International TOVS Study Conference 24 (ITSC-24)
Tromsø, Norway
16 March - 22 March 2023

Deutscher Wetterdienst
Wetter und Klima aus einer Hand

A highly efficient DFS-score based channel selection method suitable for non-diagonal R matrices and the parallel processing of large numbers of atmospheric profiles

Olaf Stiller

Abstract

The large numbers of channels from hyper-spectral sounders has motivated the development of different automated channel selection methods. These are commonly based on a figure of merit the most prominent of which is the DFS (Degrees of Freedom per Signal). Observation errors clearly play a crucial role for the channel selection and importantly, some of the proposed schemes can deal with fully non-diagonal observation error covariance matrices R as they are now used at many operational centers yielding a positive impact on the forecast accuracy. Including non-diagonal R in the selection process, however, usually comes at increased numerical cost and code complexity. It has to be noted that numerical cost is a limiting factor which requires some compromises and that finding a reasonably large subgroup of channels which strictly minimizes the total DFS for a large group of atmospheric profiles simultaneously would be prohibitive.

This poster explains a new, numerically highly effective method for computing the DFS for fully non-diagonal R matrices and large numbers of atmospheric test profiles simultaneously. The method which works sequentially (like previous schemes) is based on an application of the Cholesky decomposition to the DFS. Its high efficiency results from an appropriate ordering of the steps for computing the different terms resulting from the Cholesky partitioning.

1. Using the DFS (degrees of freedom for signal) for channel selection

Automated channel selection algorithms for instruments like IASI, CRIS or AIRS often use the **DFS** as a measure for the

- **uncertainty reduction** achieved **by a selected set of channels** (true if theoretical assumptions and employed covariance matrices were correct).

Sequential schemes have been developed

- starting from a single channel (with the highest DFS)
- choosing additional channels → DFS of the new set is maximized at each step.

This requires computing the DFS for large numbers of channel combinations.

This poster presents a newly developed method to make these computations in a numerically cheap and efficient manner. Starting from the observation space formulation of the DFS (chosen from various mathematically equivalent forms)

$$\mathrm{DFS} = Tr([\boldsymbol{H}\boldsymbol{B}H^T + \boldsymbol{R}]^{-1}\boldsymbol{H}\boldsymbol{B}H^T)$$

$$= Tr([\boldsymbol{M}^{[1]}]^{-1}\boldsymbol{M}^{[2]})$$

with $\boldsymbol{M}^{[1]} \equiv \boldsymbol{H}\boldsymbol{B}H^T + \boldsymbol{R}$ and $\boldsymbol{M}^{[2]} \equiv \boldsymbol{H}\boldsymbol{B}H^T$

2. Using Cholesky Decomposition: $\boldsymbol{M} = \boldsymbol{L}\,\boldsymbol{L}^T$ (for symmetric positive definite matrices $\boldsymbol{M}^{[1]}$, $\boldsymbol{M}^{[2]}$)

Find square root $\boldsymbol{L}$ of matrix $\boldsymbol{M}$ so that: $\boldsymbol{L}\,\boldsymbol{L}^T = \boldsymbol{M}$

$$\boldsymbol{L} = \begin{pmatrix} L_{1,1} & 0 & \cdots & \cdots & 0 \\ L_{2,1} & L_{2,2} & 0 & & \vdots \\ \vdots & & \ddots & \ddots & \vdots \\ \vdots & & & \ddots & 0 \\ L_{N,1} & \cdots & \cdots & & L_{N,N} \end{pmatrix}$$

lower triangular (LT) matrix
Note: $\boldsymbol{L}^{-1}$ is also LT.

$$\mathrm{DFS} = Tr\left(\left(\boldsymbol{M}^{[1]}\right)^{-1}\left(\boldsymbol{M}^{[2]}\right)\right)$$
$$= Tr\left(\left(\boldsymbol{L}^{[1]}\,\boldsymbol{L}^{[1]^T}\right)^{-1}\left(\boldsymbol{L}^{[2]}\,\boldsymbol{L}^{[2]^T}\right)\right)$$
$$= Tr(\boldsymbol{G}(\boldsymbol{G})^T)$$
$$= \sum_{l,m}\left(\boldsymbol{G}_{\,l,m}\right)^2$$

with LT matrix: $\boldsymbol{G} \equiv \left(\boldsymbol{L}^{[1]}\right)^{-1}\left(\boldsymbol{L}^{[2]}\right)$

2a. Sequential algorithm for matrices $\boldsymbol{L} \in \left\{\boldsymbol{L}^{[1]}, \boldsymbol{L}^{[2]}\right\}$

Cholesky method is sequential :

at step k one has:

