results (RMSD 0.0086) agree well with the true values and have very small *a posteriori* double standard deviations, representing the increased information content per spectrum due to the restriction of the effective size of the parameter space by incorporating finite and infinite *a priori* correlations. As the 1.02 µm-continuum is not retrieved, but set to its true value, wrong single-spectrum 1.02 µm-emissivity mean values are here not related to a wrong continuum in this window, but to wrong cloud modal abundances partly caused by wrong continua between 1.295–2.300 µm, and by convergence to subsidiary minima.

For the not depicted surface emissivities at 1.10 and 1.18 µm, the true values are 0.65 and 0.65 for each of the bins. The retrieved values are scattered according to 0.49 ± 0.8 and 0.69 ± 0.5 with RMSD 0.40 and 0.21 (single-spectrum), and 0.65 ± 0.04 and 0.63 ± 0.02 with RMSD 0.016 and 0.024 (multi-spectrum bins). This shows that MSR results are more reliable than single-spectrum results. Also, emissivities at 1.10 and 1.18 µm are more difficult to retrieve than those at 1.02 µm, in part a consequence of the lower surface radiance contribution in these peaks [12, Fig. 5]. But it is also due to the utilization of the 'true' 1.02 µm-continuum, and thus, neglect of its retrieval error.

Example (B) underlines that retrieved common parameters (continuum, emissivity) are not just mean values of single-spectrum results. This can be explained by observing the least-squares norms of residuals between synthetic and fitted radiances. While they are very close to the artificial noise level for MSR fits, they are about 60% larger on average for single-spectrum fits. About 30% of the single-spectrum fits are significantly worse than the corresponding MSR fits. As in Example (A), this demonstrates the superiority of MSR in avoiding subsidiary minima of the cost function. Example (B) also illustrates improved disentanglement between parameters with similar Jacobians in two examples for extreme cases of strongly differing *a priori* correlation behavior. The first example is the disentanglement of continua (infinite correlation length and time) from clouds (finite correlation lengths and times) in the first stage of the retrieval pipeline (1.295–2.300 µm). The second example is the disentanglement of surface emissivities (correlation length 0, correlation time ∞) from continua (correlation length ∞, correlation time ∞) in the second stage of the retrieval pipeline (1.000–1.235 µm).

An example to illustrate the retrieval of locally coupled parameters in presence of spatial-temporal couplings (e.g. temperature altitude profiles with vertical coupling of level temperatures and horizontal spatial-temporal coupling of temperatures from contiguous measurements, preventing arbitrarily large retrieved temperature fluctuations between contiguous levels and footprint locations) will be studied in a subsequent paper.

6. Conclusions and outlook

Currently, VIRTIS-M-IR spectra of Venus' nightside emissions establish the only data source in the infrared with high repetition and spatial resolution, where Venus' surface emissivity can be extracted from on a global scale. A radiative transfer forward model is required to simulate spectra in dependence on surface, atmospheric, and instrumental parameters. A retrieval algorithm iteratively varies these parameters until the simulated well fit the measured spectra. The so-retrieved parameters are interpreted as the surface, atmospheric, and instrumental state that led to the measurements. But single VIRTIS-M-IR spectra have a comparatively low information content, and different parameter combinations can describe the same measurement equally well. Hence, the inversion of the forward model is mathematically an ill-posed problem and must be regularized. A common approach is the minimization of a non-negative retrieval cost function that arises from the incorporation of Bayesian *a priori* mean values, standard deviations, and correlations the retrieval parameters are to respect, as well as measurement and simulation error information. This essentially rules out unlikely state vectors. Still, this cost function is a non-linear function on a possibly high-dimensional space, and it is easy to run into subsidiary local minima far away from the global minimum.

It was exemplarily shown that the multi-spectrum regularization presented in this paper (MSR) can considerably improve the data analysis of contiguous measurements with sparse spectral information content by minimizing a retrieval cost function on a parameter space that encompasses several measurements and that incorporates *a priori* spatial-temporal correlations between the state vectors of the different measurements. This naturally arising regularization decreases the probability to retrieve unlikely spatial-temporal parameter distributions. Uninteresting subsidiary minima of the retrieval cost function and unphysical jumps between them can be better avoided, and the impact of noise can be attenuated.

A detailed error analysis of the single-spectrum retrieval of local surface emissivity at 1.02, 1.10, and 1.18 µm on Venus shows that this parameter is difficult and error-prone to retrieve [23], as will be presented in a subsequent paper. But neglecting VIRTIS-observable geologic activity on Venus, surface emissivity is common to measurements that repeatedly cover the same surface spot on Venus. In the same way, corrections to the CO_2 opacity in the extreme environmental conditions in the deep atmosphere of Venus ('continuum') are common to all measurements of Venus' nightside emissions. Knowledge gain from single-spectrum retrieval of these hard-to-determine parameters is very limited due to interfering atmospheric variations. As it was shown, their single-spectrum retrieval thus statistically fails, although the spectra are sensible to these parameters.

But MSR is especially useful to disentangle retrieval parameters with similar Jacobians and strongly differing correlation lengths or times. The extreme case for parameters with infinite correlation lengths or times and identic single-spectrum *a priori* data corresponds to the retrieval of parameters common to certain selections of measurements, so for instance for surface emissivity (correlation length 0, correlation time ∞) and continuum (infinite correlation length and time). It was demonstrated that this approach leads to a better disentanglement of continua from spatially-temporally varying atmospheric parameters and of emissivities from continua, to smaller residuals between measurements and fits, and thus to more reliable retrieval results than corresponding single-spectrum retrievals.

The *a posteriori* retrieval uncertainties are lower for MSR compared to single-spectrum retrievals, as it was shown (see also Kappel et al. [11] for real-world examples). This results from the incorporation of the context of adjacent measurements, making more information available that contributes to the determination of the parameters describing a certain spectrum. Especially when single-spectrum information content is low, this is important to improve the quality of the retrieved information. Single-spectrum retrieval, on the other hand, can be regarded as an only very rough approximation of reality, since the presence of spatial-temporal correlations is rather the regular case.

It was verified that a simple smoothing or averaging of retrieved single-spectrum results does in general not lead to correct results or to a consistent parameter set describing the measurements, especially when there are unphysical jumps between different subsidiary minima or parameters close to their domain boundaries. By using MSR, all considered measurements can be parameterized by a fully consistent set of atmospheric, surface, and instrumental parameters that respects all available single- and multi-spectrum *a priori* data as well as the measurement and simulation error distributions. But as always, when regularizing an ill-posed problem, it must be checked that retrieved results do not significantly change for *a priori* and initial guess modifications.

Section 5 demonstrated that MSR allows to retrieve continua and surface emissivities from VIRTIS-M-IR measurements, when random measurement errors are assumed to be the only error sources. Note that common parameters retrieved from real-world measured spectra have to be carefully interpreted, since systematic measurement and simulation errors are also common to the spectra. The impact of systematic errors will be discussed in a subsequent paper. First tests revealed that in their presence, knowledge of continua is crucial to reliably retrieve emissivities, and it may be necessary to first assume mean surface emissivities in order to determine surface window continua. The constraint that emissivities globally must be non-negative and must not exceed unity then helps to constrain valid continua. On the other hand, it may be sufficient to utilize more diverse measurements (particularly with respect to topography) to disentangle emissivities and continua, but further studies are required.

Kappel et al. [11] already applied MSR to actual VIRTIS-M-IR spectra of Venus' nightside in order to disentangle continua from spatial-temporal atmospheric variations. As shall be presented in a subsequent paper, it will also be applied to retrieve surface emissivity maps of Venus as parameter vectors that are common to measurements that repeatedly cover the same surface bins. Compared to single-spectrum retrieval, this approach also has a higher chance of success, because relative changes in the spatial distribution of surface emissivity may be easier to detect than absolute values. This is a consequence of the underlying physical continuity in spatial-temporal variations of interfering parameters like minor gas and cloud modal distributions, temperature variations, and others. The presented multi-spectrum retrieval algorithm is ideally suited for this task, since it allows to incorporate all these continuity and consistency constraints. In conjunction with a refined consistent data calibration and preprocessing [11], retrieval reliability and accuracy can thus be pushed to their limits. However, while the processing time overhead of MSR compared to corresponding single-spectrum retrievals is negligible for up to a few thousand spectra, any retrieval based on full radiative transfer forward model simulations requires considerable computational resources. Thus, MSR will be selectively applied at first to localized targets that were beforehand identified to be of special geological interest [53–55].

A. Appendix

Some mathematical derivations and details on the implementation of MSR are presented here.

A.1. Cost function as least squares norm

This appendix reformulates the cost function in terms of a least-squares norm and discusses MSR's inputs, i.e. residual and Jacobian of the extended forward model simulations, as well as its outputs, i.e. the best estimate of the state vector and the corresponding *a posteriori* covariance matrix.

The best estimate of the state vector that is compatible with the *a priori* and error knowledge and adequately parameterizes the measurements, is the global minimum of the cost function F_c from Eq. (1). An iterative algorithm is used to identify local minima of F_c. When a cost function has the form of a least-squares residual norm $\|\mathbf{R}(\mathbf{X})\|_2$, this special structure can be exploited to improve numerical efficiency, like it is done by the numerically quite robust trust region formulation of the Levenberg-Marquardt algorithm [56]. Here, an $\mathbf{X}$ is determined that locally minimizes the least-squares norm, or equivalently $\|\mathbf{R}(\mathbf{X})\|_2^2$. To see that F_c has this structure, define the 'scaled residual' in the notation of Section 2

$$\mathbf{R}(\mathbf{X}) := \begin{pmatrix} \mathbf{S}_A^{-1/2}(\mathbf{X}-\mathbf{A}) \\ \mathbf{S}_E^{-1/2}(\mathbf{Y}-\mathbf{F}(\mathbf{X})) \end{pmatrix} \in \mathbb{R}^{N+M}, \tag{A.1}$$

which implies $\|\mathbf{R}(\mathbf{X})\|_2^2 = \mathbf{R}(\mathbf{X})^T\mathbf{R}(\mathbf{X}) = F_c(\mathbf{X})$. $\mathbf{R}$ is a mapping from $\mathbb{R}^N$ to $\mathbb{R}^{N+M}$. $\mathbf{S}_A^{-1/2}$ is such that $(\mathbf{S}_A^{-1/2})^T\mathbf{S}_A^{-1/2} = \mathbf{S}_A^{-1}$, and analogously with $\mathbf{S}_E$. $\mathbf{S}_A^{-1/2}$ is the inverse square root of the positive definite symmetric matrix $\mathbf{S}_A$ and is here not computed by using spectral decomposition, but by the numerically fast and stable Cholesky decomposition. Any real, symmetric, positive definite $N \times N$ matrix $\mathbf{S}_A$ can be Cholesky decomposed into the product $\mathbf{S}_A = \mathbf{U}^T\mathbf{U}$ of the transpose of an upper triangular matrix $\mathbf{U}$ and $\mathbf{U}$ itself by using asymptotically $N^3/3$ [57, Section 4.2] arithmetical operations, compared to $9N^3$ arithmetical operations for diagonalization by the symmetric QR algorithm [57, Section 8.3]. $(\mathbf{U}^T)^{-1} =: \mathbf{U}^{-T} =: \mathbf{S}_A^{-1/2}$ is then an inverse matrix square root of $\mathbf{S}_A$ in the required sense. See Appendix A.2 for an efficient computation for matrices $\mathbf{S}_A$ with a structure as presented in Section 3. The difference to the matrix square root, if defined by matrix diagonalization, is just an orthogonal transformation of $\mathbf{R}(\mathbf{X})$ that is consequently not observable when minimizing the least-squares norm of $\mathbf{R}(\mathbf{X})$. The inverse square root of $\mathbf{S}_E$, when assumed to be a diagonal matrix, is just the matrix $\mathbf{S}_E$ with its diagonal entries replaced by their inverse square roots.

