Vol. 20: 271–279, 2002 | CLIMATE RESEARCH Clim Res | Published April 26

Simplified EOFs—three alternatives to rotation

I. T. Jolliffe[1,*], M. Uddin[2], S. K. Vines[3]

[1]Department of Mathematical Sciences, University of Aberdeen, Meston Building, King's College, Aberdeen AB24 3UE, Scotland, United Kingdom

[2]Department of Statistics, University of Karachi, Karachi 75270, Pakistan

[3]Department of Statistics, Faculty of Mathematics and Computing, The Open University, Walton Hall, Milton Keynes MK7 6AA, United Kingdom

ABSTRACT: Principal component analysis (PCA) is widely used in atmospheric science, and the resulting empirical orthogonal functions (EOFs) are often rotated to aid interpretation. In this paper 3 methods are described which provide alternatives to the standard 2-stage procedure of PCA followed by rotation. The techniques are illustrated on a small example involving sea-surface temperatures in the Mediterranean. Each method is shown to give different simplified interpretations for the major sources of variation in the data set. All 3 techniques have advantages over standard rotation.

KEY WORDS: EOFs · Simplification · Interpretation · Rotation · LASSO · Mediterranean · SST

1. INTRODUCTION

Principal component analysis (PCA) is primarily a dimension-reduction technique, which takes observations on p correlated variables and replaces them by uncorrelated variables. These uncorrelated variables, the principal components (PCs), are linear combinations of the original variables, which successively account for as much as possible of the variation in the original variables. Typically a number m is chosen ($m << p$) such that using only the first m PCs instead of the p variables will involve only a small loss of variation. The vectors of loadings which define the PCs are known as empirical orthogonal functions (EOFs).

One problem with using PCA to replace p variables by m PCs, rather than the alternative strategy of replacing the p variables by a subset of m of the original variables, is interpretation. Each PC is a linear combination of all p variables, and to interpret a PC it is necessary to decide which variables are important, and which are unimportant, in defining that PC.

It is common practice in atmospheric science (Richman 1986) to rotate the EOFs. This idea comes from factor analysis, and proceeds by post-multiplying the ($p \times m$) matrix of PC loadings by a (usually orthogonal) matrix to make the loadings 'simple'. Simplicity is defined by 1 of a number of criteria (e.g. varimax, quartimax) which quantify the idea that loadings should be near zero or near ±1, with as few as possible intermediate values. Rotation takes place within the subspace defined by the first m PCs. Hence all the variation in this subspace is preserved by the rotation, but it is redistributed amongst the rotated components, which no longer have the successive maximisation property of the unrotated PCs.

Rotation of PCs has a number of drawbacks (Jolliffe 1987, 1989, 1995, Jolliffe & Law 1995). We defer detailed discussion of these until later in the paper, when we compare the properties of rotated PCs with the components produced by 3 new techniques. The first of these, which we call the simplified component technique (SCoT) has 1 stage, rather than the 2 needed in rotated PCA, and each simplified component optimises a criterion combining the desirable properties of large variance and simplicity. Successive components are constrained to be orthogonal to, or uncorrelated with, one another. Following a summary of PCA and rotation, SCoT is defined in Section 2.

*E-mail: itj@maths.abdn.ac.uk

The second technique is based on an idea introduced in regression by Tibshirani (1996), and known as the LASSO (least absolute shrinkage and selection operator). An important feature of this method is that some of the regression coefficients are exactly zero. An adaptation of the LASSO to PCA, which we call the Simplified Component Technique—LASSo (SCoTLASS) is described in Section 3.

The third technique, simple components analysis (SCA) is introduced in Section 4. SCA uses a different, more subjective, concept of simplicity to the other methods: simple components are those whose elements are proportional to integers, particularly if the integers are small in magnitude. Like SCoTLASS, SCA can produce loadings that are exactly zero, but, unlike the other methods, it can also give components in which all the p variables are identically weighted.

Section 5 applies all 3 new methods to a small data set involving Mediterrean sea-surface temperature (SST). Comparisons of the results of each technique for the SST data, amongst themselves and with unrotated and rotated PCA are expanded in Section 6 to a more general discussion of some properties and computational aspects of the methods.

2. THE SIMPLIFIED COMPONENT TECHNIQUE

2.1. Principal component analysis

To establish notation we formally define PCA. Let $\boldsymbol{x}$ be a vector of p random variables with covariance matrix Σ. The kth principal component, for $k = 1, 2, \ldots, p$, is the linear function of $\boldsymbol{x}$, $\boldsymbol{\alpha}'_k\boldsymbol{x}$, which maximises $\mathrm{var}(\boldsymbol{\alpha}'_k\boldsymbol{x}) = \boldsymbol{\alpha}'_k\Sigma\boldsymbol{\alpha}_k$ subject to $\boldsymbol{\alpha}'_k\boldsymbol{\alpha}_k = 1$, and, (for $k > 1$), $\boldsymbol{\alpha}'_h\boldsymbol{\alpha}_k = 0$, $h < k$. The vectors $\boldsymbol{\alpha}_k$, $k = 1, 2, \ldots, p$, are the EOFs.

In practice, we work with a $(n \times p)$ data matrix, $\mathbf{X}$, and the *sample* covariance matrix, $\mathbf{S}$, for these data. We denote the value of the kth PC for the ith observation (row of $\mathbf{X}$) by $\boldsymbol{a}'_k\boldsymbol{x}_i$ in the sample case. Often PCA is done for variables standardised to each have unit variance, so that Σ or $\mathbf{S}$ becomes the correlation matrix for the original variables.

