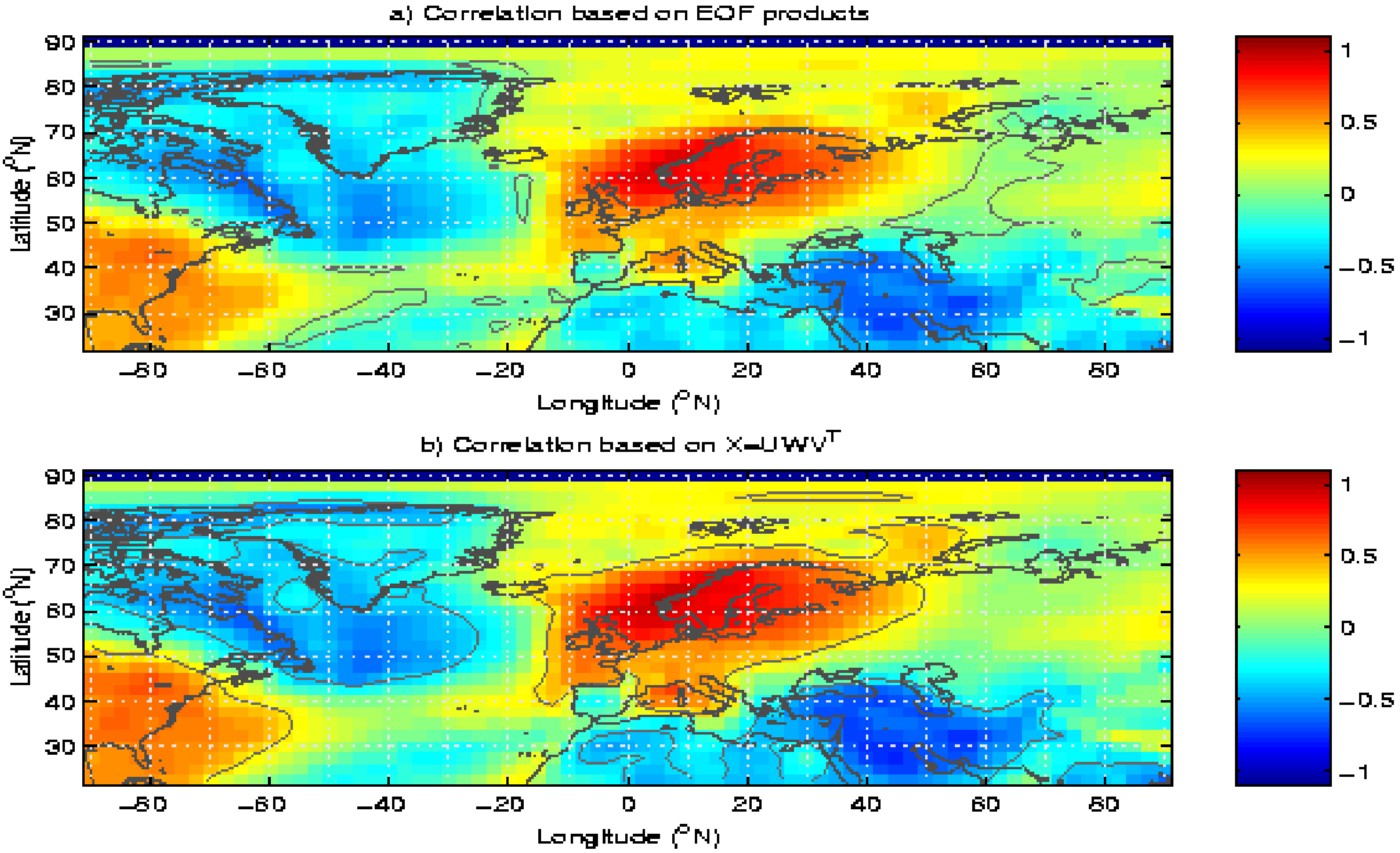
Example of use: Spatial correlation maps based on EOF products

Figure: *A comparison between correlation maps between* $\Phi_{500hPa}(60^\circ N, 5^\circ E)$ *and* $\Phi_{500hPa}(x, y)$*, based on EOFs (a) and conventional calculations (b). The confidence limit for panel b was estimated using the an MC-test on the point of maximum correlation. Note, that the confidence limits are generally higher for the conventional method, whereas only the lowest correlations in panel a are insignificant. [stats_uib_8_4.m]*



By computing the EOFs and retaining a few of the first leading EOFs ($n_{eofs} \ll n_t$**), one can compress the size of the data from** $n_x \times n_y \times n_t$ **to** $n_x \times n_y \times n_{eofs} + (n_t + 1) \times n_{eofs}$ **with a minimal loss of information (filters away much of the small-scale noise). If we have a** Φ_{500hPa} **record, such as in Fig. 8.4, stored as 100 time slices on a** 50×30 **grid and we retain the 10 leading EOFs, then the data size can be reduced from 150,000 numbers to just 16,010 numbers and still account for about 90% of the**

variance in Φ_{500hPa}.

In addition to reducing the data size, the EOFs can save computation time since there are only n_{eofs} independent numbers. For instance correlation analysis can be applied to the n_{eofs} PCs, weighted by the EOF patterns and their variance, instead of the time series from $n_x \times n_y$ points. The subscripts used below are the location vector $\vec{r}$, the EOF index k and the time index t.

$$r(\mathbf{X}_{\vec{r},t}, y_t) = \frac{\sum_t \left[\mathbf{X}'_{\vec{r},t} y'_t\right]}{\sqrt{\sum_t (\mathbf{X}'_{\vec{r},t})^2 \times \sum_t (y'_t)^2}} \tag{8.17}$$

$$\mathbf{X}'_{\vec{r},t} = \mathbf{U}_{\vec{r},k} \mathbf{W}_k \mathbf{V}^T_{k,t}$$

$\mathbf{W}_k$ is a diagonal matrix

$$r(\mathbf{X}_{\vec{r},t}, y_t) = \frac{\sum_t \mathbf{W}_k \left[\mathbf{U}_{\vec{r},k} \mathbf{W}_k \mathbf{V}^T_{k,t} y'_t\right]}{\sqrt{\sum_t \mathbf{W}_k (\mathbf{U}_{\vec{r},k} \mathbf{W}_k \mathbf{V}^T_{k,t})^2 \times \sum_t (y'_t)^2}}$$

$$r(\mathbf{X}_{\vec{r},t}, y_t) = \frac{\mathbf{W}_k \mathbf{U}_{\vec{r},k} \mathbf{W}_k \sum_t \left[\mathbf{V}^T_{k,t} y'_t\right]}{\sqrt{\mathbf{W}_k \mathbf{U}^2_{\vec{r},k} \mathbf{W}^2_k \sum_t (\mathbf{V}^2_{k,t}) \times \sum_t (y'_t)^2}} \tag{8.18}$$

The EOFs also allow better confidence limit estimates, as a confidence interval can be computed for each of the n_{eofs} PC. The geographically distributed confidence levels may be computed with a similar spatial weighting as the correlations themselves. For MC-test, using EOF products greatly reduces the computation time.

Maps of linear trends can be estimated in a simpler fashion: $\mathbf{X}_{\mathrm{trend}} = \mathbf{U}_{\vec{r},k}\mathbf{W}_{k}\mathbf{trend}^{T}_{k,t}$.

The EOF products are also often used in regressional analysis and CCA.