The Jacobian $\mathbf{J}$ of $\mathbf{R}$ is needed as input to the iterative algorithm [56] and follows from Eq. (A.1) as

$$\mathbf{J}(\mathbf{X}) = \nabla\mathbf{R}(\mathbf{X}) = \begin{pmatrix} \mathbf{S}_A^{-1/2} \\ -\mathbf{S}_E^{-1/2}\mathbf{K}(\mathbf{X}) \end{pmatrix} \in \mathbb{R}^{N+M} \times \mathbb{R}^N, \tag{A.2}$$

with the Jacobian $\mathbf{K}$ of the forward model $\mathbf{F}$. $\mathbf{J}(\mathbf{X})$ is the largest data structure in MSR and determines the problem size still manageable on a given computer hardware.

Sparse matrix formulation of $\mathbf{J}(\mathbf{X})$, and thereby of the entire retrieval algorithm, considerably increases that limit and is discussed in Appendix A.4.

When the best estimate of $\mathbf{X}$ is denoted by $\widehat{\mathbf{X}}$ (the retrieved solution), the $N \times N$-dimensional *a posteriori* covariance matrix $\widehat{\mathbf{S}}$ at the retrieved solution follows from (compare Eq. (2))

$$\widehat{\mathbf{S}}^{-1} = \big(\mathbf{J}(\widehat{\mathbf{X}})\big)^T \mathbf{J}(\widehat{\mathbf{X}}). \tag{A.3}$$

The corresponding correlation matrix $\widehat{\mathbf{C}}$ follows from Eq. (3). The block structures of $\widehat{\mathbf{S}}$ and $\widehat{\mathbf{C}}$ are inherited from the structure of $\widehat{\mathbf{X}}$. The diagonal entries $(\widehat{\mathbf{S}})_{ii}$ of $\widehat{\mathbf{S}}$ are the variances σ_i^2 of the retrieved parameters according to the *a posteriori* probability distribution. $2\sigma_i$ provides a first measure for the retrieval uncertainty, but a detailed retrieval error analysis should be performed in addition, as will be presented in a subsequent paper. The off-diagonal entries of $\widehat{\mathbf{C}}$ provide an estimate on how well a retrieved parameter is disentangled from the influences of other parameters. An absolute value close to 1 indicates bad disentanglement. $\widehat{\mathbf{C}}$ may be costly to compute and to store, and often, it is sufficient to compute the blocks corresponding to a few measurements and the related common-parameter-blocks. This already provides a good impression on the disentanglement of common from local parameters, the correlations between parameters associated to different measurements, between common parameters, and between local parameters belonging to one measurement. For an efficient computation of the diagonal entries of $\widehat{\mathbf{S}}$ and a few representative off-diagonal entries of $\widehat{\mathbf{C}}$, see Appendix A.4. Note that the correct weighting of the common parameters (Appendix A.3) affects both $\widehat{\mathbf{X}}$ and $\widehat{\mathbf{S}}$.

Violations of the retrieval parameters' physical domain boundaries are prevented by the logarithmic barrier method [58, Section 17.2], but with slight modifications to match the presented least-squares-norm formulation. The barrier function B^2 is added to the cost function F_c as a penalty for each parameter $(\mathbf{X})_i$ that approaches one of its domain boundaries (e.g. 0 or 1 for the 1.02 µm-surface-emissivity of a certain spectrum), where $B(\mathbf{X}) := -\mu \sum_{i=0}^{N} \log\big(c_i\big[(\mathbf{X})_i\big]\big)$. B^2 is chosen as barrier function instead of B, to allow better incorporation into the least-squares formulation of F_c by introducing an additional dimension to $\mathbf{R}(\mathbf{X})$ with the entry $B(\mathbf{X})$, such that $\mathbf{R}(\mathbf{X})^T\mathbf{R}(\mathbf{X})$ corresponds to the original $F_c(\mathbf{X})$ plus $\big(B(\mathbf{X})\big)^2$. c_i is continuous, piecewise differentiable, positive when $(\mathbf{X})_i$ is inside its domain, and linearly approaches 0 at the boundaries. As an additional modification, to minimize impact of the barrier function, $\log(c_i)$ shall be 0 outside a certain small boundary region within the domain. To remove the influence of B on the retrieved result, μ is decreased by a factor of 10 each time a certain number of iterations is completed, until $B(\mathbf{X})$ is neglectable compared to $F_c(\mathbf{X})$. When an iteration step would lead a number of parameters to violate a domain boundary, they are set back into their domain, close (dependent on μ) to the boundary. This prevents the trust region radius to contract to zero too early and still leads to correct results. The Jacobian of $\mathbf{R}$ including the barrier dimension follows immediately by differentiation. For μ small enough, $\widehat{\mathbf{S}}$ is unaffected by B.

A.2. Inverse square root of a priori covariance matrix

This appendix discusses the efficient computation of the inverse square root $\mathbf{S}_A^{-1/2}$ of the covariance matrix $\mathbf{S}_A$ as needed in Appendix A.1. As $\mathbf{S}_A$ is block diagonal with $\mathbf{S}_C$ and $\mathbf{S}_L$ on its diagonal (Section 3.5), the inverse square roots of $\mathbf{S}_C$ and $\mathbf{S}_L$ can be computed independently. Even for larger retrieval problems, $\mathbf{S}_C$ is of rather low dimension (see example in Appendix A.4), and it is already efficient to Cholesky decompose $\mathbf{S}_C =: \mathbf{U}^T\mathbf{U}$ and to compute $\mathbf{S}_C^{-1/2} := \mathbf{U}^{-T}$. But $\mathbf{S}_L$, for instance for thousands of measurements and ten retrieval parameters per measurement, is of the order of dimension $10,000 \times 10,000$ and costly to Cholesky decompose. However, the Kronecker product structure of $\mathbf{S}_L$ (Section 3.4) allows for a computational shortcut. Without loss of generality, retrieval of common parameters is not considered here, i.e. $\mathbf{S}_A$ is assumed to only comprise $\mathbf{S}_L$ to share notation with Section 3.4.

According to Golub and van Loan [57, Section 4.5.5], Cholesky decomposition and Kronecker multiplication commute in the sense that

$$\begin{aligned}\mathbf{U}_\mathbf{S}^T\mathbf{U}_\mathbf{S} = \mathbf{S} := \mathbf{H}\otimes\mathbf{G} = \\ (\mathbf{U}_\mathbf{H}^T\mathbf{U}_\mathbf{H})\otimes(\mathbf{U}_\mathbf{G}^T\mathbf{U}_\mathbf{G}) = (\mathbf{U}_\mathbf{H}\otimes\mathbf{U}_\mathbf{G})^T(\mathbf{U}_\mathbf{H}\otimes\mathbf{U}_\mathbf{G})\end{aligned}$$

with obvious notation. This is due to the relations $(\mathbf{A}\otimes\mathbf{B})^T = \mathbf{A}^T\otimes\mathbf{B}^T$ and $(\mathbf{A}\otimes\mathbf{B})(\mathbf{C}\otimes\mathbf{D}) = (\mathbf{AC})\otimes(\mathbf{BD})$ (which also implies $(\mathbf{A}\otimes\mathbf{B})^{-1} = \mathbf{A}^{-1}\otimes\mathbf{B}^{-1}$) [34, Section 4.2], and the fact that the Kronecker product of upper triangular matrices is upper triangular, all following immediately from definition. This can be used to rearrange the computation of $\mathbf{S}_A^{-1/2}$ as follows.

$\mathbf{S}_A$ is defined as $\mathbf{C}_A := (\mathbf{\Pi}_n^r)^T\overline{\mathbf{C}}_A\mathbf{\Pi}_n^r$ (Section 3.4) scaled with respect to the *a priori* standard deviations σ_k of the retrieval parameters (Eq. (3)) with $\overline{\mathbf{C}}_A$ from Eq. (15).

As a first step, it can be verified that $\mathbf{S}_A$ may also be computed by first scaling the $\mathbf{h}_g$ in Eq. (15) with the *a priori* standard deviations, and only then to perform the Kronecker products, form the large block diagonal matrix, and finally permute. Herefore, observe that

$$(\mathbf{S}_A)_{n(i-1)+k\,,\,n(j-1)+l} = \sigma_k\sigma_l\,(\overline{\mathbf{C}}_A)_{r(k-1)+i\,,\,r(l-1)+j}, \tag{A.4}$$

for all measurements $i,j\in\{1,\cdots,r\}$ and for all parameters $k,l\in\{1,\cdots,n\}$, corresponding to Eq. (16) and considering the permutation in the definition of $\mathbf{C}_A$.

Let $\mathbf{h}$ be the block diagonal matrix with the matrices $\mathbf{h}_1,\cdots,\mathbf{h}_G$ on its diagonal and entries h_{kl}. $\mathbf{H}$ shall be the covariance matrix that arises from $\mathbf{h}$, i.e. $H_{kl} = \sigma_k\sigma_l\,h_{kl}$, and the blocks on the diagonal of the block matrix $\mathbf{H}$ are denoted by $\mathbf{H}_1,\cdots,\mathbf{H}_G$. According to Eqs. (15) and (10), Eq. (A.4) can then be written as

$$(\mathbf{S}_A)_{n(i-1)+k\,,\,n(j-1)+l} = \sigma_k\sigma_l\,h_{kl}(\boldsymbol{\varrho}_g)_{ij} = H_{kl}(\boldsymbol{\varrho}_g)_{ij},$$

where $g\in\{1,\cdots,G\}$ depends on the partitioning of $\mathbf{h}$.

