**Atmospheric
Chemistry
and Physics**

# Technical note: Recursive rediscretisation of geo-scientific data in the Modular Earth Submodel System (MESSy)

**P. Jöckel**

Max Planck Institute for Chemistry, Air Chemistry Department, P.O. Box 3060, 55020 Mainz, Germany

**Abstract.** This technical note describes a method for the rediscretisation of "gridded" geo-scientific data. A recursive algorithm (NREGRID) is derived to solve the rediscretisation problem for orthogonal grids (including curvilinear) of arbitrary dimension. The algorithm is used within the program NCREGRID to handle geo-scientific data. These data are typically 2-dimensional (latitude-longitude grid), or 3-dimensional (latitude-longitude grid with a vertical pressure, or hybrid pressure coordinate, as used in atmospheric modelling). NCREGRID can be used as stand-alone program for the transformation ("regridding") of data from and to data files in netCDF format. Moreover, NCREGRID constitutes the core of the Modular Earth Submodel System (MESSy) data import interface, providing a powerful tool for accessing data independently of the applied grid resolution ("automatic regridding").

## 1 Introduction

In 3-dimensional global Earth system modelling the solving of sets of differential equations describing geophysical/chemical properties in the space time domain is often performed by discretisation following the Eulerian approach, i.e., on a discrete grid of spatial coordinates. Model input parameters need to be available on 2- or 3-dimensional (2-D, 3-D) grids of various resolutions. The horizontal grid space usually comprises geographical latitude and longitude. Especially in 3-D global atmospheric models the vertical pressure ($p$) coordinate is often defined by hybrid levels (with index $i$) of the form

$$p(i, x, y, t) = h_a(i) \cdot p_0 + h_b(i) \cdot p_s(x, y, t) \ , \tag{1}$$

where $p_s$ is the surface pressure, $p_0$ is a constant reference pressure, and $h_a$ and $h_b$ are the dimensionless hybrid coeffi-

cients. This representation in a curvilinear coordinate system (dependent on longitude $x$, latitude $y$, and time $t$) allows a terrain following vertical coordinate, if $h_a=0$ and $h_b=1$ for the lowest level (surface level).

A wide variety of grid resolutions is used in different models and for different purposes. Moreover, global geophysical/chemical data (including observations) are provided in different spatial resolutions. As a consequence, data sets have often to be adapted to a specific grid resolution, i.e., rediscretised, or "regridded". For this purpose, NCREGRID has been developed.

The rediscretisation is performed by a recursive algorithm (NREGRID), which is applicable to arbitrary orthogonal (including curvilinear) grids of any dimension, and not restricted to geo-hybrid-grid structures as described above. The algorithm does not apply a point-to-point interpolation, but a transformation based on overlaps between the different grid volumes.

NCREGRID can be used as a stand-alone program, and/or coupled as an import interface to a model, to rediscretise automatically the input from an arbitrary grid space onto the required grid resolution. Specifically, NCREGRID provides the core of the Modular Earth Submodel System (MESSy, Jöckel et al., 2005) data import interface.

This technical note describes the construction of the algorithm (Sects. 2 and 3), the technical implementation (Sect. 4), the application of NCREGRID as MESSy data import interface (Sect. 5), and shows some examples (Sect. 6). More detailed information, user manuals, and the complete source code (of the current version 1.4b) with installation instructions can be found in the electronic supplement (http://www.atmos-chem-phys.net/6/3557/2006/acp-6-3557-2006-supplement.zip). Future updates will become available on the internet[1].

*Correspondence to:* P. Jöckel
(joeckel@mpch-mainz.mpg.de)