2.2. Rotation

Suppose that we have decided to retain and rotate m PCs and that $\mathbf{A}$ is the $(p \times m)$ matrix whose kth column is the kth EOF, $\boldsymbol{a}_k$, $k = 1, 2, \ldots, m$. Rotating the first m EOFs or components is achieved by post-multiplying $\mathbf{A}$ by a matrix $\mathbf{T}$, to obtain rotated loadings $\mathbf{B} = \mathbf{AT}$. The choice of $\mathbf{T}$ is determined by whichever rotation criterion we use. For example, for the commonly-used varimax rotation criterion, $\mathbf{T}$ is an orthogonal matrix chosen to maximise:

$$S(\mathbf{B}) = \frac{1}{p^2}\sum_{k=1}^{m}\left[p\sum_{j=1}^{p}b_{jk}^4 - \left(\sum_{j=1}^{p}b_{jk}^2\right)^2\right] \quad (1)$$

where b_{jk} is the (j,k)th element of $\mathbf{B}$ (Krzanowski & Marriott 1995).

The idea behind the varimax criterion is to simplify the structure of the loadings by maximising the variance of squared loadings within each column of $\mathbf{B}$. This drives the loadings towards zero or ± 1; alternative rotation criteria attempt to achieve similar objectives, but using a variety of other specific definitions. Richman (1986) describes a large number of possible criteria. Concentrating the loadings close to zero or ± 1 is not the only possible definition of simplicity. For example, if all loadings are equal, that is also simple but in a completely different way. Building a criterion to take into account both types of simplicity is difficult and hence none of the techniques we describe below try to incorporate both.

2.3. The simplified component technique

SCoT reduces the 2 stages of rotated PCA into 1 step in which we successively attempt to find linear combinations of the p variables that maximise a criterion which combines variance with a penalty function that pushes the linear combination towards simplicity. Let $\boldsymbol{c}'_k\boldsymbol{x}_i$ be the value of the kth simplified component (SC) for the ith observation and suppose, for example, that we define simplicity in terms of the varimax criterion $S(\boldsymbol{c}_k)$, as defined in Eq. (1), but for the special case where $m = 1$, and scaled so that its maximum value is 1. If $V(\boldsymbol{c}_k) = \mathrm{var}(\boldsymbol{c}'_k\boldsymbol{x})$, then the SCoT successively maximises:

$$(1-\psi)V(\boldsymbol{c}_k) + \psi S(\boldsymbol{c}_k) \quad (2)$$

subject to $\boldsymbol{c}'_k\boldsymbol{c}_k = 1$ and (for $k > 1$) $\boldsymbol{c}'_h\boldsymbol{c}_k = 0$, $h < k$. Here ψ is a simplicity/complexity parameter, which needs to be chosen. For $\psi = 0$ we have PCA, and as ψ increases the SCs move away from the PCs towards greater simplicity. When $\psi = 1$, each SC is identical to one of the original variables, with zero loadings for all other variables. The choice of ψ will be discussed later.

3. MODIFIED PCA BASED ON THE LASSO

Tibshirani (1996) considered the difficulties involved in the interpretation of multiple regression equations with many predictor variables. These problems may occur due to the instability of the regression coefficients in the presence of collinearity, or simply because of the large number of variables included in the regression equation. Some current alternatives to least squares regression handle the instability problem by keeping all variables in the equation but shrinking some loadings towards zero, whereas variable selection procedures find a subset of variables and keep only the selected variables in the equation. Tibshirani (1996) proposed a new method, LASSO, which is a compromise between variable selection and shrinkage estimators. The procedure shrinks the coefficients of some of the variables not simply *towards* zero, but *exactly* to zero, giving an implicit form of variable selection. Here we adapt the LASSO idea to PCA.

3.1. The LASSO approach in regression

In standard multiple regression we have:

$$y_i = \alpha + \sum_{j=1}^{p} \beta_j x_{ij}, \quad i = 1, 2, \ldots, n \tag{3}$$

where $y_1, y_2, \ldots, y_n$ are measurements on a response variable y; x_{ij}, $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, p$, are corresponding values of p predictor variables; and α and $\beta_1, \beta_2, \ldots, \beta_p$ are parameters in the regression equation. In least squares regression, these parameters are estimated by minimising the residual sum of squares:

$$\sum_{i=1}^{n}\left(y_i - \alpha - \sum_{j=1}^{p}\beta_j x_{ij}\right)^2 . \tag{4}$$

The LASSO imposes an additional restriction on the coefficients, namely:

$$\sum_{j=1}^{p} |\beta_j| \leq t \tag{5}$$

for some 'tuning parameter' t. For suitable choices of t, this constraint has the interesting property that it forces some of the coefficients in the regression equation to zero.

3.2. The LASSO approach in PCA (SCoTLASS)

As noted above, PCA finds linear combinations, $\boldsymbol{a}'_k\boldsymbol{x}$ ($k = 1, 2, \ldots, p$), of the p measured variables $\boldsymbol{x}$ which successively have maximum variance $\boldsymbol{a}'_k \mathbf{S} \boldsymbol{a}_k$, subject to $\boldsymbol{a}'_k \boldsymbol{a}_k = 1$ and (for $k \geq 2$) $\boldsymbol{a}'_h \boldsymbol{a}_k = 0$, $h < k$.

The proposed method of LASSO-based PCs (SCoTLASS) also solves this maximisation problem, but with the *extra* constraints:

$$\sum_{j=1}^{p} |a_{kj}| \leq t \tag{6}$$

for some tuning parameter t, where a_{kj} is the jth element of the kth vector $\boldsymbol{a}_k$, ($k = 1, 2, \ldots, p$).