Hence, $\mathbf{S}_A = (\mathbf{\Pi}_n^r)^T \overline{\mathbf{S}}_A \mathbf{\Pi}_n^r$, with

$$\overline{\mathbf{S}}_A := \begin{pmatrix} \mathbf{H}_1 \otimes \boldsymbol{\varrho}_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \mathbf{H}_G \otimes \boldsymbol{\varrho}_G \end{pmatrix}. \tag{A.5}$$

Incidentally, this requires fewer multiplications for the scaling, since $\mathbf{h}$ has to be scaled just once (n^2 multiplications), whereas for the scaling of $\mathbf{C}_A$, each of the r^2 blocks of size $n \times n$ has to be scaled.

Next, observe that the block diagonal structure of $\overline{\mathbf{S}}_A$ and the uniqueness of Cholesky decomposition ensure that the upper triangular Cholesky factor $\overline{\mathbf{U}}$ of $\overline{\mathbf{S}}_A = \overline{\mathbf{U}}^T\overline{\mathbf{U}}$ is block diagonal with the upper triangular Cholesky factors $\mathbf{U}_g$ of the single blocks $\mathbf{H}_g \otimes \boldsymbol{\varrho}_g = \mathbf{U}_g^T\mathbf{U}_g$ as the blocks on its diagonal.

$$\overline{\mathbf{S}}_A = \overline{\mathbf{U}}^T\overline{\mathbf{U}} = \begin{pmatrix} \mathbf{U}_1^T & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \mathbf{U}_G^T \end{pmatrix} \begin{pmatrix} \mathbf{U}_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \mathbf{U}_G \end{pmatrix}$$

Clearly, $\overline{\mathbf{U}}$ is upper triangular. Since Cholesky decomposition and Kronecker multiplication commute, each $\mathbf{U}_g$ can be computed as $\mathbf{U}_g = \mathbf{U}_{\mathbf{H}_g} \otimes \mathbf{U}_{\boldsymbol{\varrho}_g}$ with obvious notation. Note that the $\mathbf{U}_g$ are possibly all of different dimension $n_g r \times n_g r$.

Consider the permutation $(\mathbf{\Pi}_n^r)^T(\cdot)\mathbf{\Pi}_n^r$ that maps the index pair $\big(r(k-1)+i\,,\, r(l-1)+j\big)$ to $\big(n(i-1)+k\,,\, n(j-1)+l\big)$. According to the definition of the Kronecker product, each $\mathbf{U}_g$ is a block matrix with blocks of size $r \times r$ that are all upper triangular. Thus, $\overline{\mathbf{U}}$ also is an upper triangular block matrix with blocks of size $r \times r$ that are all upper triangular. Hence, the indices of the entries $\overline{U}_{r(k-1)+i\,,\,r(l-1)+j}$ that are non-zero, satisfy $k \le l$ and $i \le j$. This implies $n(i-1)+k \le n(j-1)+l$, i.e. that $(\mathbf{\Pi}_n^r)^T\overline{\mathbf{U}}\mathbf{\Pi}_n^r$ is upper triangular. This proves that $\big((\mathbf{\Pi}_n^r)^T\overline{\mathbf{U}}\mathbf{\Pi}_n^r\big)^T(\mathbf{\Pi}_n^r)^T\overline{\mathbf{U}}\mathbf{\Pi}_n^r = (\mathbf{\Pi}_n^r)^T\overline{\mathbf{S}}_A\mathbf{\Pi}_n^r = \mathbf{S}_A$ is a (i.e. *the*) Cholesky factorization of $\mathbf{S}_A = \mathbf{U}^T\mathbf{U}$, with upper triangular Cholesky factor $\mathbf{U} = (\mathbf{\Pi}_n^r)^T\overline{\mathbf{U}}\mathbf{\Pi}_n^r$, i.e. $\mathbf{U}$ can be computed by permuting $\overline{\mathbf{U}}$.

Finally, observe that

$$\mathbf{S}_A^{-1/2} := \mathbf{U}^{-T} = \big((\mathbf{\Pi}_n^r)^T\overline{\mathbf{U}}\mathbf{\Pi}_n^r\big)^{-T} = (\mathbf{\Pi}_n^r)^T\overline{\mathbf{U}}^{-T}\mathbf{\Pi}_n^r.$$

Note that the transposed inverse of the upper triangular matrix $\mathbf{U}$ is lower triangular. Also, as $\overline{\mathbf{U}}$ is upper triangular, $\overline{\mathbf{U}}^{-T}$ is lower triangular. Due to the block diagonal structure of $\overline{\mathbf{U}}$, its inverse can simply be computed by forming the block diagonal matrix with the $\mathbf{U}_g^{-1}$ as the blocks on its diagonal. Because of $[\mathbf{A}\otimes\mathbf{B}]^{-1} = \mathbf{A}^{-1}\otimes\mathbf{B}^{-1}$ and $[\mathbf{A}\otimes\mathbf{B}]^T = \mathbf{A}^T \otimes \mathbf{B}^T$, it finally follows

$$\mathbf{S}_A^{-1/2} = (\mathbf{\Pi}_n^r)^T \begin{pmatrix} \mathbf{U}_{\mathbf{H}_1}^{-T} \otimes \mathbf{U}_{\boldsymbol{\varrho}_1}^{-T} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \mathbf{U}_{\mathbf{H}_G}^{-T} \otimes \mathbf{U}_{\boldsymbol{\varrho}_G}^{-T} \end{pmatrix} \mathbf{\Pi}_n^r. \tag{A.6}$$

Thus, direct Cholesky decomposition of $\mathbf{S}_A = \mathbf{U}^T\mathbf{U}$ and inversion of $\mathbf{U}$ can be avoided. Only the G $n_g \times n_g$-dimensional matrices $\mathbf{H}_g$ and the G $r \times r$-dimensional matrices $\boldsymbol{\varrho}_g$ need to be Cholesky decomposed and their upper triangular factors inverted. Permutation and Kronecker multiplication are computationally fast operations. Compare for large r the number of arithmetic operations for the resource dominating Cholesky decompositions: $(nr)^3/3$ for $\mathbf{S}_A$ vs. $\sum_{g=1}^G n_g^3/3 + Gr^3/3$ for the Cholesky decompositions involved in Eq. (A.6). For $r \gg n$, $\sum_{g=1}^G n_g^3/3$ can be neglected against $Gr^3/3$, and the speedup is of the order of n^3/G, which exceeds n^2 due to $G \le n$. Generically, for ten retrieval parameters per measurement, the speedup exceeds 100. Incidentally, this framework is also ideally suited for the sparse matrix formulation of MSR, as will be seen in Appendix A.4.

A.3. Limit for perfect spatial-temporal coupling

In this appendix, the limit of retrieved values for ever stronger spatial or temporal coupling between r measurements for c single-spectrum parameters with identic single-spectrum *a priori* data will be discussed. It will be shown that computing this limit is equivalent to retrieving c parameters common to these measurements in the sense of Section 2, with a certain relative weighting between common and not-common ('local', i.e. spatially-temporally varying) parameters. It suffices to show that in both approaches, the *a posteriori* probability distribution (Section 2) yields the same best estimates and uncertainties of the retrieval parameters. The main purpose of this section is the derivation of the proper relative weighting between the common and the local parameters in the *a priori* covariance matrix. From the outset, it is not clear whether there is a statistical $\sqrt{r}$-like relative weighting factor that reflects the influence of a common parameter on r spectra.

First, the setting will be defined by considering the permuted parameter space (Eq. (13)). The corresponding *a priori* covariance matrix $\overline{\mathbf{S}}_A$ can be written as in Eq. (A.5). Parameters with, in the limit, perfect spatial or temporal coupling ('perfect-coupling-parameters') must be implemented as separate from, and without intergroup coupling to parameter groups describable by constant finite spatial or temporal coupling ('finite-coupling-parameters'), as they obey different spatial-temporal correlation behavior (Section 3.4). To allow coupling between themselves, these perfect-coupling-parameters may well be combined into one single group, since they can be described by identical spatial-temporal correlation behavior. Denote this parameter group by the index P, such that $\overline{\mathbf{S}}_A = \begin{pmatrix} \overline{\mathbf{S}}_P & 0 \\ 0 & \overline{\mathbf{S}}_L \end{pmatrix}$. Here, $\overline{\mathbf{S}}_L$ is the *a priori* covariance matrix for the finite-coupling-parameters (for the permuted parameter space) and can be written as in Eq. (A.5). Because the perfect-coupling-parameters shall all have identic single-spectrum *a priori* data, their *a priori* covariance matrix can be written $\overline{\mathbf{S}}_P := \mathbf{H}_C \otimes \boldsymbol{\varrho}_P$, where the $r \times r$ matrix $\boldsymbol{\varrho}_P$ describes their spatial-temporal coupling, and $\mathbf{H}_C$ is their (for all measurements the same!) $c \times c$ single-spectrum *a priori* covariance matrix. Let the perfect-coupling-group comprise the parameter vector $\overline{\mathbf{p}} = \mathbf{\Pi}_c^r\mathbf{p} = \big(\overline{\mathbf{p}}_1, \cdots, \overline{\mathbf{p}}_c\big)$ analog to the notation in Eq. (13) for the permuted parameter space. $\overline{\mathbf{p}}_k \in \mathbb{R}^r$ is for each $k \in \{1, \cdots, c\}$ a vector describing the spatial-temporal distribution of single-spectrum parameter number k, and $(\overline{\mathbf{p}}_k)_i = (\mathbf{p}_i)_k$ is the parameter number k of measurement $i \in \{1, \cdots, r\}$.

For easier notation, the limit will be computed by using the unpermuted parameter spaces (Eq. (12)) separately for the perfect- and for the finite-coupling-parameters, i.e. the permutation $(\mathbf{\Pi}^r_{c,n})^T$ will be applied to the permuted parameter space, and correspondingly $(\mathbf{\Pi}^r_{c,n})^T(\cdot)\mathbf{\Pi}^r_{c,n}$ to $\overline{\mathbf{S}}_A$, where $\mathbf{\Pi}^r_{c,n} := \left(\begin{smallmatrix} \mathbf{\Pi}^r_c & 0 \\ 0 & \mathbf{\Pi}^r_n \end{smallmatrix}\right)$. This has the advantage of being able to work in the respective unpermuted spaces while still transparently separating perfect- from finite-coupling-parameters in the *a priori* covariance matrix. It does not change results, provided it is kept track of the associations of the entries of the matrix to the entries of the parameter space. Hence, $\mathbf{S}_A = \left(\begin{smallmatrix} \mathbf{S}_P & 0 \\ 0 & \mathbf{S}_L \end{smallmatrix}\right)$ and $(\mathbf{p}_1, \cdots, \mathbf{p}_r, \mathbf{x}_1, \cdots, \mathbf{x}_r)$ are the basic quantities to work with, where $\mathbf{p}_i \in \mathbb{R}^c$ and $\mathbf{x}_i \in \mathbb{R}^n$, and the i-independent (identic single-spectrum *a priori* data!) *a priori* mean value vector for the $\mathbf{p}_i$ is $\mathbf{a}_P \in \mathbb{R}^c$ and that for the $\mathbf{x}_i$ is $\mathbf{a} \in \mathbb{R}^n$.