---

[1]http://www.mpch-mainz.mpg.de/~joeckel/ncregrid/

## 2 The rediscretisation problem

A function, which is defined on a grid should be redefined on another grid. $\mathbb{X} \subset \mathbb{R}^k$ is the $k$-dimensional ($k \in \mathbb{N}$) domain decomposed into a grid $\mathcal{X} = \{X_1, \ldots, X_J\}$, i.e., a set of a finite number $J$ of distinct $k$-dimensional grid-boxes (or discretisation intervals) $X_j$:

$$\bigcup_{j=1}^{J} X_j = \mathbb{X} , \tag{2}$$

$$X_i \cap X_j = \emptyset \quad \forall i \neq j , \tag{3}$$

$$X_j \neq \emptyset \quad \forall j = 1 \ldots J . \tag{4}$$

With the Lebesgue measure denoted by $| \bullet |$ it is further

$$|X_j| \neq 0 \quad \forall j = 1 \ldots J , \tag{5}$$

and

$$|\mathcal{X}| := \sum_{j=1}^{J} |X_j| = |\mathbb{X}| < \infty . \tag{6}$$

The set of all possible grids is $\Gamma$. Let $\Omega$ be the set of functions

$$\Omega : \Gamma \longrightarrow \mathbb{R}, \quad \mathcal{X} \mapsto \Omega(\mathcal{X}) \tag{7}$$

represented by $F \in \Omega$ on a specific grid $\mathcal{A} \in \Gamma$ with $N$ grid-boxes, i.e.,

$$F : \mathcal{A} \longrightarrow \mathbb{R}, \quad A_n \mapsto F(A_n), \quad n = 1 \ldots N . \tag{8}$$

The problem is then to construct the representation $G \in \Omega$ defined on another grid $\mathcal{B} \in \Gamma$ with $M$ grid-boxes

$$G : \mathcal{B} \longrightarrow \mathbb{R}, \quad B_m \mapsto G(B_m), \quad m = 1 \ldots M . \tag{9}$$

The weighted average of $F$ on its grid $\mathcal{A}$ is

$$\bar{F} = \frac{\sum_{n=1}^{N} F_n \cdot |A_n|}{\sum_{n=1}^{N} |A_n|} , \tag{10}$$

where $F_n$ is the value of $F$ within the grid-box $A_n$, i.e., $F_n := F(A_n)$ for simpler notation. Likewise, the unweighted sum of $F$ is simply

$$\hat{F} = \sum_{n=1}^{N} F_n . \tag{11}$$

With the measure introduced above, two different overlap matrices **P** and **Q** can be defined:

$$P_{n,m} := \frac{|A_n \cap B_m|}{|A_n|} \tag{12}$$

and

$$Q_{n,m} := \frac{|A_n \cap B_m|}{|B_m|} , \tag{13}$$

resulting in two different rediscretised distributions of $F$ on $\mathcal{B}$, denoted as $G^E$ and $G^I$, respectively:

$$G_m^E := \sum_{n=1}^{N} P_{n,m} \cdot F_n = \sum_{n=1}^{N} \frac{|A_n \cap B_m|}{|A_n|} F_n \tag{14}$$

and

$$G_m^I := \sum_{n=1}^{N} Q_{n,m} \cdot F_n = \sum_{n=1}^{N} \frac{|A_n \cap B_m|}{|B_m|} F_n . \tag{15}$$

For further considerations, the general relation

$$|x \cup y| = |x| + |y| - |x \cap y| \tag{16}$$

can be expanded (e.g., by mathematical induction) for a summation, yielding

$$\sum_{m=1}^{M} |A_n \cap B_m| = |\bigcup_{m=1}^{M} (A_n \cap B_m)| \tag{17}$$

$$= |A_n \cap (\bigcup_{m=1}^{M} B_m)|$$

$$= |A_n \cap \mathbb{X}|$$

The second conversion is applicable because of Eq. (3). Since furthermore $\mathcal{A}$ and $\mathcal{B}$ cover the same domain $\mathbb{X}$ by definition, it further holds

$$A_n \subseteq \mathbb{X} \Leftrightarrow A_n \cap \mathbb{X} = A_n \tag{18}$$

and therefore

$$\sum_{m=1}^{M} |A_n \cap B_m| = |A_n| . \tag{19}$$

Finally, using Eq. (19), for the overlap matrices $P$ and $Q$ it is

$$\sum_{m=1}^{M} P_{n,m} = \sum_{n=1}^{N} Q_{n,m} = 1 . \tag{20}$$

With the definitions above and relation (19) it is straightforward to show for the unweighted sum of $G^E$ (cf. Eq. 11) that

