VEÐURSTOFA ÍSLANDS

Report
03007

Halldór Björnsson

Eigenvectors of local topography in Iceland

VÍ-ÚR07
Reykjavík
February 2003

Abstract

This report describes the calculation of eigenvectors of local topography for Iceland. The motivation was to generate predictands to use with regression. In this report "local" refers to a scale of 5 by 5 km. The results for the scale show that half of local topographic variance can be explained by one eigenvector, and the first four combined explain almost 90% of the variance. These four eigenvectors 'localize' the following features of the landscape: a) east-west slope, b) north-south slope, c) east west ridge and d) a saddle like topography.

Data and method

The topographic data was the same one as used in the regression study by Gylfadóttir [1]. This data is obtained from the U. S. National Oceanic and Atmospheric Administration (NOAA) web server.[1] The data is on a grid that is approximately 1 by 1 km (see figure 1). The size of the grid is 1440 by 480 cells, but of those 59% lie outside the coast of Iceland. The methodology for calculating the eigenvectors is fairly standard [2] and will not be described in detail in this report.

Eigenvectors of 5 by 5 km elements.

The method used is based on the French AURELHY method [3]. To calculate the eigenvectors of local topography, each grid point on the map was sampled along with surrounding points. For a 5 by 5 km grid we chose as surrounding points those that lie within a box that extends two grid-points to the north, south, east and west (see figure 2). Henceforth, the grid-point along with this halo region will be referred to as a local topographic element.

For each grid-point on the map a topographic element is obtained and stored. In some cases a part of an element will extend outside the coastline. In such cases the element is tagged, and not used in the ensuing analysis. This process yielded 259,104 elements, each consisting of 25 numbers. Notice that the total number of elements is slightly less than the number of grid-points (1440×480) since grid-points on the periphery of the domain cannot be used. The element mean was subtracted from each element, and the resulting data stored in a matrix M. We will refer to this data as the centered element matrix.

Figure 3 shows one line from the centered element matrix. The line chosen is the one corresponding to the central grid-point. The resulting map shows the "local" topography, in that positions of small scale valleys and mountain ridges will appear quite clearly, but without larger scale features. The local topography has several interesting features. Strong erosional features are evident both in the eastern part and the western part of the map. This is especially clear in parts of west-Iceland, where glacier carved valleys predominate the landscape. In a region extending from the southwest of Iceland to the northeast, the erosional features seem to be absent, but this area corresponds to the youngest part of the surface.