It will now be shown that in the limit of perfect spatial or temporal coupling between the $\mathbf{p}_i$ that have identic single-spectrum *a priori* data, the retrieved values of the $\mathbf{p}_i$ all coincide, i.e. that $\mathbf{p}_1$ is common to the r measurements. Also, it will follow directly from the *a posteriori* probability distribution that the proper relative weighting between the common and the local parameters is exactly 1.

First, some additional notational conventions shall be fixed. Denote $\mathbf{p}_i - \mathbf{a}_P =: \mathbf{z}_i$ and $\mathbf{x}_i - \mathbf{a} =: \mathbf{Z}_i$ with $\mathbf{z} = (\mathbf{z}_1, \cdots, \mathbf{z}_r)$ and $\mathbf{Z} = (\mathbf{Z}_1, \cdots, \mathbf{Z}_r)$, such that $(\mathbf{z}, \mathbf{Z}) =: \mathbf{W}$ is the partial permutation of the parameter vector $\mathbf{X}$ translated by its *a priori* mean value vector. The *a posteriori* probability distribution $P_p(\mathbf{X}|\mathbf{Y}) = \frac{1}{N}\exp\big(-\frac{1}{2}F_c(\mathbf{X})\big)$, compare Eq. (1) with normalization factor N can now be written as

$$-2\log\big(N\widetilde{P}_p(\mathbf{W}|\mathbf{Y})\big) = \\ \mathbf{W}^T\mathbf{S}_A^{-1}\mathbf{W} + \big(\mathbf{Y} - \widetilde{\mathbf{F}}(\mathbf{W})\big)^T\mathbf{S}_E^{-1}\big(\mathbf{Y} - \widetilde{\mathbf{F}}(\mathbf{W})\big)$$

with correspondingly transformed functions marked by tildes. In particular, $\widetilde{\mathbf{F}}(\mathbf{W}) := \mathbf{F}\big((\mathbf{p}_1, \mathbf{x}_1), \cdots, (\mathbf{p}_r, \mathbf{x}_r)\big) = \big(\mathbf{f}_1(\mathbf{p}_1, \mathbf{x}_1), \cdots, \mathbf{f}_r(\mathbf{p}_r, \mathbf{x}_r)\big)$, compare notation in Section 2 not considering $\mathbf{x}_C$.

Let the strength of the spatial-temporal coupling of the parameters $\mathbf{z}$ be parameterized by ε. In the limit of $\varepsilon \downarrow 0$, it shall continuously approach perfect spatial or temporal coupling ('or' also allows 'spatial and temporal'). Of the two blocks on the diagonal of the block diagonal matrix $\mathbf{S}_A$, the first block $\mathbf{S}_P(\varepsilon)$ is associated to the parameters $\mathbf{z}$ and depends on ε, and the other block $\mathbf{S}_L$ is associated to the parameters $\mathbf{Z}$ and does not depend on ε. Clearly, $\lim_{\varepsilon\downarrow 0}\mathbf{S}_P(\varepsilon) = \lim_{\varepsilon\downarrow 0}(\boldsymbol{\varrho}_P(\varepsilon) \otimes \mathbf{H}_C) = \mathbf{1}_{r\times r} \otimes \mathbf{H}_C$, where $\mathbf{1}_{r\times r}$ is the $r \times r$ matrix with all entries 1. $\mathbf{1}_{r\times r}$ is degenerate and thus has no inverse. This is the reason, why the *a posteriori* probability distribution for perfect coupling has to be defined by a limit. Note that the normalization factor N depends on ε. For $\varepsilon > 0$, the inverse of $\mathbf{S}_A$ is the block diagonal matrix with the inverses of the blocks of $\mathbf{S}_A$ as its blocks on the diagonal. The *a posteriori* probability distribution for $\mathbf{z}$, $P_z(\mathbf{z}, \varepsilon) := \int \widetilde{P}_p\,(\mathbf{z}, \mathbf{Z})|\mathbf{Y})\,\mathrm{d}\mathbf{Z}$, for finite ε can thus be written

$$P_z(\mathbf{z}, \varepsilon) =: \frac{1}{N_1(\varepsilon)}\exp\left(-\frac{1}{2}\mathbf{z}^T\mathbf{S}_P^{-1}(\varepsilon)\mathbf{z}\right)\frac{1}{N_2(\varepsilon)}G(\mathbf{z}), \tag{A.7}$$

where $G(\mathbf{z}) := \int \exp\big(-\frac{1}{2}G_0(\mathbf{z}, \mathbf{Z})\big)\,\mathrm{d}\mathbf{Z}$ with auxiliary term

$$G_0(\mathbf{z}, \mathbf{Z}) := \mathbf{Z}^T\mathbf{S}_L^{-1}\mathbf{Z} + \big(\mathbf{Y} - \widetilde{\mathbf{F}}(\mathbf{W})\big)^T\mathbf{S}_E^{-1}\big(\mathbf{Y} - \widetilde{\mathbf{F}}(\mathbf{W})\big). \tag{A.8}$$

Here, $N_1(\varepsilon) := \int \exp\big(-\frac{1}{2}\mathbf{z}^T\mathbf{S}_P^{-1}(\varepsilon)\mathbf{z}\big)\,\mathrm{d}\mathbf{z}$ is the normalization factor for the exponential term in Eq. (A.7) that is going to be singular in the limit, and $N_2(\varepsilon)$ is such that $\int P_z(\mathbf{z}, \varepsilon)\,\mathrm{d}\mathbf{z} = 1$. Note that $N_2(\varepsilon)$ does not normalize $G(\mathbf{z})$ in general, and that the normalization is split into two factors $N_1(\varepsilon)$ and $N_2(\varepsilon)$ in order to separate singular from regular terms to properly manage the limit functions.

In order to properly define the *a priori* covariance matrix for the retrieval of common parameters, $\lim_{\varepsilon\downarrow 0} P_z(\mathbf{z}, \varepsilon)$ has to be evaluated. Since $\mathbf{S}_P(\varepsilon)$ degenerates in the limit, $P_z(\cdot, \varepsilon)$ shall be regarded as tempered distribution. See Reed and Simon [59, Section V.3] for distribution theory. Let test function φ be an element of Schwartz space, the space of rapidly decreasing infinitely differentiable functions. Then $\lim_{\varepsilon\downarrow 0}\int P_z(\mathbf{z}, \varepsilon)\varphi(\mathbf{z})\,\mathrm{d}\mathbf{z}$ provides the *a posteriori* probability distribution in the sense of distribution theory. This term shall now be rearranged in order to compute the limit.

Let $\varepsilon > 0$. The first factor of $\mathbf{S}_P(\varepsilon) = \boldsymbol{\varrho}_P(\varepsilon) \otimes \mathbf{H}_C$, the matrix $\boldsymbol{\varrho}_P(\varepsilon)$, is real symmetric and can thus be diagonalized by an orthogonal matrix $\mathbf{Q}(\varepsilon)$ that consists of eigenvectors of $\boldsymbol{\varrho}_P(\varepsilon)$ [25, Theorem 4.1.5], i.e. $\boldsymbol{\varrho}_P(\varepsilon) = \mathbf{Q}(\varepsilon)\mathbf{D}(\varepsilon)\mathbf{Q}^T(\varepsilon)$, where the diagonal matrix $\mathbf{D}(\varepsilon)$ has the corresponding eigenvalues $D_{ii}(\varepsilon)$ on its diagonal. The eigenvalues are continuous functions of the entries of $\boldsymbol{\varrho}_P(\varepsilon)$ [60, Theorem 5.2]. Moreover, an algebraically simple eigenvalue is an analytic function of perturbations of the matrix entries, and so is the corresponding eigenvector [60, Theorem 5.3]. For eigenvalues that are not algebraically simple, the eigenvectors need not be continuous. The eigenvalues of $\lim_{\varepsilon\downarrow 0}\boldsymbol{\varrho}_P(\varepsilon) = \mathbf{1}_{r\times r}$ are r and 0, and r is an algebraically simple eigenvalue with eigenvector $(1, \cdots, 1)^T = \mathbf{1}_r$. Here, $\mathbf{1}_r \in \mathbb{R}^r$ is the vector with all entries 1. Therefore, and since $\boldsymbol{\varrho}_P(\varepsilon)$ continuously depends on ε, $\mathbf{D}(\varepsilon)$ and $\mathbf{Q}(\varepsilon)$ can be written as

$$\mathbf{D}(\varepsilon) = \begin{pmatrix} r + d_1(\varepsilon) & 0 & \cdots & 0 \\ 0 & d_2(\varepsilon) & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & d_r(\varepsilon) \end{pmatrix} \quad \text{and}$$

$$\mathbf{Q}(\varepsilon) = \left(\begin{array}{c|c} (1 + q_1(\varepsilon))\,/v(\varepsilon) & \\ (1 + q_2(\varepsilon))\,/v(\varepsilon) & \\ \vdots & * \\ (1 + q_r(\varepsilon))\,/v(\varepsilon) & \end{array}\right), \tag{A.9}$$

where $D_{ii}(\varepsilon) > 0$ for $\varepsilon > 0$ and $\lim_{\varepsilon\downarrow 0} d_i(\varepsilon) = 0$. Furthermore, $\lim_{\varepsilon\downarrow 0} q_i(\varepsilon) = 0$, and $v(\varepsilon)$ normalizes the first column of $\mathbf{Q}(\varepsilon)$, whence $\lim_{\varepsilon\downarrow 0} v(\varepsilon) = \sqrt{r}$. While v and all d_j and q_i are continuous in ε, the remaining columns $*$ of $\mathbf{Q}(\varepsilon)$ need not be continuous functions of ε. However, the absolute values of their entries are bounded from above by 1, since $\mathbf{Q}(\varepsilon)$ is orthogonal.