$$\hat{G}^E = \sum_{m=1}^{M} G_m^E \tag{21}$$

$$= \sum_{m=1}^{M} \sum_{n=1}^{N} \frac{|A_n \cap B_m|}{|A_n|} F_n$$

$$= \sum_{n=1}^{N} \frac{F_n}{|A_n|} \cdot \sum_{m=1}^{M} |A_n \cap B_m|$$

$$= \sum_{n=1}^{N} F_n = \hat{F} .$$

Similarly, the weighted average of $G^I$ yields (cf. Eq. 10)

$$\bar{G}^I = \frac{\sum_{m=1}^{M} G_m^I |B_m|}{\sum_{m=1}^{M} |B_m|}$$

$$= \frac{\sum_{m=1}^{M} \sum_{n=1}^{N} |A_n \cap B_m| F_n}{\sum_{m=1}^{M} |B_m|}$$

$$= \frac{\sum_{n=1}^{N} |A_n| F_n}{\sum_{n=1}^{N} |A_n|} = \bar{F}. \tag{22}$$

Note that in the denominator Eqs. (2) and (6) for $\mathcal{A}$ and $\mathcal{B}$ have been used.

The choice of the overlap matrix ($\mathbf{P}$ or $\mathbf{Q}$) for the grid transformation needs to consider which moment of the original discretised function $F$ should be conserved: $\mathbf{P}$ conserves the unweighted sum and is therefore suitable for the rediscretisation of extensive variables (EXT, therefore the superscript $E$ of $G^E$), whereas $\mathbf{Q}$ conserves the weighted average, suitable for the rediscretisation of intensive variables (INT, therefore the superscript $I$ of $G^I$). Intensive quantities are defined at each point within the grid-box and do not depend on the grid-box size, whereas extensive quantities are representative for the grid-box as it is defined, and therefore scale with the grid-box size. Examples for intensive quantities are temperature, tracer mixing ratios, or an average altitude. In contrast, absolute tracer masses are extensive. Figure 1 illustrates the rediscretisation of INT and EXT quantities.

Finally, it should be noted that the algorithm can also be used to rediscretise functions $F$, which can only take discrete values:

$$F : \mathcal{A} \longrightarrow \{f_1, f_2, \ldots, f_u\} \subset \mathbb{Z} \tag{23}$$

The non-fractional land-sea mask with $u=2$, $f_1=1$ (land), and $f_2=0$ (ocean) is a simple example. For these "index functions", averaging over grid-boxes is not necessarily defined. However, with the definition

$$\tilde{F}[v] = \begin{cases} 1 & : & F = f_v \\ 0 & : & F \neq f_v \end{cases} \quad \forall v \in \{1, \ldots, u\} \tag{24}$$

a rediscretisation is still possible, by first transforming all $\tilde{F}[v]$ as intensive quantities (INT), i.e., by calculating $\tilde{G}^I[v]$ using the overlap matrix $\mathbf{Q}$. The result $\tilde{G}_m^I[v] \in [0, 1]$ is the fraction of index value $f_v$ in grid-box $m$ of the destination grid. If a single index value is desired, it can for instance be defined as the index with the highest fraction, i.e.,

$$\tilde{G}_m^I = f_w \; ; \; \tilde{G}_m^I[w] := \max_v \tilde{G}_m^I[v] . \tag{25}$$

## 3  The recursive algorithm

So far, no assumptions on the specific shape or dimension of the discretisation $A_n \subseteq \mathcal{A}$ and $B_m \subseteq \mathcal{B}$ have been made. Only the more general constraints of distinct grid-boxes (Eq. 3) and of full coverage (Eq. 2) have been taken into account.

In practice, the discretisation domain in $k$ dimensions often describes a grid which is spanned by orthogonal axes. In these cases, an efficient recursive algorithm for the calculation of the overlap matrices can be derived, since then the overlap calculation can be separated into a multiplication of 1-dimensional interval overlaps along the orthogonal axes.

The scalar indices $n$ or $m$ (from 1 to $N$ and $M$, respectively), for the unambiguous enumeration of the abstract discretisation intervals (grid-boxes), expand to vector indices as

$$n \rightarrow \boldsymbol{n} = (n_1, n_2, \ldots, n_k) , \tag{26}$$

and likewise for $m$, running from $(1, 1, \ldots, 1)$ to $(N_1, N_2, \ldots, N_k)$ and $(M_1, M_2, \ldots, M_k)$, respectively.