[1] http://www.noaa.gov/topography.html

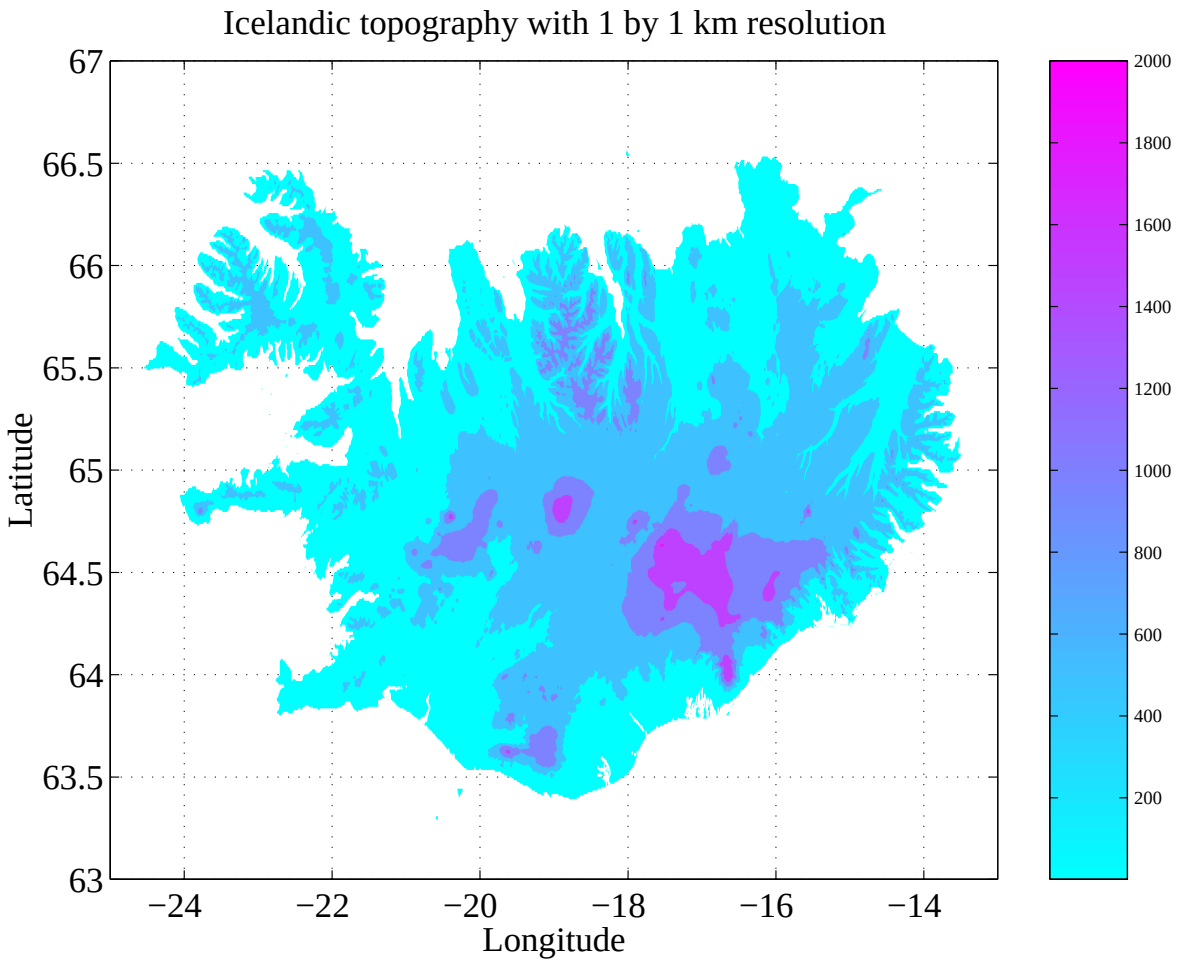

Figure 1: The topography of Iceland on a grid that is approximately 1 by 1 km.

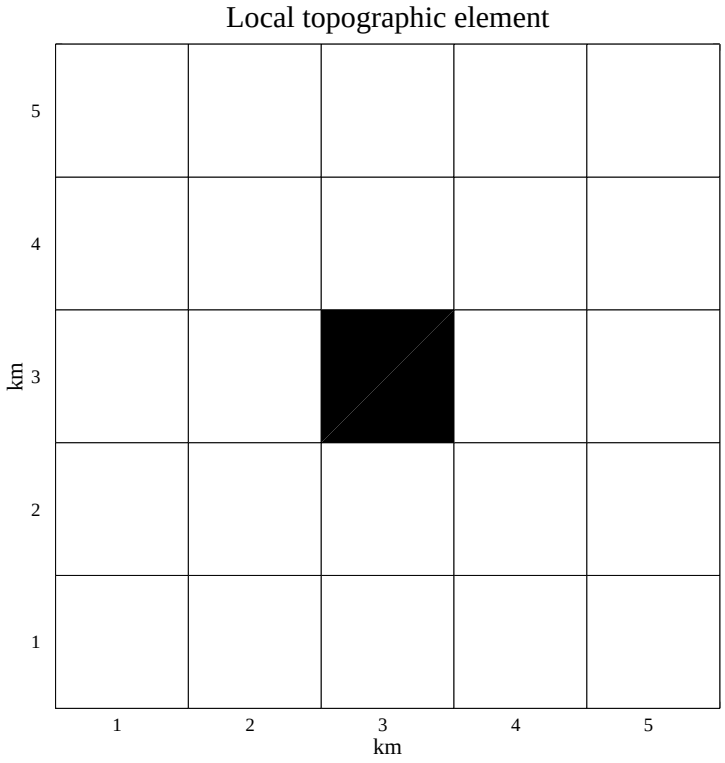

Figure 2: A local topographic element on the 5 by 5 km scale consists of a central grid-point and a two grid-point wide halo.