3.3. Some properties

SCoTLASS differs from PCA in the inclusion of the constraints defined in Eq. (6), so a decision must be made on the value of the tuning parameter, t. It is easy to see that (1) for $t \geq \sqrt{p}$, we get PCA; (2) for $t < 1$, there is no solution; (3) for $t = 1$, we must have exactly 1 non-zero α_{kj} for each k.

As t decreases from $\sqrt{p}$, we move progressively away from PCA and eventually to a solution where only 1 variable has a non-zero loading on each component, as with $\psi = 1$ in SCoT. Loadings for all other variables will shrink (not necessary monotonically) with t and ultimately reach zero. Examples of this behaviour are given in Section 5, and there will be further discussion of the choice of t in Section 6.

4. SIMPLE COMPONENTS

The iterative approach at the heart of the SCA algorithm is substantially different to that of SCoT and SCoTLASS. The algorithm starts with a set of p vectors of loadings chosen for their particular simpleness without regard to the variances of the associated components. Typically this is the set of vectors $\boldsymbol{a}_k$ where $a_{kk} = 1$ and $a_{kj} = 0$ ($k = 1, 2, \ldots, p$; $j = 1, 2, \ldots, p$; $j \neq k$), a_{kj} being the jth element of $\boldsymbol{a}_k$. A series of 'simplicity-preserving' transformations are applied to these vectors. Each transformation affects just a pair of the vectors, orthogonally rotating them so that the variance associated with the already higher variance component is increased. The algorithm finishes when no non-trivial simplicity-preserving transformation leads to an improvement in variance.

In general, the elements of the transformed vectors after orthogonal rotation could be any real numbers between −1 and +1. To achieve 'simple' components, only a restricted set of angles is considered for each rotation, namely those for which the elements of the transformed vectors are proportional to integers, as were the elements of original loading vectors. These 'preserve simplicity' if we define 'simple' vectors as those whose elements are proportional to integers. Furthermore the transformed vector associated with the higher variance tends to be simpler (proportional to smaller magnitude integers) than the other transformed vector. Cumulatively, this means that when the algorithm halts, all the vectors of loadings are simple, with those for the first few components tending to be simpler than those for the others.

The performance of the SCA algorithm is controlled by a tuning parameter, c. For any given value of $c = 0, 1, 2, \ldots$, the number of angles considered for each simplicity preserving transformation is 2^{c+2}. As c

increases, the simple components generated tend to be become closer to the PCs. However this closer approximation is at the cost of subjective simplicity, as the elements of the vectors also tend to become proportional to larger magnitude integers as c increases. In practice it has been found that $c = 0$ usually gives the best balance between simplicity and large variances for the components, and consequently $c = 0$ was used for examples in the next section.

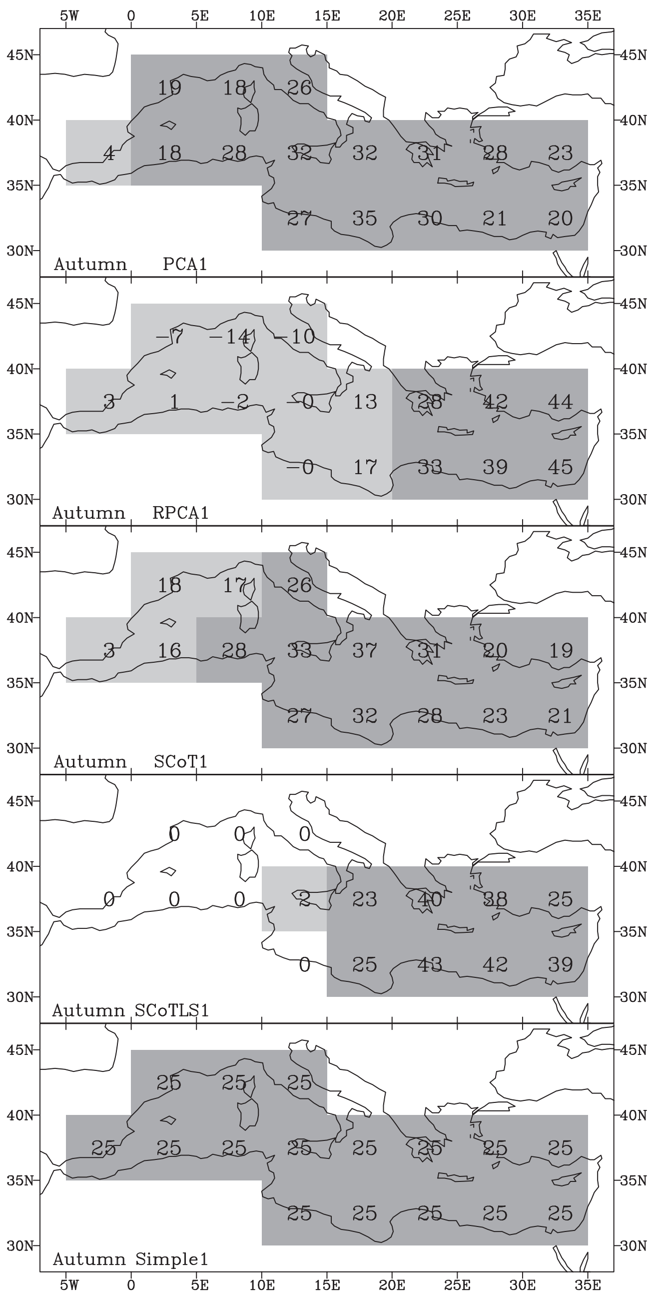

Fig. 1. Loadings of first autumn components for PCA (principal component analysis), RPCA (rotated PCA), SCoT (simplified component technique), SCoTLASS (LASSO [least absolute shrinkage and selection operator] approach in PCA) and SCA (simple components analysis). Dark shading: loadings whose values are greater than 50%, in absolute value, of the largest loading for that component; lighter shading: smaller but non-zero loadings; unshaded areas: exactly zero loadings. The loadings themselves (×100) are included