$L_{i,j}$ for $j < k$ and for *all i*

we also compute for $l, m > k$:

$$B^{(k)}_{l,m} \equiv M_{l,m} - \sum_{s=1}^{k} L_{l,s}\,L_{m,s}$$

step k → k+1

$$L_{k+1,k+1} = \sqrt{B^{(k)}_{k+1,k+1}}$$

for all $m > k$:

$$L_{m,k+1} = \frac{B^{(k)}_{m,k+1}}{L_{k+1,k+1}}$$

Components which have been computed (are known) at step *k* of our Cholesky algorithm of a matrix.

$$\begin{pmatrix} L_{1,1} & 0 & \cdots & 0 & & & \\ L_{2,1} & L_{2,2} & 0 & \vdots & & & \\ \vdots & & \ddots & 0 & & & \\ L_{k,1} & \cdots & \cdots & L_{k,k} & & & \\ L_{k+1,1} & \cdots & \cdots & L_{k+1,k} & B^{(k)}_{k+1,k+1} & \cdots & B^{(k)}_{k+1,N} \\ \vdots & & & \vdots & \vdots & & \vdots \\ \vdots & & & \vdots & \vdots & & \vdots \\ L_{N,1} & \cdots & \cdots & L_{N,k} & B^{(k)}_{N,k+1} & \cdots & B^{(k)}_{N,N} \end{pmatrix}$$

2b. Sequential algorithm for $\boldsymbol{G} \equiv \left(\boldsymbol{L}^{[1]}\right)^{-1}\left(\boldsymbol{L}^{[2]}\right)$

This allows a sequential computation of $\boldsymbol{G}$

at step k one has: $G_{i,j}$ for $i, j < k$

we also compute for $l > k$ and $m \le k$:

$$S^{(k)}_{l,m} \equiv L^{[2]}_{l,m} - \sum_{s=m}^{k} L^{[1]}_{l,s}\,G_{s,m}$$

step k → k+1:

$$G_{k+1,\boldsymbol{k+1}} = \frac{L^{[2]}_{k+1,\boldsymbol{k+1}}}{L^{[1]}_{k+1,\boldsymbol{k+1}}}$$

and $G_{k+1,m} = \dfrac{S^{(k)}_{k+1,m}}{L^{[1]}_{k+1,\boldsymbol{k+1}}}$ for all $m \le k$:

Components which have been computed (are known) at step *k* of our Cholesky algorithm for $\boldsymbol{G} \equiv \left(\boldsymbol{L}^{[1]}\right)^{-1}\left(\boldsymbol{L}^{[2]}\right)$.

$$\begin{pmatrix} G_{1,1} & 0 & \cdots & 0 \\ G_{2,1} & G_{2,2} & 0 & \vdots \\ \vdots & & \ddots & 0 \\ G_{k,1} & \cdots & \cdots & G_{k,k} \\ S^{(k)}_{k+1,1} & \cdots & \cdots & S^{(k)}_{k+1,k} \\ \vdots & & & \vdots \\ S^{(k)}_{N,1} & \cdots & \cdots & S^{(k)}_{N,k} \end{pmatrix}$$

Summary of sequential code for *G*

Step *1*

For matrixes [1] and [2]

for all m : $L_{m,1} = M_{m,1}/\sqrt{M_{1,1}}$

for all $l, m > 1$:

$$B^{(k=1)}_{l,m} \equiv M_{l,m} - L_{l,1}\,L_{m,1}$$

For $\boldsymbol{G} \equiv \left(\boldsymbol{L}^{[1]}\right)^{-1}\left(\boldsymbol{L}^{[2]}\right)$

$$G_{1,1} = L^{[2]}_{1,1}/L^{[1]}_{1,1}$$

for all $m > 1$:

$$S^{(k=1)}_{m,1} \equiv L^{[2]}_{m,1} - L^{[1]}_{m,1}\,G_{1,1}$$

Update step : *k* → *k+1*

For matrixes [1] and [2]

$$L_{k+1,k+1} = = \sqrt{B^{(k)}_{k+1,k+1}}$$

for all $m > k+1$:

$$L_{m,k+1} = B^{(k)}_{m.k+1}/L_{k+1,\boldsymbol{k+1}}$$

for all $l, m > k$:

$$B^{(k+1)}_{l,m} \equiv B^{(k)}_{l,m} - L_{l,k}\,L_{m,k}$$

For $\boldsymbol{G} \equiv \left(\boldsymbol{L}^{[1]}\right)^{-1}\left(\boldsymbol{L}^{[2]}\right)$

$$G_{k+1,k+1} = L^{[2]}_{k+1,k+1}/L^{[1]}_{k+1,k+1}$$

for $m \le k$:

$$G_{k+1,m} = S^{(k)}_{k+1,m}/L^{[1]}_{k+1,k+1}$$

for $m > k$, $l \le k$:

$$S^{(k+1)}_{m,l} \equiv S^{(k)}_{m,l} - L^{[1]}_{m,k+1}\,G_{l,k+1}$$

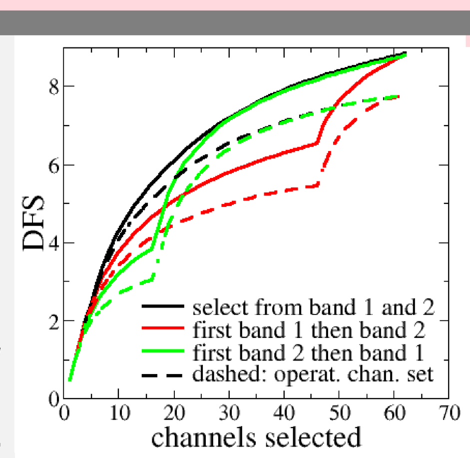
Update of the DFS: (*k* → *k+1*) :