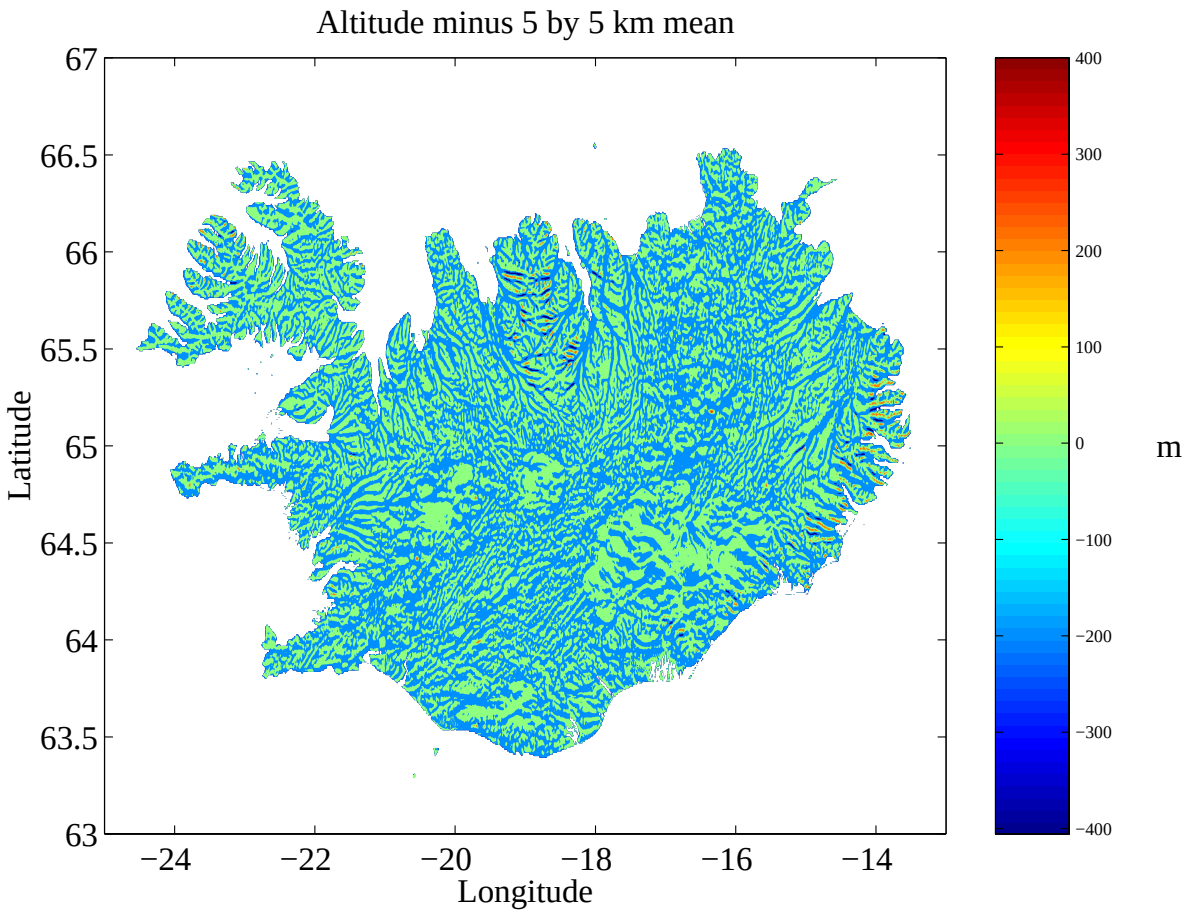

Figure 3: A subset of the centered element matrix M, i.e. line 13. This line corresponds to the middle grid-point of each topographic element. One can think of this figure as showing the "local" topography, i.e., a map that clearly shows positions of valleys and ridges.

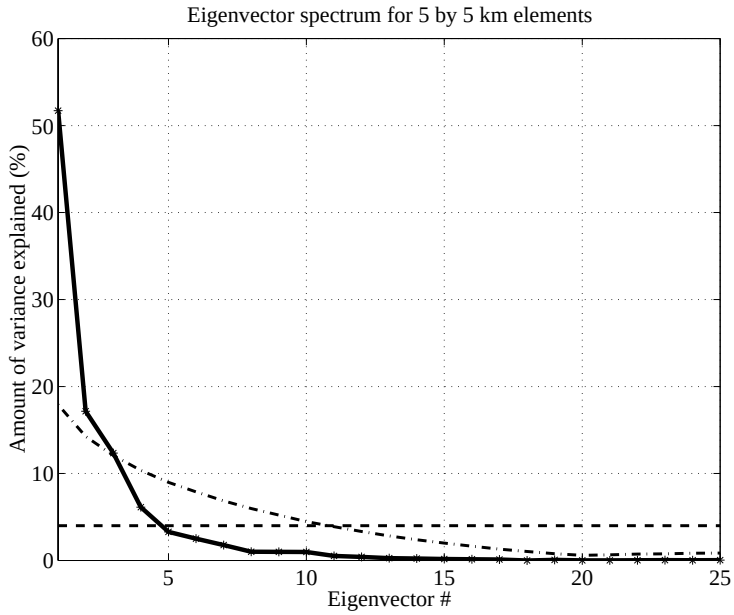

Figure 4: Eigenvalues of R (solid line and *). The two dotted lines show Monte Carlo estimations of eigenvalue sizes expected by chance (see appendix for discussion).

From the centered element matrix a covariance matrix R can now be defined[2]

$$R = MM^t.$$

If the size of M is $(25 \times 259,104)$ then the size of R will be (25×25). The eigenvectors of R are solutions to

$$RB = B\Lambda,$$

where Λ is a diagonal matrix with the eigenvalues of R, and B is a matrix with the corresponding eigenvectors of R as column vectors. The relative importance of each eigenvector can be gauged from the size of the eigenvalue. If λ_{ii} is the eigenvalue corresponding to eigenvector number i, then the amount of variance explained by the eigenvector given by $\lambda_{ii}/tr(\Lambda)$, where tr is the trace (i.e., the sum of the diagonal) of Λ.