5. AN EXAMPLE—MEDITERREAN SSTS

The data examined in this section were originally analysed by Bartzokas et al. (1994), using PCA and rotated PCA. Jolliffe & Law (1995) demonstrated how the choice of different normalization constraints affects the results of rotated PCA for the same data set. The data consist of values of SST, averaged over 3 mo seasons, for the years 1946–1988, and for each of sixteen $5° \times 5°$ grid boxes covering most of the Mediterranean. Thus p is 16, and, for each season, there are measurements for 43 yr. Here we analyse the autumn and winter data sets.

Each of Figs. 1–4 displays the loadings in 5 different varieties of 'components', from PCA, RPCA, SCoT, SCoTLASS and SCA. In all cases the vectors of loadings are normalized to have unit lengths. This makes comparisons between techniques easier, but disguises the simple integer structure produced by SCA. Figs. 1–4 give the first and second components for autumn (Figs. 1 & 2 respectively) and for winter (Figs. 3 & 4). For all but RPCA, the order of components is well defined. For RPCA, 2 of the rotated components obtained by rotating the first 3 PCs are displayed. For autumn, the 2 out of 3 with the largest variances are presented. The choice for winter was less clear-cut. RPC1 has a clearly larger variance than the other 2, but the latter have very similar variances (3.96, 3.94). The RPC with the slightly smaller variance is presented as RPC2 because it has a larger score with respect to the varimax criterion and matches better with RPC2 for autumn. The loadings differ from those of Bartzokas et al. (1994) because a different normalization, in which EOFs have unit length, is used in the present case (see Jolliffe & Law 1995), and because, for winter, 3 PCs have been rotated, rather than the 4 in Bartzokas et al. (1994).

First consider Figs. 1 & 2. The first 2 PCs and rotated PCs take characteristic forms for a fairly small spatial region. The first PC has positive loadings in all grid boxes, and the second PC is, necessarily because of the orthogonality constraint, a contrast between 2 sub-regions, in this case the Eastern and Western Mediterranean. Rotated PCA gives components which correspond to 2 separate sub-regions. One component has its largest loadings in the east of the Mediterranean, the other slightly west of the centre. There has been extensive debate in the literature on the merits or otherwise of rotation—see, for example, Jolliffe (1987) and Richman (1987). This will not be discussed further here, except to say that both unrotated and rotated PCA have advantages and disadvantages. The choice between them depends on which of these pros and cons are most important in the context of the objectives of an analysis.

Returning to the example, we have given initial interpretations of the first 2 PCs and RPCs, but closer inspection shows that they are not quite that simple. The loadings in PC1, for example, are not uniform; they range from 0.04 to 0.35. In the same way, the loadings outside the Central Mediterranean are not zero for RPC2; there are areas in the east and west with loadings −0.16 and −0.20. Similar comments can be made about PC2 and RPC1.

Of our 3 alternatives, SCA certainly lives up to its name for the autumn data. The first component has all

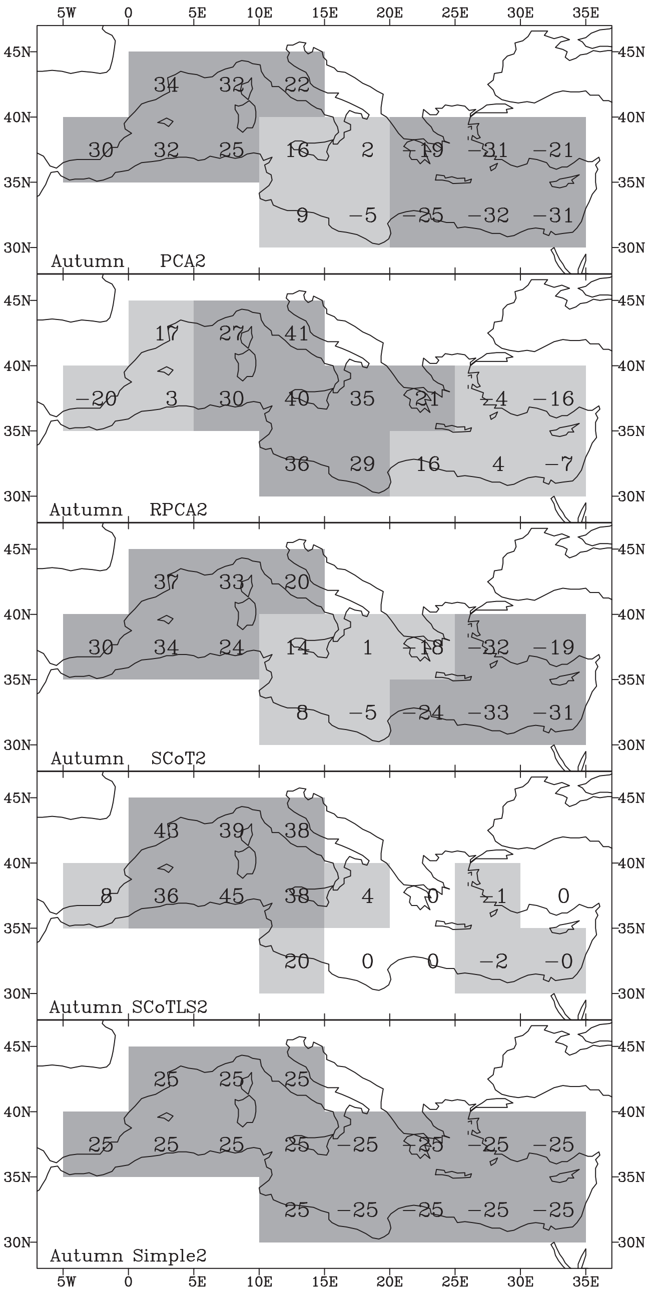

Fig. 2. Loadings of second autumn components for PCA, RPCA, SCoT, SCoTLASS and SCA. Shading as in Fig. 1