With this, the overlap matrices can be decomposed into $k$ sub-matrices of the form

$$P_{n,m} \rightarrow P_{\boldsymbol{n},\boldsymbol{m}} = P_{n_1,m_1} \bullet P_{n_2,m_2} \bullet \ldots \bullet P_{n_k,m_k} \tag{27}$$

(and likewise for $Q$), whereby the bullet ($\bullet$) indicates the multiplication of the sub-matrix elements $P_{n_i,m_i}$ corresponding to the vector indices $\boldsymbol{n}$ and $\boldsymbol{m}$. It is therefore clear that an overlap matrix element $P_{\boldsymbol{n},\boldsymbol{m}}$ is zero, if the matrix element of at least one sub-matrix $1 \leq i \leq k$ is zero, i.e., if $P_{n_i,m_i}=0$.

Using this information, a recursive algorithm can be designed which calculates the overlap sub-matrix for the first dimension, i.e., for $P_{n_1,m_1}$. For all elements where $P_{n_1,m_1}$ is zero, also $P_{\boldsymbol{n},\boldsymbol{m}}=0$ and the calculation is finished. For those elements where $P_{n_1,m_1} \neq 0$, the algorithm calls itself to calculate the overlap sub-matrix for the second dimension, i.e., $P_{n_2,m_2}$, and so on.

Furthermore, the dimensions can be ordered such that potentially invariant dimensions, defined by

$$P_{n_i,m_i} = \begin{cases} 1 & : & n_i = m_i \\ 0 & : & n_i \neq m_i \end{cases} \tag{28}$$

(i.e., where $P_{n_i,m_i}$ is the identity matrix), are skipped at the end of the computation.

## 4  Technical implementation

NCREGRID is coded according to the Fortran95 standard (ISO/IEC-1539-1). The code quality has been tested by applying the Fortran analyser forcheck[2]. The program can be applied in two different modes:

- As a stand-alone program (stand-alone mode) for the rediscretisation of data files. In this case, the grid information of the destination grid is imported from an additional data file.

- Coupled to a model (interface mode, e.g., linked to a Chemistry Transport Model (CTM), a General Circulation Model (GCM), etc.) for the resolution independent

---

[2]http://www.forcheck.nl
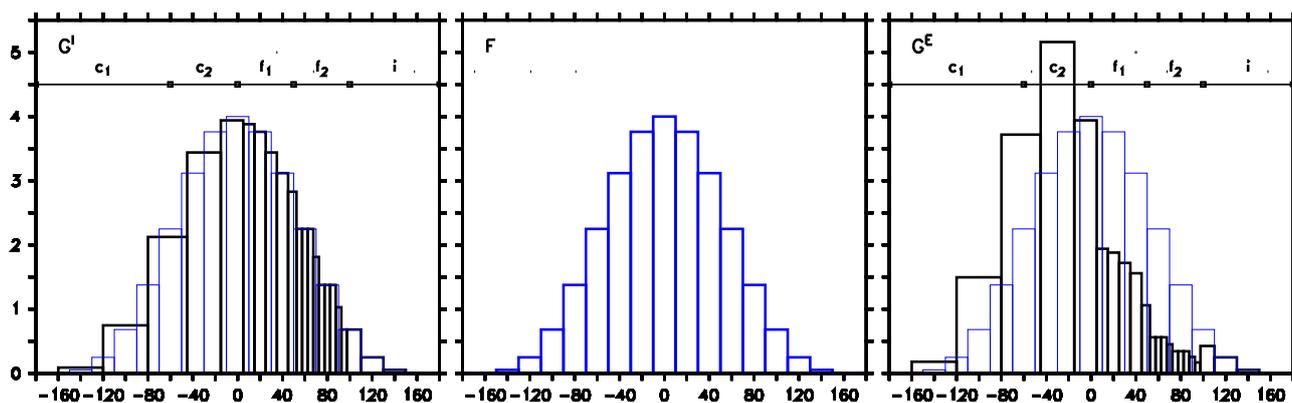
**Fig. 1.** Sketch of the rediscretisation procedure for a function $F$ (blue, middle panel) in 1 dimension (horizontal axis). The original grid is regular with equidistant intervals. The destination grid is coarser within the intervals indicated by $c_1$, and $c_2$, and finer in the intervals $f_1$, and $f_2$, respectively. In interval $i$, destination and original grid are the same (invariant). The left panel shows the result (black), if $F$ describes an intensive quantity ($G^I$, see text), and the right panel shows the result (black) for an extensive quantity ($G^E$, see text).

import of data from external files (automatic regridding). In this case, the grid information of the destination grid (= model grid) is provided by a model-specific interface routine.