Figure 4 shows the amount of variance explained by each eigenvector. Clearly the first few explain most of the variance. To estimate the significance of each eigenvalue, we performed two different Monte-Carlo (MC) type experiments, the details of which are given in the appendix. The results of the MC experiment indicated that the first four eigenvectors should be retained. Eigenvectors $1-4$ explain respectively 52%,17%, 12% and 6% of the variance.

Figure 5 shows the four eigenvectors corresponding to these eigenvalues. The first two eigenvectors describe a dipole in the north-south (east -west for eigenvector 2) direction. Eigenvector three describes a unimodal feature, whereas eigenvector four is a saddle like feature.

[2]Strictly speaking R is not a covariance matrix, since M is has had the column averages subtracted, but not the line averages.

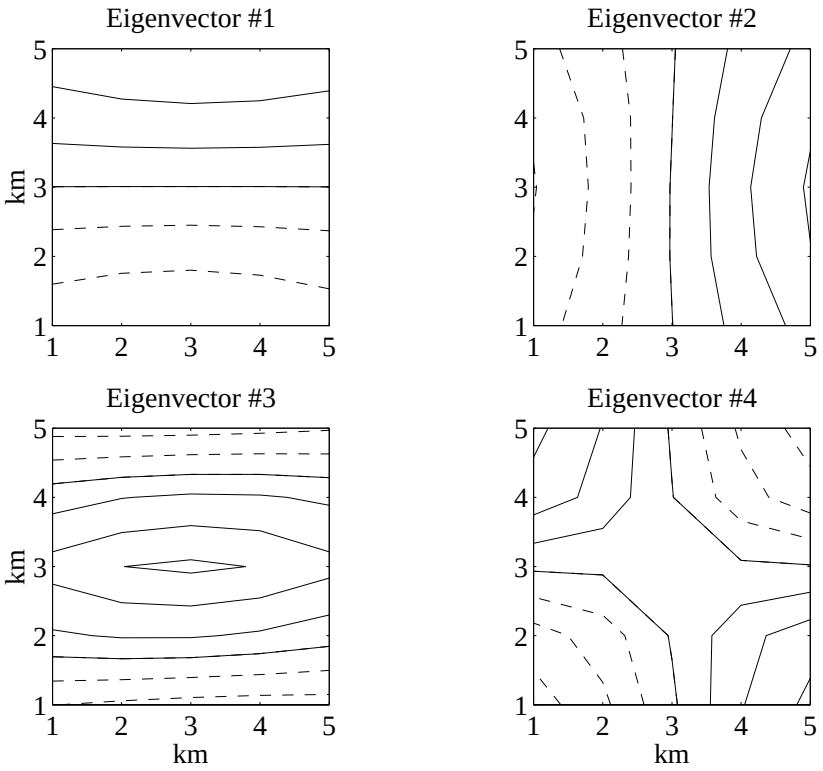

Figure 5: The shape of the four eigenvectors corresponding to the four largest eigenvalues.