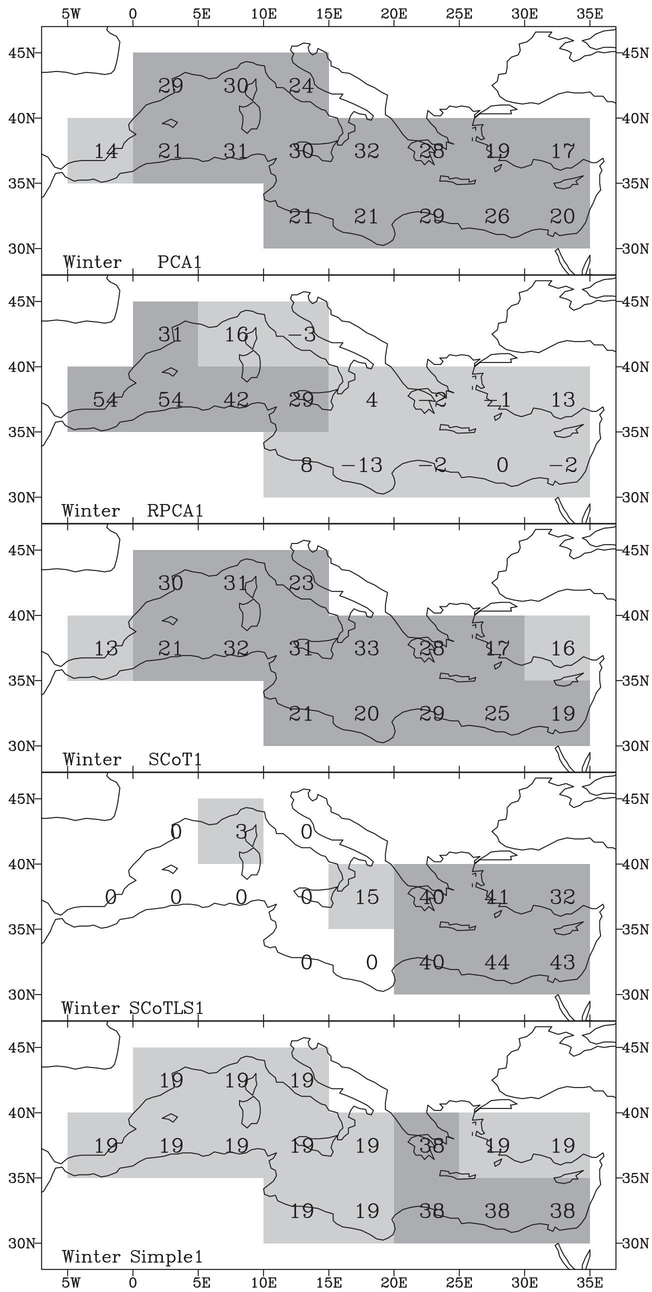

Fig. 3. Loadings of first winter components for PCA, RPCA, SCoT, SCoTLASS and SCA. Shading as in Fig. 1

its loadings equal (to 1, when expressed as integers), and the second is a straightforward contrast between east and west, with all loadings equal to +1 or –1 when expressed as integers. Both can be viewed as simplified versions of PC1 and PC2.

SCoT1 and SCoT2 can also be viewed as simple versions of PC1 and PC2, though the simplicity is of a different form. The largest loadings become larger and the smallest become smaller, as is sometimes observed