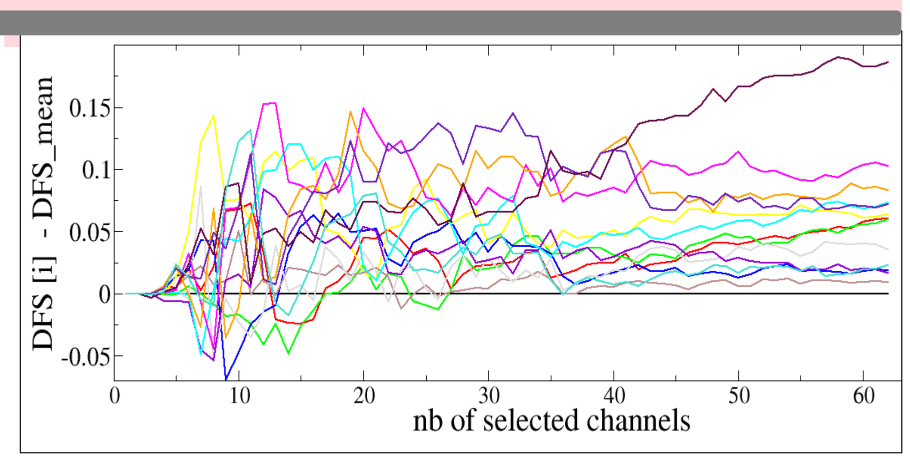
One has:

$$\Delta\mathrm{DFS}_{k+1}(m) \equiv \mathrm{DFS}_{k+1}(m) - \mathrm{DFS}_k = \frac{\left(B^{(k)}_{m,m}\right)^{[2]}}{\left(B^{(k)}_{m,m}\right)^{[1]}} + \sum_{s=1}^{k}\frac{\left(S^{(k)}_{m,s}\right)^2}{\left(B^{(k)}_{m,m}\right)^{[1]}}$$

$\Delta\mathrm{DFS}_{k+1}(m)$: value of DFS added at step *k+1* if channel m is selected at step *k+1*

Fig.1: Example selecting 62 IASI channels (46: band 1, 16: band 2)
Left: DFS of selected IASI channels
solid lines: maximal DFS selection
dashed lines: operational set
green: first only band 1 ch.selected
red : first only band 2 ch.selected
black : bands select.simultaneously
Right: difference of *DFS for individual FoVs* from average DFS (all FoVs) (corresponding to simultan.selection).





3. Recommended selection strategy

Compute DFS for large number of FoVs simultaneously

At each step (with k selected channels in each FoV) compute $\Delta\mathrm{DFS}_{k+1}(m)$

- for all not yet selected channels m (fulfilling possible constraints)
- in all FoVs

Choose channel m to maximize average $\Delta\mathrm{DFS}_{k+1}$ (averaged over all FoVs - *black solid curve Fig.1, left)*

Compare simultaneous DFS with results for individual FoVs (DFS maximized for single FoV – Fig.1 right)

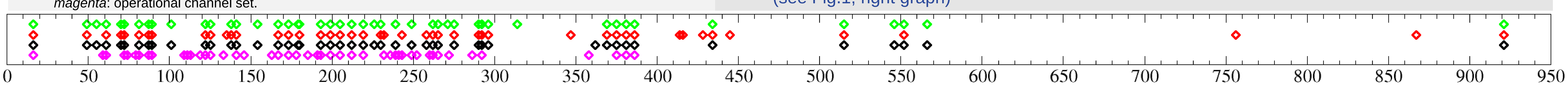
Left graph shows that: total DFS differs little whether channels are chosen from

- available channels in both bands (black curves),) or
- first only from band 1 (2) – red (green) curve

Still: the channels selected by these different procedures differ (Fig.2).

Here: set selected simultaneously for all FoVs is *only slightly less optimal* for individual FoVs, respectively.
(see Fig.1, right graph)

Fig.2: *black, red, green symbols:*, channels selected from band 1 by selection processes in Fig. 1 (left, solid curves) *magenta*: operational channel set.



DWD, German Meteorological Service, Data assimilation section
Author: Olaf Stiller, Olaf.Stiller@dwd.de