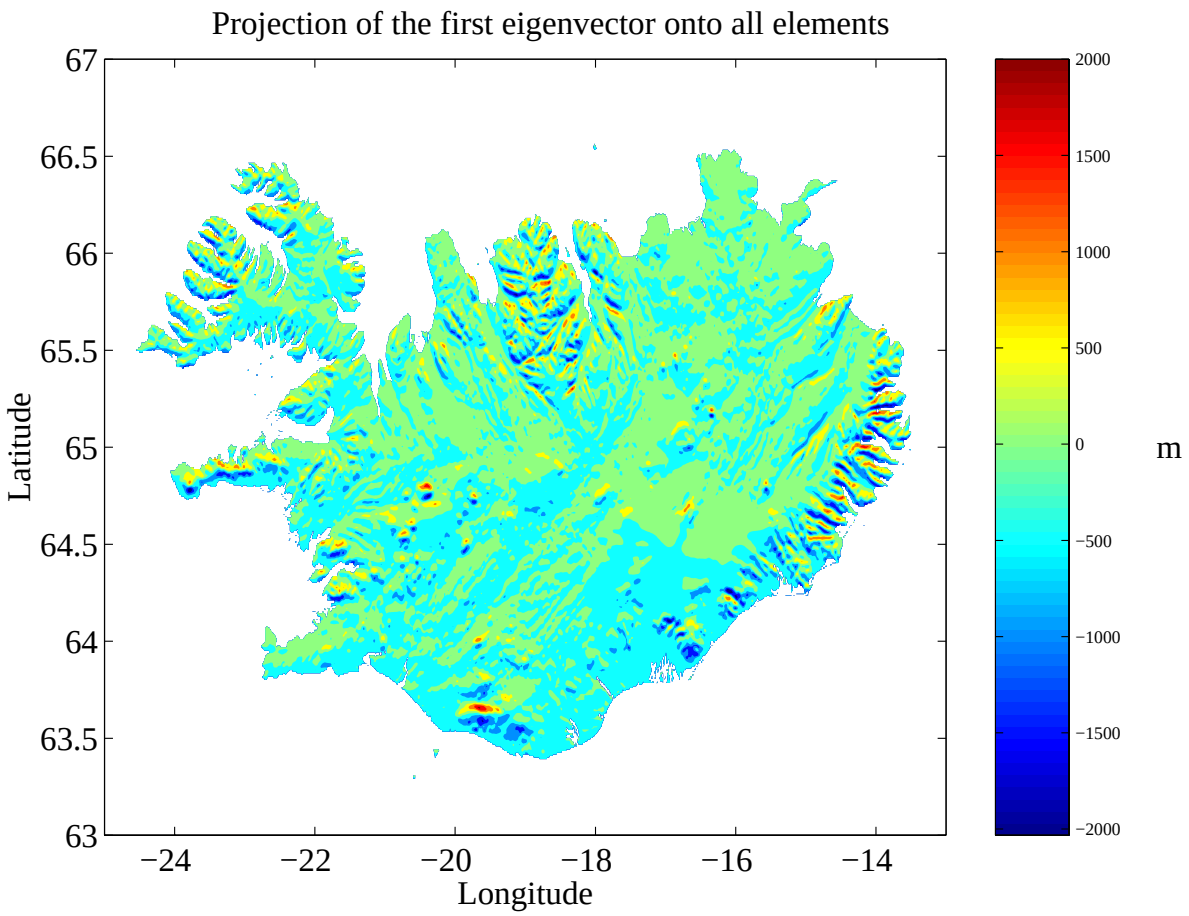

Figure 6: The projection of the first eigenvector onto the centered element matrix.

If $\mathbf{b}_i$ is the eigenvector corresponding to eigenvalue λ_{ii} the projection of the eigenvector onto the centered data matrix is simply found by calculating

$$\mathbf{a}_i = M^t \mathbf{b}_i.$$

With M is $(25 \times 259, 104)$ and $\mathbf{b}_i$ is be (25×1) the size of $\mathbf{a}_i$ will be $(259, 104 \times 1)$, which is identical to the number of elements obtained from the original map.

Figures 6-9 show the projections of these eigenvectors onto the centered element matrix. The figures show that the projections will have the largest amplitudes in the fjord areas on the east coast, and in the northwest peninsula. Large amplitudes are also apparent on the mid-peninsula to the north of the country. In the fjord areas local topographic features have a predominant east west orientation, and hence the first eigenvector projects strongly onto this area.

These projections will be used as predictands for an empirical model of mean monthly temperature (to be used with the standard predictands of location (position and altitude) and distance from coast, see [1] for details). As such, they can be compared with those obtained in [4] where 5 physically based predictands