In both modes, NCREGRID is controlled via the standard Fortran95 user interface (namelists). The user specifies the input data file, the name of the output data file, and describes the grid information of the input file. In the stand-alone mode additionally the destination grid information has to be provided.

The program is hierarchically organised into the following modules:

1. BASE: This module contains NREGRID, which is the implementation of the recursive algorithm described in Sect. 3 for an arbitrary number $k$ of orthogonal dimensions. To allow an arbitrary number of dimensions, all arrays are internally transformed to arrays of rank 1 with index $n$ by an unambiguous mapping procedure

$$(n_1, n_2, ..., n_k) \leftrightarrow n . \tag{29}$$

This is required, since Fortran95 does not allow the specification of the rank of an array during runtime.

2. NETCDF: This module organises the input and output from and to the external data files. The data file format for gridded data is netCDF[3].

3. GEOHYB: This module meets the specific requirements of geo-hybrid grids (spherical geographic coordinates, vertical curvilinear pressure or hybrid pressure coordinates), and transforms the information (for instance including adequate weights for the latitude axis) for subsequent application of NREGRID.

4. DIAG: This module contains diagnostic routines which are only required for debugging.

5. CONTROL: This module contains the sequence control system of the overall regridding procedure: reading the user information, input data and grid information, reading output grid information, regridding, output of data. Since it is the main entry point for using NCREGRID, this module comprises the low-level (level 1) user-interface for the usage of NCREGRID coupled to a model. In the stand-alone mode, the regridding procedure is called via this level-1 user interface from the main program.

6. TOOLS: This module contains high-level (level 2) interface routines for the application of NCREGRID in interface mode.

7. INTERFACE: This module defines the grid information, if NCREGRID is coupled to a model (see Fig. 2). In the stand-alone mode, the grid information is imported from a data file; in this case this module is empty.

Installation instructions for the stand-alone mode, details about the namelist control, information for using the interface-mode (including the level 1 and level 2 interfaces), more details about the implementation, and the complete source code (of current version 1.4b) are contained in the electronic supplement (http://www.atmos-chem-phys.net/6/3557/2006/acp-6-3557-2006-supplement.zip).

## 5   The MESSy data import interface

The coding of NCREGRID summarised in Sect. 4 strictly adheres to the MESSy standard (Jöckel et al., 2005). Specifically, the modules clearly separate into the submodel core

---

[3]http://www.unidata.ucar.edu/software/netcdf/

layer (SMCL), namely BASE, NETCDF, GEOHYB, and DIAG, and into the submodel interface layer (SMIL), namely CONTROL, and TOOLS. The stand-alone mode with the main program of NCREGRID applies the level 1 interface (CONTROL) and constitutes a MESSy box model.

Moreover, NCREGRID is a generic MESSy submodel, since it provides (in interface mode) an essential part of the MESSy infrastructure, namely a general data import interface (for gridded data) with an automatic regridding facility. Technically, this is achieved by two additional modules for the base model interface layer (BMIL) organising the communication with the base model (e.g., a general circulation model (GCM)). These two additional BMIL modules are

– the module INTERFACE providing the base model grid information to NCREGRID, and

– a module providing the level 3 interface for passing the imported data from NCREGRID back to the base model.

As defined in the MESSy standard, the BMIL takes into account the coding aspects of the base model. Specifically, the level 3 interface routines need to transform the output of NCREGRID (Fortran95 structures) into the data fields of the base model, taking into account, for instance, the domain decomposition in a parallel model, the configuration and order of the dimensions, etc. In doing so, the level 3 module makes use of the level 1 and 2 routines of the SMIL. Figure 2 summarises the structural design aspects of NCREGRID in stand-alone and interface mode. The documentation of the BMIL for the the atmospheric chemistry GCM ECHAM5/MESSy (version 1.1) is provided in the supplement (http://www.atmos-chem-phys.net/6/3557/2006/acp-6-3557-2006-supplement.zip).