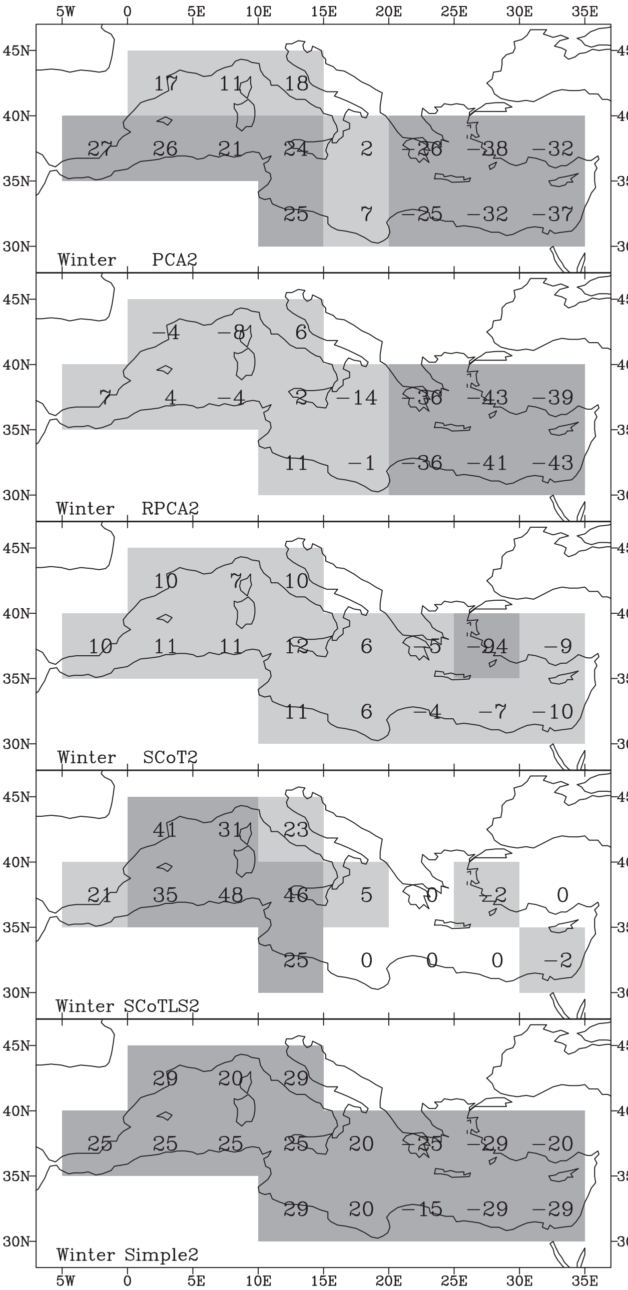

Fig. 4. Loadings of second winter components for PCA, RPCA, SCoT, SCoTLASS and SCA. Shading as in Fig. 1

during rotation. Finally, SCoTLASS behaves like RPCA, except that its patterns are more clear-cut, with several grid boxes having zero, rather than small, loadings. Its first component concentrates on the Eastern Mediterranean (like RPCA), but its second component is centred a little further west than RPC2.

It is clear that, in different ways, each of the 3 techniques simplifies the results of PCA or RPCA. The price paid for this simplification is a reduction in the proportion of total variation accounted for. A defining property of PCA is that it successively maximizes this variation, so any other technique will be inferior in this respect. Whether or not simplification is deemed worthwhile will depend on how much variation is lost. In the present example, the first 2 PCs account for 78.2% of total variation. This drops to 65.8, 78.0, 70.5 and 70.1% for RPCA, SCoT, SCoTLASS and SCA, respectively. The drop for RPCA is because rotation is within a space of more than 2 dimensions. If only the first 2 PCs were rotated, the total percentage of variation accounted for would be the same after rotation as before.

Turning to the winter data (Figs. 3 & 4), the patterns are similar to autumn, though a little more complex. The first 2 PCs and RPCs are not too different in the 2 seasons, though RPC2 is centred in the Western rather than Central Mediterranean. SCoTLASS also has similar patterns in autumn and winter, but there are bigger differences for SCoT and for SCA. SCoT2 is very much dominated by a single grid box in the Eastern Mediterranean, whilst the SC1 and SC2 are less simple in winter. SC1 is no longer uniform—4 grid boxes have loadings of 2, compared to 1 for the remaining 12 boxes—and SC2 has loadings whose absolute values are proportional to 3, 4, 5 and 6, rather than the simple ±1 of autumn. For winter the first 2 PCs account for 71.0% of the total variation, and this falls to 57.4, 55.8, 65.4 and 67.4% for RPCA, SCoT, SCoTLASS and SCA respectively.

6. DISCUSSION

The descriptions above of the 3 new techniques are fairly brief. Further details and discussion can be found in Jolliffe & Uddin (2000), Vines (2000) and Joliffe et al. (unpubl.) which also include a number of examples from outside atmospheric science. We conclude with a few remarks resulting from our experience in using the techniques on a variety of examples, as well as on simulated artificial data.

Remark 1. It is of interest to examine the geometry of the new techniques and compare it with that of the established methods. In a written paper only 2 dimensions are easily visualized, so we restrict attention to that case. Fig. 5 shows the constraints imposed on $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ in 2 dimensions by SCoTLASS for the limiting cases $t = 1$ and $t = \sqrt{p}$ (indicated by solid squares) and the non-trivial case where $1 < t < \sqrt{2}$ (dashed square), together with the unit circle corresponding to the constraints $\boldsymbol{a}'_k\boldsymbol{a}_k = 1$. A SCoTLASS solution should lie on the unit circle and within the relevant square. If the correlation coefficient between the 2 variables is positive, the first PC defines a direction in the first (top right) quadrant, and the direction of the second PC is at right angles to the first. If the square corresponding to SCoTLASS's constraint lies outside the unit circle in the direction of the first PC, the first SCoTLASS component will be identical to the first PC. If t is decreased so that the square lies inside the unit circle in that direction, then SCOTLASS's optimal direction moves towards one the axes (that is towards simplification) and corresponds to the direction where the circle and square intersect.

Ordinary rotation also rotates from the direction of the first PC towards one of the axes. If the normalization $\boldsymbol{a}'_k\boldsymbol{a}_k = 1$ is used and all p components are rotated, then the varimax method rotates exactly to the original axes. Note, however, that with different normalizations (see Jolliffe 1995) varimax rotation gives different solutions. SCoT loses variance and increases simplicity compared to PCA and hence also rotates towards the axes. As $\psi \rightarrow 1$, so the SCoT directions approach the axes. Finally, simple components allow only a finite set of rotations *away* from their initial directions which are along the axes themselves. The rotation seeks to maximize variance amongst the allowed rotations, but as the first PC is not usually one of the allowed simple directions, rotation will not be as far this direction.