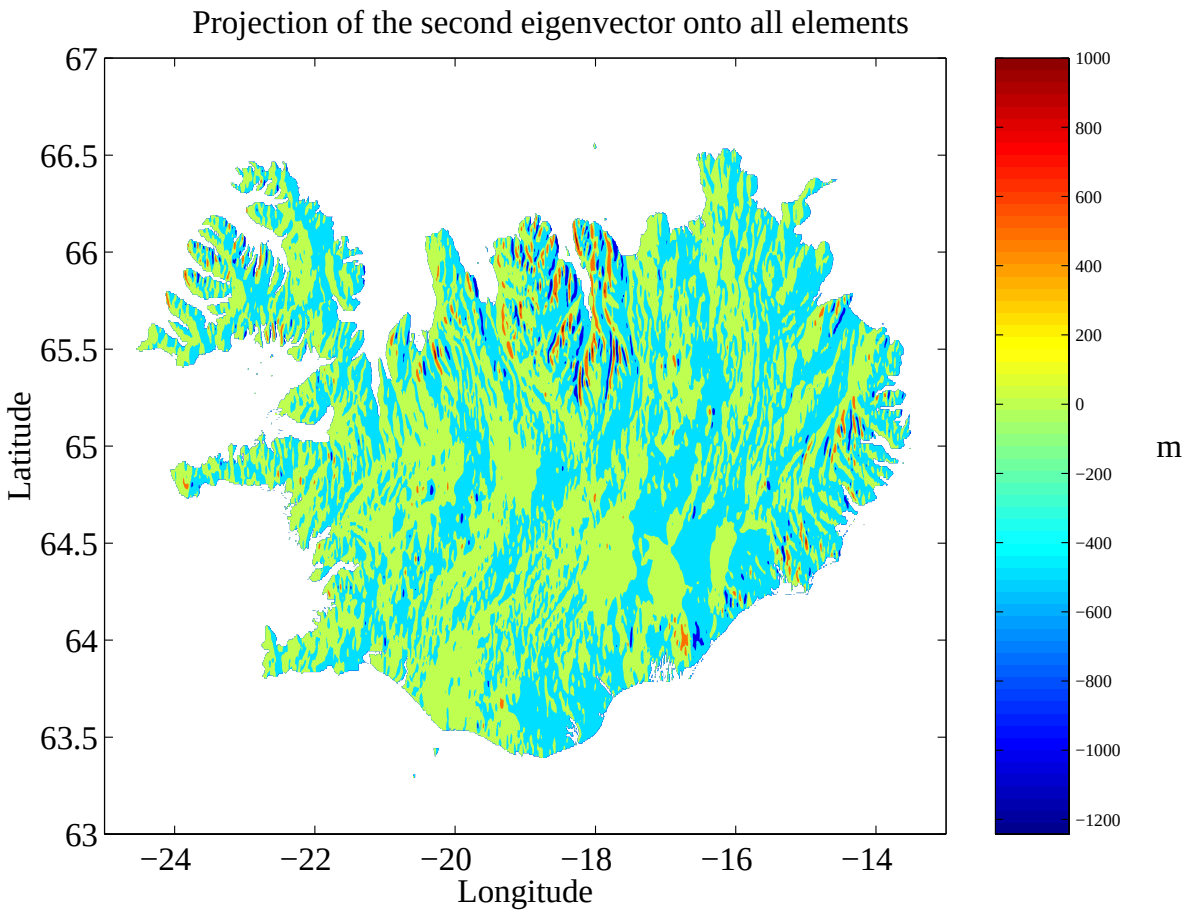

Figure 7: The projection of the second eigenvector onto the centered element matrix.