## 6 Examples

Two example results of the application of NCREGRID are shown in Fig. 3. The upper left panel shows the Earth topography on a regular 2° by 2° latitude-longitude grid. The colour code represents the orography in km. The result after rediscretisation of this intensive quantity onto a regular 10° by 10° grid is shown in the upper right panel.

The lower left panel depicts a typical vertical distribution of an atmospheric tracer. A cross section along one longitude of a 3-dimensional dataset is shown. The quantity (colour code) is given as mixing ratio in arbitrary normalised units. In the vertical direction, the original data are on constant pressure levels (in hPa). The lower right panel shows the result (at the same longitude) after rediscretisation onto a different grid. The latitudinal, longitudinal, and the vertical grid intervals changed, specifically, the vertical axis of the destination grid is defined in hybrid pressure coordinates (Eq. 1). Note that this is not visible in the figure, since a
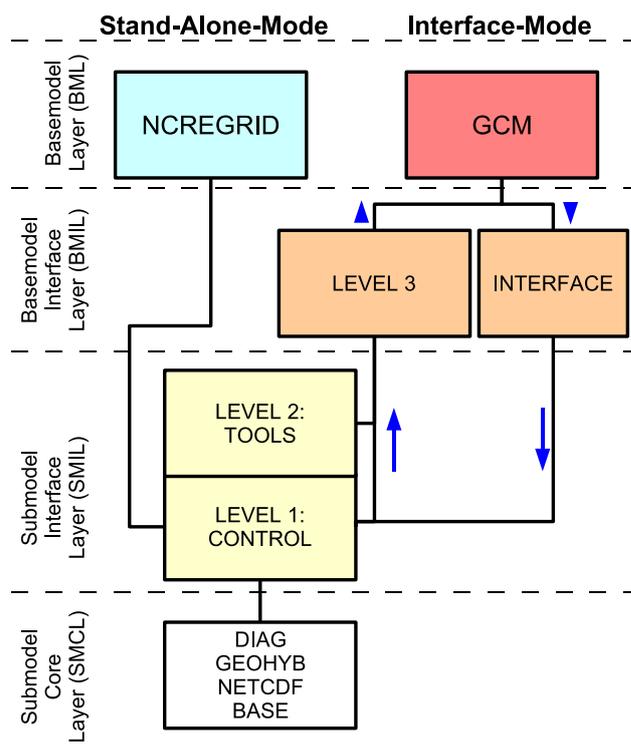


**Fig. 2.** Modular structure of NCREGRID and the MESSy data import interface. The modules in the submodel core layer (SMCL) and submodel interface layer (SMIL) are described in Sect. 4. For the stand-alone mode, NCREGRID provides the main program (constituting a MESSy "box" model). The alternative interface mode (here indicated by the coupling to a general circulation model (GCM) in the base model layer (BML)) requires the base model grid description to be provided in the module INTERFACE, and an additional module with level 3 interface routines, providing the base model interface layer (BMIL). The level 3 routines need to take into account the specific design aspects of the base model, here the GCM. The blue arrows indicate the information flow.

constant surface pressure has been assumed for plotting. The black crosses in the lower panels indicate the mid points of the boxes of the underlying grids.