Remark 2. Both SCoT and SCoTLASS require the choice of a value for a tuning parameter, t or ψ respectively. To save space we have presented the results for only 1 value of t and ψ in Section 5. In practice, a user of either technique will need to examine results for at least 2 or 3 values of the tuning parameter, and eventually choose one (or possibly more) that provides a suitable compromise between simplicity and loss of variance. There is, at present, no way to choose the value of the parameter automatically, and it would be difficult to devise one. The choice of t in Figs. 1–4 seems to achieve the compromise mentioned above quite nicely. The choice of ψ is trickier. As ψ increases, there is often a rapid jump from simplified components which are very close to the corresponding PC, as with both components in autumn, to those which are dominated by a single variable, as exemplified by SCoT2 in winter. This is usually accompanied, as seen in the example above, by a large drop in the proportion of variation accounted for, compared to PCA. Choosing different values of ψ for SCoT1, SCoT2, ..., only partially alleviates this problem and, mainly for this reason, SCoT seems to be the least promising of the 3 new techniques.

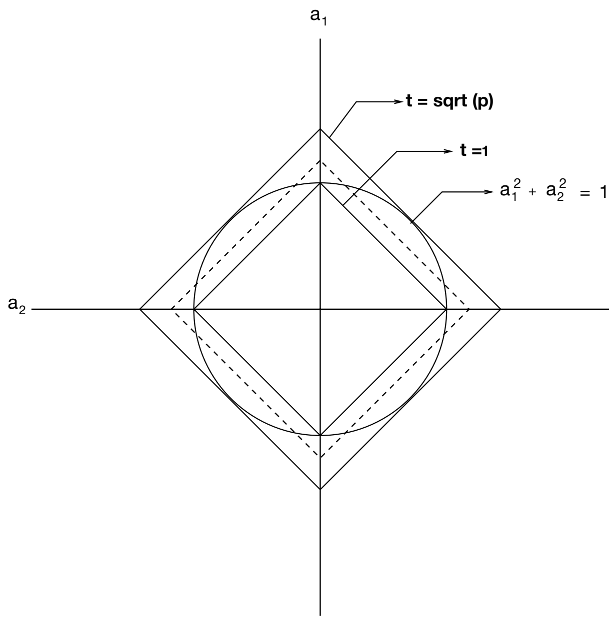

Fig. 5. Geometry of SCoTLASS when $p = 2$

SCA also requires the choice of a tuning parameter, c. Like t and ψ, there is also no way of choosing the value of c automatically. However the choice of c is more straightforward than t and ψ. Firstly, c is discrete, not continuous. Secondly, in our experience so far there is little to be gained from choosing anything other than $c = 0$.

Remark 3. PCs have the special property that the EOFs are orthogonal and the components are uncorrelated. RPCs lose at least one of these properties, depending on which normalization constraint is used (Jolliffe 1995, Mestas-Nuñez 2000). None of the new techniques is able to retain both properties either. SCoT, SCoTLASS and SCA, as defined above, and implemented in the example, all retain orthogonality of the loading vectors, but sacrifice the uncorrelatedness of the components. For SCoT and SCoTLASS, it is straightforward, though a little more complicated computationally, to implement versions that keep uncorrelatedness, rather than orthogonality. All that is required is to substitute the conditions $\boldsymbol{c}'_h\mathbf{S}\boldsymbol{c}_k = 0$, $h < k$ (or $\boldsymbol{a}'_h\mathbf{S}\boldsymbol{a}_k = 0$, $h < k$) for $\boldsymbol{c}'_h\boldsymbol{c}_k = 0$, $h < k$ (or $\boldsymbol{a}'_h\boldsymbol{a}_k = 0$, $h < k$) in the definitions of the techniques in Sections 2 and 3.

Remark 4. An alternative strategy to simplification is to replace the original p variables by a subset of variables, rather than by a small number of PCs. There has been work on this problem in the statistical literature—see, for example, Cadima & Jolliffe (2001), who discuss 2 main strategies for doing so, and also make the often-neglected point that relative sizes of loadings do not always reflect the importance of each variable in a component. SCoTLASS has an implicit variable selection by setting some of the loadings to zero for sufficiently small values of *t*. Another form of variable selection (empirical orthogonal teleconnections) has been suggested recently in the climate literature (van den Dool et al. 2000).

Despite being an old and well-established procedure, there is still much active research on PCA and related techniques. As a recent example, Monahan (2000) described a non-linear version of PCA based on neural networks, and other non-linear extensions of PCA have been discussed by Tenenbaum et al. (2000) and Roweis & Saul (2000). These ideas are not explicitly concerned with simplification, although if the underlying structure is simple but non-linear, then a non-linear technique has a better chance of finding this simple structure than the usual linear PCA.

Remark 5. Ordinary 2-stage rotation of EOFs has a number of drawbacks, including (1) dependence on the choice of normalization constraint before rotation; (2) sensitivity to the number of EOFs rotated; and (3) loss of the sequential optimization property of PCA.

The nature of the new techniques means that all 3 of these difficulties are avoided, though the first is replaced by a choice between orthogonal EOFs and uncorrelated components. The gains here are paid for by a loss in variation accounted for, though, as seen in the example, the reduction is not great, in the case of SCA and SCoTLASS, for solutions which provide considerable simplification. One way to quantify simplicity is to calculate the value of the varimax criterion (or whatever other criterion is used in RPCA) for any component derived from any of the methods. It has been found that SCoT and SCoTLASS often do better than 2-stage rotation with respect to the latter's own simplicity criterion, with only a small reduction in variance, giving another reason for preferring the new techniques over rotation. In contrast, for the autumn data, the clearly simple SC1 gives the worst possible value of the varimax criterion. Thus SCA is capable of producing very simple vectors of loadings, with little loss in variation, that would be completely missed by varimax-based rotation.