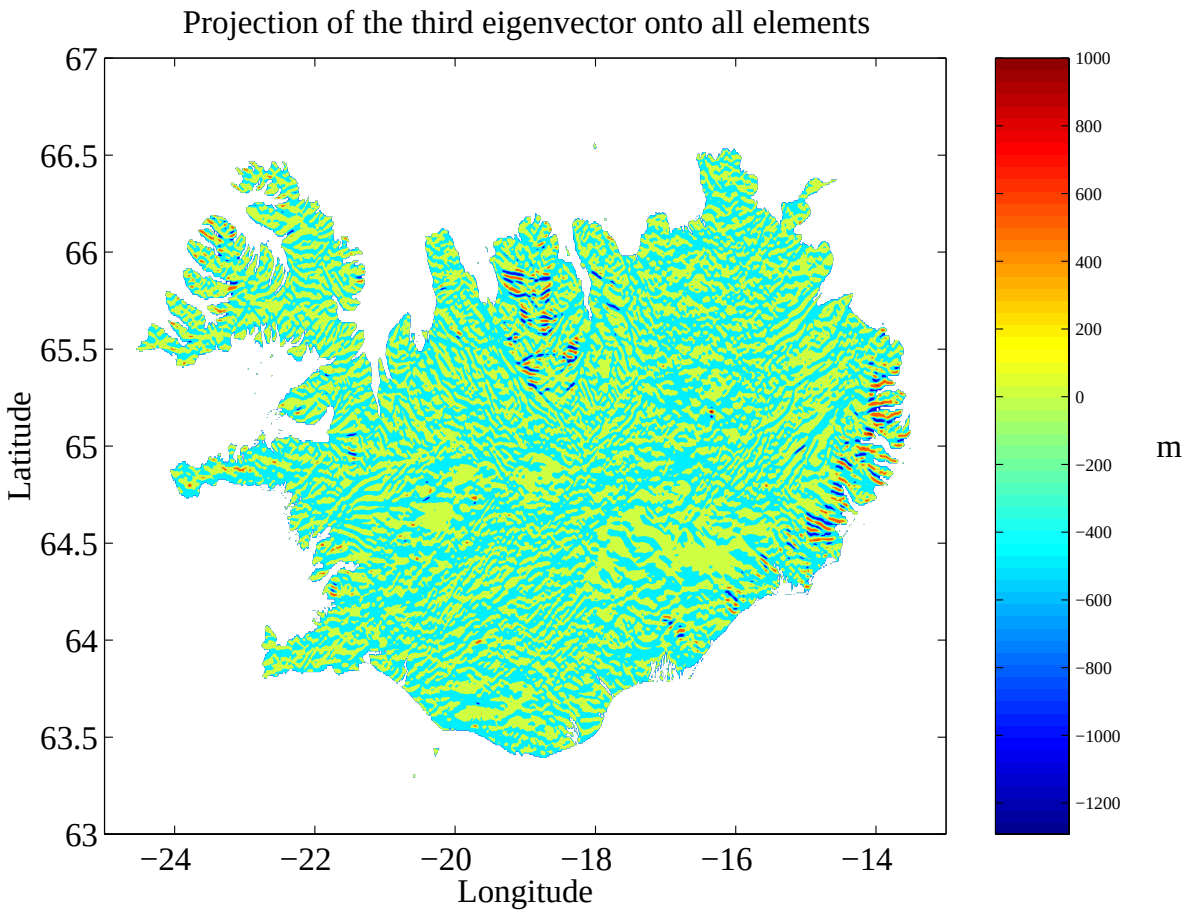

Figure 8: The projection of the third eigenvector onto the centered element matrix.

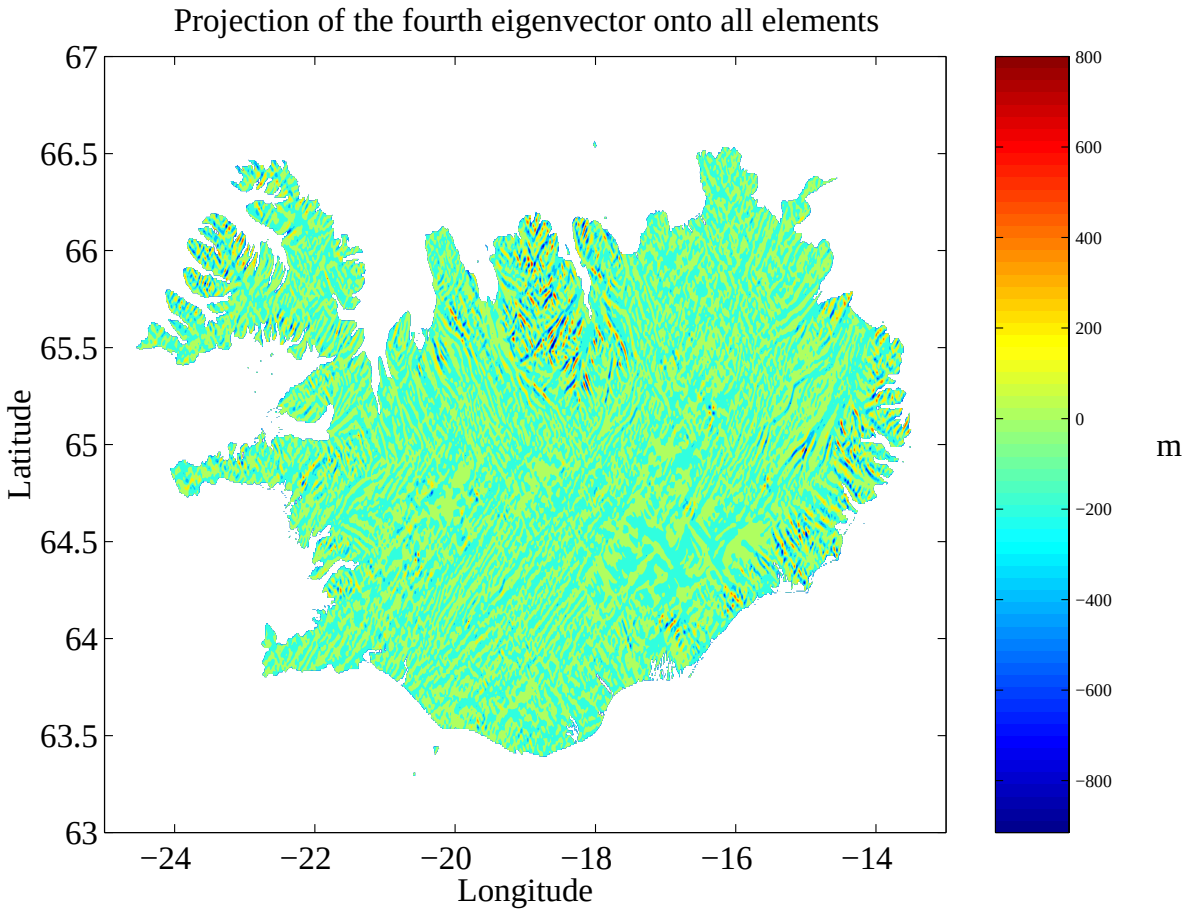

Figure 9: The projection of eigenvector 4 onto the centered element matrix.

for an empirical model of precipitation were calculated. These were based on a) standard deviation of local elevation, b) steepness of topography, c) slope orientation, d) standard deviation of the slope orientation and e) the greatest difference elevation within each 5 by 5 km element. Three of these 5 predictands (a,b,e) loosely resemble the eigenvector projections in figures $6-9$ in that the amplitude is largest in similar areas of the country. The predictands having to do with slope orientation and standard deviation (c,d) do not resemble any of the eigenvector projections above.