Let $\sqrt{\mathbf{D}^{-1}(\varepsilon)}$ be the diagonal matrix with the entries $\big(D_{ii}(\varepsilon)\big)^{-1/2}$ on its diagonal. Then by using the identity matrix $\mathbb{1}_{c\times c}$ of dimension $c\times c$, the substitution $\boldsymbol{\xi} := \big[\big(\sqrt{\mathbf{D}^{-1}(\varepsilon)}\mathbf{Q}^T(\varepsilon)\big)\otimes \mathbb{1}_{c\times c}\big]\mathbf{z}$ yields for $\int P_z(\mathbf{z},\varepsilon)\varphi(\mathbf{z})\,\mathrm{d}\mathbf{z}$

$$\int \frac{\big(\det \mathbf{D}(\varepsilon)\big)^{c/2}}{N_1(\varepsilon)} \exp\left(-\frac{1}{2} \boldsymbol{\xi}^T\big[\mathbb{1}_{r\times r}\otimes \mathbf{H}_C^{-1}\big]\boldsymbol{\xi}\right) \cdot \frac{1}{N_2(\varepsilon)} G\big(\boldsymbol{\psi}(\varepsilon,\boldsymbol{\xi})\big)\cdot\varphi\big(\boldsymbol{\psi}(\varepsilon,\boldsymbol{\xi})\big)\,\mathrm{d}\boldsymbol{\xi}, \qquad (A.10)$$

with $\boldsymbol{\psi}(\varepsilon,\boldsymbol{\xi}) = \big[\big(\mathbf{Q}(\varepsilon)\sqrt{\mathbf{D}(\varepsilon)}\,\big)\otimes\mathbb{1}_{c\times c}\big]\boldsymbol{\xi}$. This is a consequence of the change-of-variables formula $\int_{\boldsymbol{\psi}(V)} f(\mathbf{z})\,\mathrm{d}\mathbf{z} = \int_V f(\boldsymbol{\psi}(\boldsymbol{\xi}))|\det \mathbf{J}_{\boldsymbol{\psi}}(\boldsymbol{\xi})|\,\mathrm{d}\boldsymbol{\xi}$ [61, Theorems 8.26, 8.28] and due to

$$\big|\det\big[\big(\mathbf{Q}(\varepsilon)\sqrt{\mathbf{D}(\varepsilon)}\,\big)\otimes\mathbb{1}_{c\times c}\big]\big| = \big|\det\big(\mathbf{Q}(\varepsilon)\sqrt{\mathbf{D}(\varepsilon)}\,\big)\big|^c = \big(\det\mathbf{D}(\varepsilon)\big)^{c/2},$$

which holds because of $\det(\mathbf{A}\otimes\mathbf{B}) = (\det\mathbf{A})^n(\det\mathbf{B})^m$ for matrices $\mathbf{A}$ of dimension $m\times m$ and $\mathbf{B}$ of dimension $n\times n$ [62, Section 2.3, X] and $\det\mathbf{Q}(\varepsilon)=1$.

As the $*$ in Eq. (A.9) are bounded, it follows that

$$\lim_{\varepsilon\downarrow0}\mathbf{Q}(\varepsilon)\sqrt{\mathbf{D}(\varepsilon)} = \begin{pmatrix}1&0&\cdots&0\\ \vdots&\vdots&&\vdots\\ 1&0&\cdots&0\end{pmatrix}\in\mathbb{R}^{r\times r},$$

and therefore

$$\lim_{\varepsilon\downarrow0}\boldsymbol{\psi}(\varepsilon,\boldsymbol{\xi}) = \big(\boldsymbol{\xi}_1,\cdots,\boldsymbol{\xi}_1\big) = \mathbf{1}_r\otimes\boldsymbol{\xi}_1\in\mathbb{R}^{rc}.$$

To now compute $\lim_{\varepsilon\downarrow0}\int P_z(\mathbf{z},\varepsilon)\varphi(\mathbf{z})\,\mathrm{d}\mathbf{z}$, Lebesgue's Dominated Convergence Theorem [61, Theorem 1.34] shall be applied to evaluate $\lim_{\varepsilon\downarrow0}$ of Eq. (A.10) by interchanging limit and integral. Thereto, it shall be checked whether the Theorem's assumptions apply (i.e. whether for the sequence of functions that point-wise converges to the limit function under the integral, there exists an integrable dominating function).

First, observe that Eq. (A.8) implies that G is bounded, since it can be written as $\int\exp\big(-\frac{1}{2}\mathbf{Z}^T\mathbf{S}_L^{-1}\mathbf{Z}\big)K(\mathbf{z},\mathbf{Z})\,\mathrm{d}\mathbf{Z}$, with K bounded by 1 and the positive definite $\mathbf{S}_L^{-1}$ providing an exponential damping for the integration. Here, K is bounded and continuous, since $\mathbf{S}_E^{-1}$ is positive definite and the forward model outcome $\mathbf{F}$ is continuous (see Section 2), and therefore also $\widetilde{\mathbf{F}}$. Furthermore, $|G(\mathbf{z})-G(\widetilde{\mathbf{z}})| \leq \int\exp\big(-\frac{1}{2}\mathbf{Z}^T\mathbf{S}_L^{-1}\mathbf{Z}\big)\cdot 2\,\mathrm{d}\mathbf{Z}$, providing an integrable dominating function for the verification that G is continuous because K is, by applying Lebesgue's Dominated Convergence Theorem to $|G(\mathbf{z})-G(\widetilde{\mathbf{z}})|$.

Next, the properties of $N_1(\varepsilon)$ and $N_2(\varepsilon)$ as normalizing factors in Eq. (A.7) for small (i.e. Eq. (A.9) holds) $\varepsilon\geq0$ shall be investigated.

By using the same substitution that was used to arrive at Eq. (A.10), one obtains

$$N_1(\varepsilon) = (2\pi)^{rc/2}\big(\det\mathbf{H}_C\big)^{r/2}\big(\det\mathbf{D}(\varepsilon)\big)^{c/2},$$

since $\int\exp\big(-\frac{1}{2}\mathbf{x}^T\mathbf{A}\mathbf{x}\big)\,\mathrm{d}\mathbf{x} = (2\pi)^{n/2}/\sqrt{\det\mathbf{A}}$ for a real symmetric positive definite matrix $\mathbf{A}$ of dimension $n\times n$.

By the same substitution, N_2 can be written as

$$N_2(\varepsilon) = (2\pi)^{-rc/2}\big(\det\mathbf{H}_C\big)^{-r/2} \cdot\int\exp\left(-\frac{1}{2}\boldsymbol{\xi}^T\big[\mathbb{1}_{r\times r}\otimes\mathbf{H}_C^{-1}\big]\boldsymbol{\xi}\right)G\big(\boldsymbol{\psi}(\varepsilon,\boldsymbol{\xi})\big)\,\mathrm{d}\boldsymbol{\xi}. \qquad (A.11)$$

The arguments that yielded boundedness and continuity of G, carry over to the verification that N_2 is in ε continuous at $\varepsilon=0$, because $\mathbf{Q}(\varepsilon)\sqrt{\mathbf{D}(\varepsilon)}$ and hence $\boldsymbol{\psi}(\varepsilon,\boldsymbol{\xi})$ are. As G is bounded, there exists some finite $\widetilde{K}>0$ such that for all $\varepsilon\geq0$, $\widetilde{K}\exp\big(-\frac{1}{2}\boldsymbol{\xi}^T\big[\mathbb{1}_{r\times r}\otimes\mathbf{H}_C^{-1}\big]\boldsymbol{\xi}\big)$ is an integrable dominating function in $\boldsymbol{\xi}$ for the integrand in Eq. (A.11). Thus Lebesgue's Dominated Convergence Theorem can be applied to Eq. (A.11) to obtain $N_2(0) := \lim_{\varepsilon\downarrow0}N_2(\varepsilon)$.

$$N_2(0) = (2\pi)^{-rc/2}\big(\det\mathbf{H}_C\big)^{-r/2} \cdot\int\exp\left(-\frac{1}{2}\boldsymbol{\xi}^T\big[\mathbb{1}_{r\times r}\otimes\mathbf{H}_C^{-1}\big]\boldsymbol{\xi}\right)G\big(\mathbf{1}_r\otimes\boldsymbol{\xi}_1\big)\,\mathrm{d}\boldsymbol{\xi},$$

as G is continuous and $\boldsymbol{\psi}(\varepsilon,\boldsymbol{\xi})$ converges point-wise to $\mathbf{1}_r\otimes\boldsymbol{\xi}_1$ for $\varepsilon\downarrow0$. For generic $\mathbf{Y}$ and positive definite $\mathbf{S}_L$ and $\mathbf{S}_E$, $G(\mathbf{z})$ is positive for finite $\widetilde{\mathbf{F}}$, see Eq. (A.8). But for finite arguments, $\widetilde{\mathbf{F}}$ should be finite (Section 2). Hence, for small $\varepsilon\geq0$, $N_2(\varepsilon)$ is some finite positive number.

Lebesgue's Dominated Convergence Theorem can now be applied to evaluate $\lim_{\varepsilon\downarrow0}$ of Eq. (A.10). As a test function in Schwartz space, φ is continuous and bounded, as is G. $N_2(\varepsilon)$ is positive for small $\varepsilon\geq0$, finite, and continuous in ε at $\varepsilon=0$. Thus, for some finite $\overline{K}>0$, $\overline{K}\exp\big(-\frac{1}{2}\boldsymbol{\xi}^T\big[\mathbb{1}_{r\times r}\otimes\mathbf{H}_C^{-1}\big]\boldsymbol{\xi}\big)$ is an integrable dominating function in $\boldsymbol{\xi}$ for the integrand for small $\varepsilon\geq0$, and by applying the theorem, it follows that $\lim_{\varepsilon\downarrow0}\int P_z(\mathbf{z},\varepsilon)\varphi(\mathbf{z})\,\mathrm{d}\mathbf{z}$ is equal to

$$(2\pi)^{-rc/2}\big(\det\mathbf{H}_C\big)^{-r/2}\cdot\frac{1}{N_2(0)}\cdot \int\exp\left(-\frac{1}{2}\boldsymbol{\xi}^T\big[\mathbb{1}_{r\times r}\otimes\mathbf{H}_C^{-1}\big]\boldsymbol{\xi}\right)G\big(\mathbf{1}_r\otimes\boldsymbol{\xi}_1\big)\varphi\big(\mathbf{1}_r\otimes\boldsymbol{\xi}_1\big)\,\mathrm{d}\boldsymbol{\xi}.$$

Since $\boldsymbol{\xi}^T\big[\mathbb{1}_{r\times r}\otimes\mathbf{H}_C^{-1}\big]\boldsymbol{\xi} = \sum_{i=1}^r\boldsymbol{\xi}_i^T\mathbf{H}_C^{-1}\boldsymbol{\xi}_i\in\mathbb{R}$, evaluating the integral over $\mathrm{d}\boldsymbol{\xi}_2\cdots\mathrm{d}\boldsymbol{\xi}_r$ and absorbing the uninteresting terms into the constant $\widetilde{N}$, yields

$$\frac{1}{\widetilde{N}}\int\exp\left(-\frac{1}{2}\boldsymbol{\xi}_1^T\mathbf{H}_C^{-1}\boldsymbol{\xi}_1\right)G\big(\mathbf{1}_r\otimes\boldsymbol{\xi}_1\big)\varphi\big(\mathbf{1}_r\otimes\boldsymbol{\xi}_1\big)\,\mathrm{d}\boldsymbol{\xi}_1.$$