## 7 Conclusions

This technical note describes a recursive algorithm and its implementation (NCREGRID) for the rediscretisation of ("gridded") geo-scientific data from one grid to another. The method presented here is conceptually different from standard point-to-point interpolation methods, as for instance linear-, bilinear- or spline-interpolation methods (e.g., Press et al., 1996). Those point-to-point interpolation methods suffer three major disadvantages: First, they are strictly speaking only applicable for intensive quantities; second, the result depends on the chosen interpolation method (e.g., the order
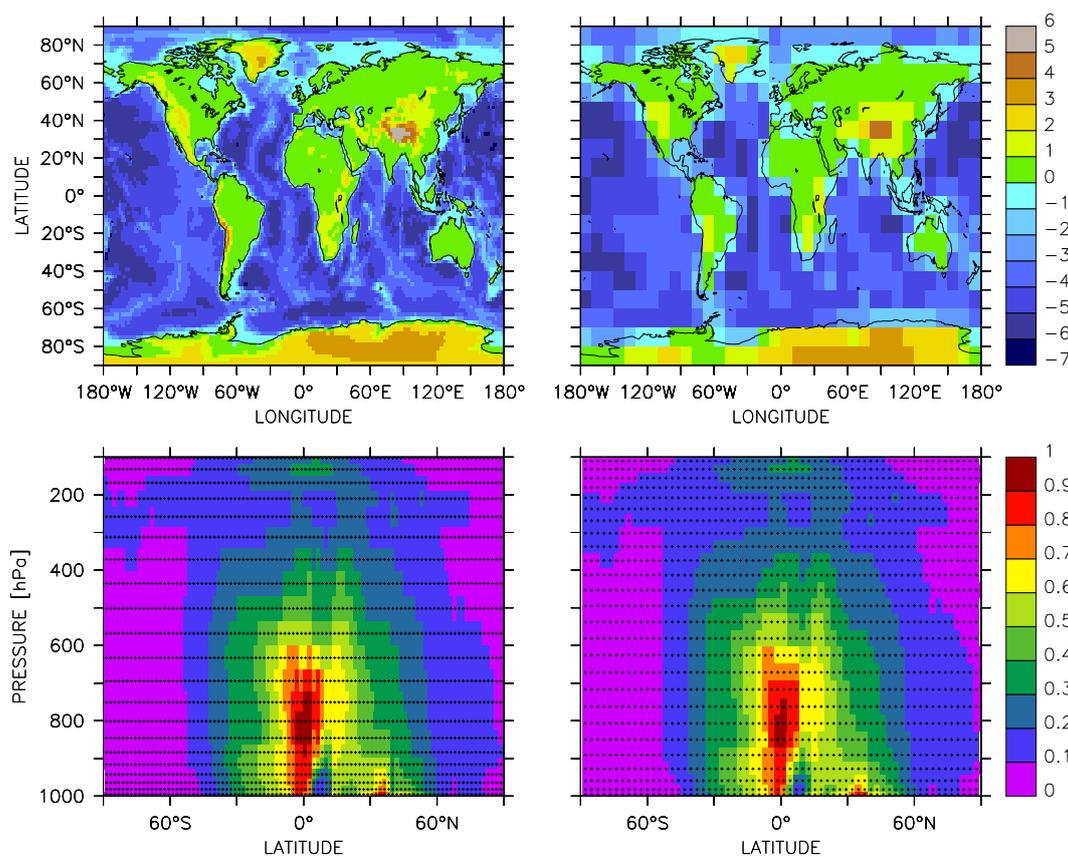
**Fig. 3.** Examples for the application of NCREGRID showing topographical height in km (top) and a vertical tracer distribution in normalised units (bottom). Details are given in the text.

of the interpolating polynomial); and third, they do by construction not necessarily conserve integral moments of the distribution. In contrast to this, the presented algorithm is applicable to intensive and extensive quantities, it is unambiguous, and it conserves integral moments, such as the weighted average of intensive quantities, and the unweighted sum of extensive quantities. No arbitrary information is added if data are rediscretised from a coarser to a finer grid, and vice versa, only minimal information is lost. Implemented in a modular structure following the MESSy standard, NCRE-GRID constitutes a generic MESSy submodel: NCREGRID itself is structured as a MESSy submodel, i.e., separated in submodel core layer and submodel interface layer, including a MESSy box-model (i.e., the stand-alone mode). With two additional modules in the base model interface layer, NCRE-GRID serves at the same time as the overall MESSy data import interface for automatic regridding of imported gridded data.

## References

Jöckel, P., Sander, R., Kerkweg, A., Tost, H., and Lelieveld, J.: Technical Note: The Modular Earth Submodel System (MESSy) – a new approach towards Earth System Modeling, Atmos. Chem. Phys., 5, 433–444, 2005, http://www.atmos-chem-phys.net/5/433/2005/.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., Flannery, B. P., and Metcalf, M.: Numerical Recipes in Fortran 90, Vol. 2, Cambridge University Press, 1996.