In figures $6-9$ the difference between the maximum and minimum values (the range of the figure) is reduced as we go to higher eigenvectors (thus is largest for 6 but smallest for figure 9). This is due to the fact that the variance associated with each eigenvalue is reduced as we go to higher eigenvectors. Although not crucial for regression purposes it is still preferable that this difference be similar in all the four predictands produced by this method. In accordance with the methodology described in appendix B, we therefore chose to divide each projection with the square root of the corresponding eigenvalue. This yields predictand patterns that all have similar range.

References

[1] Gylfadóttir S.S. *Spatial interpolation of Icelandic monthly mean temperature data* Veðurstofa Íslands (2003) Report 03006

[2] Bjornsson, H. and Venegas, S. A. *A Manual for EOF and SVD analyses of Climatic Data*. Center for Climate and Global Change Research, McGill University, Montreal. (1997) C^2GCR Report 97-2.

[3] Wotling, G., Bouvier, Ch., Danloux, J. and Fritsch. *Regionalization of extreme precipitation distribution using the principal components of the topographical environment.* J. Hydrology (2000), **233**:86-101.

[4] Crochet, P. *A linear model for mapping precipitation in Iceland.* Veðurstofa Íslands (2002) Report 02028.

A Monte Carlo estimation of the significance of eigenvalues

Assessing the significance level of a quantity using Monte Carlo methods is usually performed in the following manner: First surrogate data is generated, next the quantity of interest is calculated from the surrogate data. This is repeated enough times to get a distribution of the quantity of interest. The original value of the quantity is then compared with the distribution and the significance level estimated.

Here, two methods were used to generate the surrogate data. First we generated surrogate data by shuffling the element matrix and centering it. The eigenvalues were then calculated and stored. This process was repeated 100 times and following that the distribution of values obtained for each eigenvalue was examined. In general the surrogate data yielded covariance matrices with

uniform eigenvalues, each explaining about 4% of the data. This is not surprising since this method generates random surrogate covariance matrices R_s. Such matrices do not have any directional preference, and thus for a random covariance matrix of size N, one can expect each eigenvalue to explain $(100/N)\%$ of the variance.

The second method consisted of selectively reordering the original covariance matrix R, in such a fashion that the reordered one was still a quadratic (positive definite) matrix. The simplest way to do this is to first calculate the matrix square root of R, i.e., find S so that

$$S^t S = R.$$

The MC experiment consists of randomly shuffling S, reforming the covariance matrix and calculating the eigenvalues. We did this 1000 times, and following that, we examined the distribution of each eigenvalue. Since this method better preserves directional information inherent in R it yields eigenvalues that are not uniformly distributed. Figure 4 shows the 95% level for the distribution of each eigenvalue.

From the figure it is clear that eigenvalues 1,2 and 3 (that explain 52%,17% and 12% of the total variance, respectively) are all judged significant by both methods. Eigenvalue 4 (explaining 6% of the variance) fails according to the second method, and all other eigenvalues fail according to both methods. In what follows we will accordingly, focus our attention on the first four eigenvalues only (which combined explain 87% of the variance), keeping in mind that of those the fourth one is the least significant one.

B Eigenvectors of the covariance matrix

As noted earlier the matrix formed to calculate the eigenvalues $R = MM^t$ is not a true covariance matrix, since M is centered on the columns, not the lines. However, as the size of M is $(25 \times 259, 104)$ the size of the true covariance matrix

$$K = M^t M$$

is $(259, 104 \times 259, 104)$, which is too large for most eigenvalue solvers.

In practice, it is not neccessary to solve for the eigenvalues of K, since it is easy to show (e.g. using singular value decomposition, see [2] for details) that R and K have the same eigenvalues. The eigenvectors of K can be found by calculating the projection of the eigenvectors of R onto the centered element matrix M, and scaling the resuls. The scaling factor for each eigenvector is given by the square root of the corresponding eigenvalue of R (and of K). Thus the four dominant eigenvectors of K are represented by maps identical to those in figures $6 - 9$, apart from the scaling.

The maps shown in these figures differ considerably in the their *range* (i.e., the difference between maxima and minima, easily read of the color-bar axis of each figure). This scaling has the effect of reducing the changes in range that occur from one map to another. Thus the range of the map in figure 6 is about 4000 while the range of the map in figure 9 is about 1750. With each map scaled with the square root of the corresponding eigenvalue, the range of both maps becomes close to 0.03.