This term can be rewritten by using the c-dimensional δ-

distribution $\left(\int \delta^c(\mathbf{t})f(\mathbf{t})\,\mathrm{d}\mathbf{t} = f(\mathbf{0}) \text{ for } \mathbf{t} \in \mathbb{R}^c\right)$, such that

$$\lim_{\varepsilon \downarrow 0} \int P_z(\mathbf{z}, \varepsilon)\varphi(\mathbf{z})\,\mathrm{d}\mathbf{z} = \frac{1}{\widetilde{N}} \int \delta^c(\mathbf{z}_1 - \mathbf{z}_2) \cdots \delta^c(\mathbf{z}_1 - \mathbf{z}_r) \cdot \exp\left(-\frac{1}{2}\mathbf{z}_1^T \mathbf{H}_C^{-1} \mathbf{z}_1\right) G(\mathbf{1}_r \otimes \mathbf{z}_1)\varphi(\mathbf{z})\,\mathrm{d}\mathbf{z}.$$

This equation shows that the *a posteriori* probability distribution for $\mathbf{z}$ concentrates to the plane $\mathbf{1}_r \otimes \mathbf{z}_1$ (all $\mathbf{z}_i$ coincide). This means that $\mathbf{z}_1$ is spatially or temporally constant, i.e. it is common to the r measurements. For the whole space (including the $\mathbf{Z}$-dimensions), the induced probability distribution on that plane can be written $\frac{1}{\widetilde{N}} \exp\left(-\frac{1}{2}\widetilde{F}_c\right)$, with cost function $\widetilde{F}_c(\mathbf{z}, \mathbf{Z})$ that reads

$$\mathbf{z}_1^T \mathbf{H}_C^{-1} \mathbf{z}_1 + \mathbf{Z}^T \mathbf{S}_L^{-1} \mathbf{Z} + \big(\mathbf{Y} - \widetilde{\mathbf{F}}(\mathbf{1}_r \otimes \mathbf{z}_1, \mathbf{Z})\big)^T \mathbf{S}_E^{-1} \big(\mathbf{Y} - \widetilde{\mathbf{F}}(\mathbf{1}_r \otimes \mathbf{z}_1, \mathbf{Z})\big). \quad \text{(A.12)}$$

Define the new extended parameter space as the dimensionally reduced parameter space on that plane, i.e. $(\mathbf{1}_r \otimes \mathbf{z}_1, \mathbf{Z})$ is dimensionally reduced to obtain $(\mathbf{z}_1, \mathbf{Z})$. Rename the vector of the common parameters $\mathbf{z}_1 =: \mathbf{z}_C$ and its *a priori* covariance matrix $\mathbf{H}_C =: \mathbf{S}_C$. Here and in the following, the subscript 'C' is intended to flag quantities that involve common parameters. Revert the translational substitutions with respect to the *a priori* mean values at the beginning of this section, and denote the vector of the parameters associated to $\mathbf{S}_L$ by $\mathbf{X}_L := (\mathbf{x}_1, \cdots, \mathbf{x}_r)$, along with its *a priori* mean value vector $\mathbf{A}_L := (\mathbf{a}, \cdots, \mathbf{a})$, such that $\mathbf{Z} = \mathbf{X}_L - \mathbf{A}_L$, and similarly with $\mathbf{z}_C =: \mathbf{x}_C - \mathbf{a}_C$, where $\mathbf{a}_C := \mathbf{a}_P$. The extended parameter vector shall be denoted by $\mathbf{X} := (\mathbf{x}_C, \mathbf{X}_L)$, and its *a priori* mean value vector by $\mathbf{A} := (\mathbf{a}_C, \mathbf{A}_L)$, leading to the same notation as in Section 2. $\widetilde{\mathbf{F}}(\mathbf{1}_r \otimes \mathbf{z}_1, \mathbf{Z})$ can be written as $\mathbf{F}\big((\mathbf{x}_C, \mathbf{x}_1), \cdots, (\mathbf{x}_C, \mathbf{x}_r)\big) = \big(\mathbf{f}_1(\mathbf{x}_C, \mathbf{x}_1), \cdots, \mathbf{f}_r(\mathbf{x}_C, \mathbf{x}_r)\big)$ and will be denoted by $\mathbf{F}_C(\mathbf{X})$.

Thus, finally the *a posteriori* probability distribution can be written $\frac{1}{N} \exp\left(-\frac{1}{2}F_c\right)$, with normalization factor N and cost function $F_c(\mathbf{X})$ that reads

$$(\mathbf{x}_C - \mathbf{a}_C)^T \mathbf{S}_C^{-1} (\mathbf{x}_C - \mathbf{a}_C) + (\mathbf{X}_L - \mathbf{A}_L)^T \mathbf{S}_L^{-1} (\mathbf{X}_L - \mathbf{A}_L) + \big(\mathbf{Y} - \mathbf{F}_C(\mathbf{X})\big)^T \mathbf{S}_E^{-1} \big(\mathbf{Y} - \mathbf{F}_C(\mathbf{X})\big), \quad \text{(A.13)}$$

and which has to be minimized by the retrieval. This shows that the relative weighting between the common and the local parameters is exactly 1, and not some $\sqrt{r}$-like value. Also, it follows that spatially or temporally perfectly coupled parameters can be treated as common parameters in the sense of Section 2, where $\mathbf{S}_A$ is a block diagonal matrix with $\mathbf{S}_C$ and $\mathbf{S}_L$ as its blocks on the diagonal.

For VIRTIS-M-IR measurements of Venus, some relevant common parameters along with a suitable $\mathbf{S}_C$ are discussed in Section 4.3.

A.4. Sparse matrix formulation

Most entries of the largest structure in MSR, the Jacobian $\mathbf{J}$ (Eq. (A.2)), are zero. This can be exploited by using sparse matrix storage formats that only store the non-zero entries and their positions [63]. The 'coordinate list' format (COO) requires 16 bytes of storage space for each non-zero entry (4 for integer row index, 4 for integer column index, 8 for value in double precision). The 'compressed column' format (CCS) requires about 12 bytes per entry (4 for row index, 8 for value) for matrices with many more non-zero entries than the number of columns (pointer to where in the value list the columns start, is then neglectable). A matrix stored in 'dense' format as $M \times N$ array, requires 8 bytes per entry. Matrix creation is convenient with COO, matrix computations are efficient with CCS. Sparse matrix formulation of MSR saves computer memory (use sparse matrix storage) and processing time (apply sparse matrix operations), and for VIRTIS data, it allows to treat retrieval problems that are larger by one order of magnitude.

Sparse matrices can be manipulated by using SuiteSparse, a suite of sparse matrix routines that include format conversion and QR factorization [64–73]. SuiteSparse relies on METIS [74].

As an example that is on current standard desktop hardware easily treatable in sparse format, let there be $r = 1{,}000$ spectra of an effective size of 100 wavelengths each, $c = 100$ common retrieval parameters (surface emissivities in a surface patch of 10×10 bins) and $n = 10$ local retrieval parameters per spectrum. Then $\mathbf{J}$ is of size $110{,}100 \times 10{,}100$, corresponding to about 10^9 (dense) entries, and can be compiled as follows, see Eq. (A.2) and Section 3.5.

$\mathbf{S}_C^{-1/2}$ as a small dense matrix (100×100 for the example) is created in dense format according to Section 4.3 and then converted to CCS.

$\mathbf{S}_L^{-1/2}$ ($10{,}000 \times 10{,}000$ for the example) as given by Eq. (A.6) is ideally suited to be directly created in sparse format. First, the dense $\mathbf{H}_g$ (Appendix A.2) and $\boldsymbol{\varrho}_g$ have to be computed (Sections 3.3 and 3.2.3) for the G parameter groups g, yielding the dense $\mathbf{U}_{\mathbf{H}_g}^{-T}$ and $\mathbf{U}_{\boldsymbol{\varrho}_g}^{-T}$. This is efficient due to the small sizes of these matrices ($n_g \times n_g$ where $\sum_{g=1}^{G} n_g = n$, and $r \times r$). The Kronecker products $\mathbf{U}_{\mathbf{H}_g}^{-T} \otimes \mathbf{U}_{\boldsymbol{\varrho}_g}^{-T}$ are computed to yield COO matrices. These G sub-matrices are combined and permuted according to Eq. (A.6) in COO representation and converted to CCS. At no point, the dense representation of the full matrix is needed. Note that, the sparser the $\mathbf{H}_g$ or $\boldsymbol{\varrho}_g$ are, the sparser $\mathbf{S}_L^{-1/2}$ tends to be, but in general the inverse of a sparse matrix needs not to be sparse anymore. The worst case relative population of $\mathbf{S}_L^{-1/2}$ is about $\frac{1}{2}\sum_{g=1}^{G} n_g^2/n^2$ for large r and n.

$\mathbf{J}$ in CCS format can now be assembled column-wise by combining $\mathbf{S}_C^{-1/2}$ and $\mathbf{S}_L^{-1/2}$, and by collecting the columns of the sparse $-\mathbf{S}_E^{-1/2}\mathbf{K}$. In the worst case ($G = 1$, maximal population of all single-spectrum Jacobians), $\mathbf{J}$ is about 3.5% populated in the example above, translating to about 5.2% required storage space compared to dense format.

The single most expensive matrix operation in MSR is the QR factorization of $\mathbf{J}$ as needed in the trust region algorithm [56]. A sparse $\mathbf{J} = \mathbf{Q}\mathbf{R}\mathbf{P}^T$ can be factorized by SuiteSparse into an orthogonal matrix $\mathbf{Q}$ and an upper triangular matrix $\mathbf{R}$ [57, Section 5.2], [64], where $\mathbf{P}$ is a permutation matrix that can lead to fewer non-zero entries of the sparse $\mathbf{R}$, and $\mathbf{Q}$ needs not to be stored. The actual determination of the Levenberg-Marquardt step [56, Section 3] is performed as an additional sparse QR factorization.

Appendix A.1 discusses the interpretation of the covariance matrix $\widehat{\mathbf{S}}$ (Eq. (A.3)) and the correlation matrix $\widehat{\mathbf{C}}$ at the retrieved result. Only the diagonal and a few off-diagonal entries $\widehat{S}_{ij}$ of $\widehat{\mathbf{S}}$ may be needed, and the corresponding entries of $\widehat{\mathbf{C}}$ follow from Eq. (3). In the last retrieval iteration, $\mathbf{J}$ is evaluated at $\widehat{\mathbf{X}}$ to yield $\widehat{\mathbf{J}}$ with sparse QR factorization $\widehat{\mathbf{J}} = \widehat{\mathbf{Q}}\widehat{\mathbf{R}}\widehat{\mathbf{P}}^T$. $\widehat{\mathbf{J}}^T\widehat{\mathbf{J}} = \widehat{\mathbf{P}}\widehat{\mathbf{R}}^T\widehat{\mathbf{R}}\widehat{\mathbf{P}}^T$ immediately (neglecting the structurally empty lines to yield a $N \times N$ matrix from $\widehat{\mathbf{R}}\widehat{\mathbf{P}}^T$) provides a permuted Cholesky decomposition of $\widehat{\mathbf{S}}^{-1} = \widehat{\mathbf{J}}^T\widehat{\mathbf{J}}$ which is needed to efficiently invert $\widehat{\mathbf{S}}^{-1}$. But the inverse of a sparse matrix needs not to be sparse, and it may not be possible to store all of its entries. To avoid the costly computation of $\widehat{\mathbf{J}}^T\widehat{\mathbf{J}}$ and its inversion via Cholesky decomposition, note that

$$\begin{aligned}\widehat{S}_{ij} = \langle \mathbf{e}_i, \widehat{\mathbf{S}}\mathbf{e}_j \rangle &= \langle \mathbf{e}_i, \widehat{\mathbf{P}}\widehat{\mathbf{R}}^{-1}\widehat{\mathbf{R}}^{-T}\widehat{\mathbf{P}}^T\mathbf{e}_j \rangle \\ &= \langle \widehat{\mathbf{R}}^{-T}\widehat{\mathbf{P}}^T\mathbf{e}_i, \widehat{\mathbf{R}}^{-T}\widehat{\mathbf{P}}^T\mathbf{e}_j \rangle =: \langle \mathbf{z}_i, \mathbf{z}_j \rangle.\end{aligned}$$

with the standard basis vectors $\mathbf{e}_i$ and Euclidean standard scalar product $\langle \cdot, \cdot \rangle$. $\mathbf{z}_i$ can be determined by solving the sparse linear equation $\widehat{\mathbf{R}}^T\mathbf{z}_i = \widehat{\mathbf{P}}^T\mathbf{e}_i$.

A.5. Further notes on implementation

The retrieval of parameters whose impacts on the simulated spectra can be separated, may be arranged into several stages. Each stage corresponds to a complete run of the retrieval algorithm in order to determine the best estimate for the corresponding retrieval parameter subset by considering a suitable stage-specific spectral range. The *a priori* and error covariance matrices have to be newly constructed for each stage, as has the covariance matrix at the retrieved result. The retrieved values from earlier stages can either be used as fixed input values with now known uncertainties, or as initial guesses for a refined determination of parameters from earlier stages. Such refinements might be necessary, since the adjustment of additional parameters and the inclusion of different wavelength ranges can cause previously retrieved parameters to become suboptimal. This partitioning into stages results in Jacobians of smaller dimensions and consequently decreased maximal computer memory usage, and also processing time that tends to increase faster than linear with problem size. A tight choice of parameters and wavelength ranges for each retrieval stage can decide the processable size of the retrieval problem.

In order to take advantage of multi-core computer hardware, the program is parallelized by using the Message Passing Interface (MPI) [75] as implemented, for instance in OpenMPI [76]. As computer memory is a limiting factor for the processable problem size, only one of the parallel processes has all information needed for the retrieval algorithm, including the *a priori* covariance matrix and the Jacobian, and manages and performs the retrieval iterations. The remaining processes act as co-processors that evaluate their share of the single-spectrum simulations and single-spectrum Jacobians and communicate them via MPI to the main process.

Acknowledgements

This work is funded by the German Aerospace Center e.V. (DLR e.V.).

[1] Avduevskii VS, Marov MI, Kulikov IN, Shari VP, Gorbachevskii AI, Uspenskii GR, et al. Structure and parameters of the Venus atmosphere according to Venera probe data. In: Hunten, D. M., Colin, L., Donahue, T. M., & Moroz, V. I. , editor. Venus. University of Arizona Press; 1983, p. 280–98.

[2] Pettengill GH, Ford PG, Johnson WTK, Raney RK, Soderblom LA. Magellan: Radar Performance and Data Products. Science 1991;252(5003):260–5. doi:\bibinfo{doi}{10.1126/science.252.5003.260}.

[3] Seiff A, Schofield J, Kliore A, Taylor F, Limaye S, Revercomb H, et al. Models of the structure of the atmosphere of Venus from the surface to 100 kilometers altitude. Advances in Space Research 1985;5(11):3 – 58. doi:\bibinfo{doi}{10.1016/0273-1177(85)90197-8}.

[4] Zasova L, Moroz V, Linkin V, Khatuntsev I, Maiorov B. Structure of the Venusian atmosphere from surface up to 100 km. Cosmic Research 2006;44(4):364–83. doi:\bibinfo{doi}{10.1134/S0010952506040095}.

[5] Pollack JB, Dalton JB, Grinspoon D, Wattson RB, Freedman R, Crisp D, et al. Near-Infrared Light from Venus' Nightside: A Spectroscopic Analysis. Icarus 1993;103(1):1 – 42. doi:\bibinfo{doi}{10.1006/icar.1993.1055}.

[6] Meadows VS, Crisp D. Ground-based near-infrared observations of the Venus nightside: The thermal structure and water abundance near the surface. J Geophys Res 1996;101(E2):4595–622. doi:\bibinfo{doi}{10.1029/95JE03567}.

[7] Drossart P, Piccioni G, Adriani A, Angrilli F, Arnold G, Baines K, et al. Scientific goals for the observation of Venus by VIRTIS on ESA/Venus express mission. Planetary and Space Science 2007;55(12):1653 –72. doi:\bibinfo{doi}{10.1016/j.pss.2007.01.003}.

[8] Arnold GE, Drossart P, Piccioni G, Haus R. Venus atmospheric and surface studies from VIRTIS on Venus Express. In: Infrared Remote Sensing and Instrumentation XIX; vol. 8154. 2011, p. 81540W–81540W–17. doi:\bibinfo{doi}{10.1117/12.892895}.

[9] Arnold GE, Haus R, Kappel D, Piccioni G, Drossart P. VIRTIS/VEX observations of Venus: overview of selected scientific results. Journal of Applied Remote Sensing 2012;6(1):063580–1–063580–20. doi:\bibinfo{doi}{10.1117/1.JRS.6.063580}.

[10] Cardesin Moinelo A, Piccioni G, Ammannito E, Filacchione G, Drossart P. Calibration of Hyperspectral Imaging Data: VIRTIS-M Onboard Venus Express. IEEE Transactions on Geoscience and Remote Sensing 2010;48:3941–50. doi:\bibinfo{doi}{10.1109/TGRS.2010.2064325}.

[11] Kappel D, Arnold G, Haus R, Piccioni G, Drossart P. Refinements in the data analysis of VIRTIS-M-IR Venus nightside spectra. Advances in Space Research 2012;50(2):228 –55. doi:\bibinfo{doi}{10.1016/j.asr.2012.03.029}.

[12] Haus R, Arnold G. Radiative transfer in the atmosphere of Venus and application to surface emissivity retrieval from VIRTIS/VEX measurements. Planetary and Space Science 2010;58(12):1578 –98. doi:\bibinfo{doi}{10.1016/j.pss.2010.08.001}.

[13] Tikhonov A. Numerical Methods for the Solution of Ill-Posed Problems. Mathematics and Its Applications; Kluwer Academic Publishers; 1995. ISBN 9780792335832.

[14] Marcq E, Encrenaz T, Bézard B, Birlan M. Remote sensing of Venus' lower atmosphere from ground-based IR spectroscopy: Latitudinal and vertical distribution of minor species. Planetary and Space Science 2006;54(13-14):1360 –70. doi:\bibinfo{doi}{10.1016/j.pss.2006.04.024}.

[15] de Bergh C, Bézard B, Crisp D, Maillard JP, Owen T, Pollack J, et al. Water in the deep atmosphere of Venus from high-resolution spectra of the night side. Advances in Space Research 1995;15(4):79 – 88. doi:\bibinfo{doi}{10.1016/0273-1177(94)00067-B}.

[16] de Kok R, Irwin P, Tsang C, Piccioni G, Drossart P. Scattering particles in nightside limb observations of Venus' upper atmosphere by Venus Express VIRTIS. Icarus 2011;211(1):51 –7. doi:\bibinfo{doi}{10.1016/j.icarus.2010.08.023}.

[17] Tellmann S, Pätzold M, Häusler B, Bird MK, Tyler GL. Structure of the Venus neutral atmosphere as observed by the Radio Science experiment VeRa on Venus Express. J Geophys Res 2009;114(E9):E00B36. doi:\bibinfo{doi}{10.1029/2008JE003204}.

[18] Rodgers CD. Inverse Methods for Atmospheric Sounding : Theory and Practice (Series on Atmospheric, Oceanic and Planetary Physics). World Scientific Publishing Company; 2000. ISBN 981022740X.

[19] Kappel D, Arnold G, Haus R. Multispectrum retrieval techniques applied to Venus deep atmosphere and surface problems. In: 38st COSPAR Scientific Assembly. 2010,C33-0008-10.

[20] Kappel D, Arnold G, Haus R, Piccioni G, Drossart P. Results from multispectrum retrieval of VIRTIS-M-IR measurements of Venus' nightside. In: EPSC Abstracts Vol. 5, European Planetary Science Congress 2010. 2010,EPSC2010-390.

[21] Kappel D, Arnold G, Haus R. Retrieval of Surface Emissivity in a Venus Coordinate Patch as Parameter Common to Repeated Measurements by VIRTIS/VEX. In: EGU General Assembly 2012; vol. 14. 2012, p. 9708–.

[22] Spurr R. LIDORT and VLIDORT: Linearized pseudo-spherical scalar and vector discrete ordinate radiative transfer models for use in remote sensing retrieval problems. vol. 3 of *Light Scattering Reviews*. Springer; 2008,.

[23] Kappel D, Arnold G, Haus R. Sensitivity of Venus surface emissivity retrieval to model variations of CO2 opacity, cloud features, and deep atmosphere temperature field. In: 39th COSPAR Scientific Assembly. 2012,B0.8-0008-12.

[24] Anderson T. An introduction to multivariate statistical analysis. Wiley publications in statistics; Wiley; 1958.

[25] Horn RA, Johnson CR. Matrix Analysis. Cambridge University Press; 1990. ISBN 0521386322.

[26] Schoenberg IJ. Selected papers. Ed C de Boor 1988;(1).

[27] Baxter B. Positive definite functions on hilbert space. 2002.

[28] Miller KS, Samko SG. Completely monotonic functions. Integral Transforms and Special Functions 2001;12(4):389–402.

[29] Daley R. Atmospheric data analysis (Cambridge Atmospheric and Space Science Series). World Scientific Publishing Company; 1992. ISBN 0521 382157. doi:\bibinfo{doi}{10.1002/joc.3370120708}.

[30] Balgovind R, Dalcher A, Ghil M, Kalnay E. A stochastic-dynamic model for the spectral structure of forecast errors. Mon Wea Rev 1983;111:701–21.

[31] Gelb A. Applied Optimal Estimation. Cambridge, MA., USA: MIT Press; 1974.

[32] Rood RB, Gaspari G, Cohn SE. Construction of Correlation Functions in Two and Three Dimensions. 1996.

[33] Reed M, Simon B. II: Fourier analysis, self-adjointness. Methods of Modern Mathematical Physics; Academic Press; 1975. ISBN 9780125850025.

[34] Horn RA, Johnson CR. Topics in Matrix Analysis. Cambridge: Cambridge University Press; 1991.

[35] Stamnes K, Tsay SC, Wiscombe W, Jayaweera K. Numerically stable algorithm for discrete-ordinate-method radiative transfer in multiple scattering and emitting layered media. Appl Opt 1988;27(12):2502–9. doi:\bibinfo{doi}{10.1364/AO.27.002502}.

[36] Spurr RJ. Linearized radiative transfer theory : a general discrete ordinate approach to the calculation of radiances and analytic weighting functions , with application to atmospheric remote sensing. Ph.D. thesis; 2001. URL http://library.tue.nl/csp/dare/LinkToRepository.csp?recordnumber=545442.

[37] Marov M, Lystsev V, Lebedev V, Lukashevich N, Shari V. The structure and microphysical properties of the Venus clouds: Venera 9, 10, and 11 data. Icarus 1980;44(3):608 –39. doi:\bibinfo{doi}{10.1016/0019-1035(80)90131-1}.

[38] Tashkun SA, Perevalov VI, Teffo JL, Bykov AD, Lavrentieva NN. CDSD-1000, the high-temperature carbon dioxide spectroscopic databank. Journal of Quantitative Spectroscopy and Radiative Transfer 2003;82(1-4):165 –96. doi:\bibinfo{doi}{10.1016/S0022-4073(03)00152-3}.

[39] Rothman LS, Wattson RB, Gamache R, Schroeder JW, McCann A. HITRAN HAWKS and HITEMP: high-temperature molecular database. In: Dainty JC, editor. Atmospheric Propagation and Remote Sensing IV; vol. 2471. 1995, p. 105–11. doi:\bibinfo{doi}{10.1117/12.211919}.

[40] Rothman L, Gordon I, Barber R, Dothe H, Gamache R, Goldman A, et al. HITEMP, the high-temperature molecular spectroscopic database. Journal of Quantitative Spectroscopy and Radiative Transfer 2010;111(15):2139 –50. doi:\bibinfo{doi}{http://dx.doi.org/10.1016/j.jqsrt.2010.05.001}.

[41] Rothman L, Gordon I, Barbe A, Benner D, Bernath P, Birk M, et al. The HITRAN 2008 molecular spectroscopic database. Journal of Quantitative Spectroscopy and Radiative Transfer 2009;110(9 - 10):533 –72. doi:\bibinfo{doi}{10.1016/j.jqsrt.2009.02.013}.

[42] Hansen JE, Travis LD. Light scattering in planetary atmospheres. Space Science Reviews 1974;16(4):527–610. doi:\bibinfo{doi}{10.1007/BF00168069}.

[43] Piccioni G, Zasova L, Migliorini A, Drossart P, Shakun A, Garcia Munoz A, et al. Near-IR oxygen nightglow observed by VIRTIS in the Venus upper atmosphere. Journal of Geophysical Research: Planets 2009;114(E5). doi:\bibinfo{doi}{10.1029/2008JE003133}.

[44] Bézard B, Fedorova A, Bertaux JL, Rodin A, Korablev O. The 1.10- and 1.18-µm nightside windows of Venus observed by SPICAV-IR aboard Venus Express. Icarus 2011;216(1):173 –83. doi:\bibinfo{doi}{10.1016/j.icarus.2011.08.025}.

[45] Palmer KF, Williams D. Optical Constants of Sulfuric Acid; Application to the Clouds of Venus? Appl Opt 1975;14(1):208–19. doi:\bibinfo{doi}{10.1364/AO.14.000208}.

[46] Carlson R, Anderson M. Absorption properties of sulfuric acid in Venus's infrared spectral windows region. In: EPSC-DPS Joint Meeting 2011. 2011, p. 1171.

[47] Haus R, Kappel D, Arnold G. Self-consistent retrieval of temperature profiles and cloud structure in the northern hemisphere of Venus using VIRTIS/VEX and PMV/VENERA-15 radiation measurements. submitted to Planetary and Space Science 2013;.

[48] Wiscombe WJ. Improved Mie scattering algorithms. Appl Opt 1980;19(9):1505–9. doi:\bibinfo{doi}{10.1364/AO.19.001505}.

[49] Tsang C, Taylor F, Wilson C, Liddell S, Irwin P, Piccioni G, et al. Variability of CO concentrations in the Venus troposphere from Venus Express/VIRTIS using a Band Ratio Technique. Icarus 2009;201(2):432 –43. doi:\bibinfo{doi}{10.1016/j.icarus.2009.01.001}.

[50] Müller N, Helbert J, Stofan ER, Smrekar S, Piccioni G, Drossart P. Search for active lava flows with VIRTIS on Venus Express. In: American Geophysical Union, 45th Fall Meeting. 2012, p. P24B–02.

[51] Moroz VI. Estimates of visibility of the surface of Venus from descent probes and balloons. Planetary and Space Science 2002;50(3):287 –97. doi:\bibinfo{doi}{10.1016/S0032-0633(01)00128-3}.

[52] Lebonnois S, Hourdin F, Forget F, Eymet V, Fournier R. The LMD Venus General Circulation Model: Improvements and Questions. 2010,URL http://lesia.obspm.fr/venus2010/IMG/pdf/04-04_Lebonnois.pdf.

[53] Arnold G, Haus R, Kappel D, Drossart P, Piccioni G. Venus surface data extraction from VIRTIS/Venus Express measurements: Estimation of a quantitative approach. J Geophys Res 2008;113(E5):E00B10. doi:\bibinfo{doi}{10.1029/2008JE003087}.

[54] Müller N, Helbert J, Hashimoto GL, Tsang CCC, Erard S, Piccioni G, et al. Venus surface thermal emission at 1 µm in VIRTIS imaging observations: Evidence for variation of crust and mantle differentiation conditions. J Geophys Res 2008;113(E5):E00B17. doi:\bibinfo{doi}{10.1029/2008JE003118}.

[55] Smrekar SE, Stofan ER, Müller N, Treiman A, Elkins-Tanton L, Helbert J, et al. Recent Hotspot Volcanism on Venus from VIRTIS Emissivity Data. Science 2010;328(5978):605–8. doi:\bibinfo{doi}{10.1126/science.1186785}. http://www.sciencemag.org/content/328/5978/605.full.pdf.

[56] Moré J. The Levenberg-Marquardt algorithm: Implementation and theory. In: Watson GA, editor. Numerical Analysis; vol. 630 of *Lecture Notes in Mathematics*; chap. 10. Springer Berlin Heidelberg. ISBN 978-3-540-08538-6; 1978, p. 105–16. doi:\bibinfo{doi}{10.1007/BFb0067700}.

[57] Golub G, van Loan C. Matrix computations. Johns Hopkins studies in the mathematical sciences; Johns Hopkins University Press; 1996. ISBN 9780801854149.

[58] Nocedal J, Wright S. Numerical Optimization. Springer Series in Operations Research; Springer; 1999. ISBN 9780387987934.

[59] Reed M, Simon B. I: Functional Analysis. Methods of Modern Mathematical Physics; Academic Press; 1981. ISBN 0125850506.

[60] Serre D. Matrices: Theory and Applications. Graduate Texts in Mathematics; Springer; 2010. ISBN 9781441976826.

[61] Rudin W. Real and complex analysis. McGraw-Hill series in

higher mathematics; McGraw-Hill; 1974. ISBN 9780070542334.

[62] Graham A. Kronecker products and matrix calculus: with applications. Ellis Horwood series in mathematics and its applications; Horwood; 1981. ISBN 9780470273005.

[63] Davis T. Direct methods for sparse linear systems. Fundamentals of algorithms; Society for Industrial and Applied Mathematics; 2006. ISBN 9780898716139.

[64] Davis TA. Algorithm 915, SuiteSparseQR: Multifrontal multithreaded rank-revealing sparse QR factorization. ACM Trans Math Softw 2011;38(1):8:1–8:22. doi:\bibinfo{doi}{10.1145/2049662.2049670}.

[65] Davis TA, Gilbert JR, Larimore SI, Ng EG. A column approximate minimum degree ordering algorithm. ACM Trans Math Softw 2004;30(3):353–76. doi:\bibinfo{doi}{10.1145/1024074.1024079}.

[66] Davis TA, Gilbert JR, Larimore SI, Ng EG. Algorithm 836: COLAMD, a column approximate minimum degree ordering algorithm. ACM Trans Math Softw 2004;30(3):377–80. doi:\bibinfo{doi}{10.1145/1024074.1024080}.

[67] Amestoy PR, Davis TA, Duff IS. Algorithm 837: AMD, an approximate minimum degree ordering algorithm. ACM Trans Math Softw 2004;30(3):381–8. doi:\bibinfo{doi}{10.1145/1024074.1024081}.

[68] Amestoy PR, Davis TA, Duff IS. An Approximate Minimum Degree Ordering Algorithm. SIAM J Matrix Anal Appl 1996;17(4):886–905. doi:\bibinfo{doi}{10.1137/S0895479894278952}.

[69] Davis TA, William , Hager W. Row modifications of a sparse Cholesky factorization. SIAM J Matrix Anal Appl 2005;:997–1013.

[70] Davis TA, William , Hager W. Multiple-rank modifications of a sparse Cholesky factorization. SIAM J Matrix Anal Appl 2001;22:997–1013.

[71] Davis TA, Hager WW. Modifying a Sparse Cholesky Factorization. SIAM J Matrix Anal Appl 1999;20(3):606–27. doi:\bibinfo{doi}{10.1137/S0895479897321076}.

[72] Davis TA, Hager WW. Dynamic Supernodes in Sparse Cholesky Update/Downdate and Triangular Solves. ACM Trans Math Softw 2009;35(4):27:1–27:23. doi:\bibinfo{doi}{10.1145/1462173.1462176}.

[73] Chen Y, Davis TA, Hager WW, Rajamanickam S. Algorithm 887: CHOLMOD, Supernodal Sparse Cholesky Factorization and Update/Downdate. ACM Trans Math Softw 2008;35(3):22:1–22:14. doi:\bibinfo{doi}{10.1145/1391989.1391995}.

[74] Karypis G, Kumar V. A Fast and High Quality Multilevel Scheme for Partitioning Irregular Graphs. SIAM J Sci Comput 1998;20(1):359–92. doi:\bibinfo{doi}{10.1137/S1064827595287997}.

[75] Gropp W, Lusk E, Skjellum A. Using MPI: portable parallel programming with the message-passing interface. Scientific and engineering computation; MIT Press; 1999. ISBN 9780262571326.

[76] Gabriel E, Fagg GE, Bosilca G, Angskun T, Dongarra JJ, Squyres JM, et al. Open MPI: Goals, Concept, and Design of a Next Generation MPI Implementation. In: Proceedings, 11th European PVM/MPI Users' Group Meeting. Budapest, Hungary; 2004, p. 97–104.