Remark 6. PCA is a straightforward optimization problem, reducing to finding eigenvalues and eigenvectors of a symmetric positive semi-definite matrix, a problem for which there is plenty of efficient software. The optimization problems posed by the 3 new techniques are computationally harder and more time-consuming, and at present the SST example is near to the upper limit of the number of variables which can be handled. We are confident that more efficient algorithms can be devised which will speed up the computations and make the methods practicable for much larger data sets, but further research is needed. Once these computational hurdles are surmounted, the techniques have great potential as simplifying tools in the many cases where EOFs are used in climate research. The main problem associated with SCoT and SCoTLASS is the existence of multiple optima. A number of optimization methods were tried, and the current implementations are now described briefly.

For SCoT we used the built-in quasi-Newton function of S-PLUS, interfaced with modified FORTRAN functions based on algorithms described by Gay (1983, 1984). To ensure that the criterion values for SC1, SC2, SC3, … were in decreasing order, and to avoid reporting local optima as far as possible, the problem was solved for each dimension by replicating the optimization 11 times for different randomly chosen starting values. The maximum value of the objective function and the corresponding vectors thus obtained were selected. Adopting this strategy, the convergence is quicker than the simulated annealing method. Another important aspect of SCoT is the implementation of orthogonality or uncorrelatedness. In SCoT, the Gram-Schmidt process of orthogonalisation was used during the maximisation so that the successive vectors ($\boldsymbol{c}_1$, $\boldsymbol{c}_2$, …, $\boldsymbol{c}_m$) were orthogonal. A modification is needed to achieve uncorrelatedness instead of orthogonality.

SCoTLASS is implemented using S-PLUS, interfaced with a FORTRAN routine for simulated annealing (Goffe et al. 1994). The S-PLUS code requires as input the data matrix, $\mathbf{X}$, of order ($n \times p$), the value of the tuning parameter, *t*, and the number of components to be retained, *m*. It then calculates the correlation (or covariance) matrix and calls the FORTRAN routine for simulated annealing to optimize the objective function subject to the constraints. The FORTRAN routine returns a loading matrix to the S-PLUS code, which calculates a number of relevant statistics. To obtain an appropriate solution, a number of parameters of the simulated annealing method (starting points, temperature, number of cycles, maximum number of iterations and tolerance limit) also need to be defined. To achieve orthogonality for each $\boldsymbol{a}_k$, $k \geq 2$, each iteration within the simulated annealing method is followed by a Gram-Schmidt orthogonalization step.

The computation involved in finding simple components is less complex, as it requires a finite search over all possible pairwise rotations of current axes that lead to integer solutions. The only complication is how best to combine several pairwise rotations into a single step. This is discussed further in Vines (2000).

Remark 7. We conclude with perhaps the most important remark of all from a practical point of view. Most techniques used to simplify climate data, including those described in this paper, are designed to optimize some mathematical or statistical criterion and take no account of the physics of the atmosphere or ocean. A simple pattern found by one them does not necessarily have a physical interpretation. If it is known in advance what physical patterns are likely to exist, then methods can be devised that will find them, in the same way that fingerprinting techniques are tailored to detect expected climate changes (Zweirs 1999). The battery of methods that look for major, yet simple, sources of variation do just that and no more. Physical interpretation of the patterns they find needs justification through knowledge of the underlying physics.

Acknowledgements. We are grateful to Dr A. Bartzokas for providing the Mediterranean SST data. Much of this paper was written while the first author was visiting the Bureau of Meteorology Research Centre (BMRC), Melbourne. He is grateful to BMRC for the support and facilities provided and to the Leverhulme Trust for partial support of the visit through their Study Abroad Fellowship scheme. Special thanks are due to Wasyl Drosdowsky, who constructed most of the figures. Comments from the editor and 2 referees helped us to improve the paper substantially.

LITERATURE CITED

Bartzokas A, Metaxas DA, Ganas IS (1994) Spatial and temporal sea-surface temperatures in the Mediterranean. Int J Climatol 14:201–213

Cadima J, Jolliffe IT (2001) Variable selection and the interpretation of principal subspaces. J Agric Biol Environ Stat 6:62–79

Gay DM (1983) Algorithm 611. Subroutines for unconstrained optimization using a model/ trust-region approach. ACM Trans Math Software 9:503–524

Gay DM (1984) A trust region approach to linearly constrained optimization. In: Lootsma FA (ed) Numerical analysis. Proceedings, Dundee 1983. Springer-Verlag, Berlin, p 171–189

Goffe WL, Ferrier GD, Rogers J (1994) Global optimizations of statistical functions with simulated annealing. J Econometrics 60:65–99

Jolliffe IT (1987) Rotation of principal components: some comments. J Climatol 7:507–510

Jolliffe IT (1989) Rotation of ill-defined principal components. Appl Stat 38:139–147

Jolliffe IT (1995) Rotation of principal components: choice of normalization constraints. J Appl Stat 22:29–35

Jolliffe IT, Law B (1995) Normalisation constraints in rotated principal component analysis: illustrations in climatology. In: Sixth International Meeting on Statistical Climatology, p 127–130

Jolliffe IT, Uddin M (2000) The simplified component technique—an alternative to rotated principal components. J Comput Gr Stat 9:689–710

Krzanowski WJ, Marriott FHC (1995) Multivariate analysis, Part II. Arnold, London

Mestas-Nuñez AM (2000) Orthogonality properties of rotated empirical modes. Int J Climatol 20:1509–1516

Monahan AH (2000) Nonlinear principal component analysis by neural networks: theory and application to the Lorenz system. J Clim 13:821–835

Richman MB (1986) Rotation of principal components. J Climatol 6:293–335

Richman MB (1987) Rotation of principal components: a reply. J Climatol 7:511–520

Roweis ST, Saul LK (2000) Nonlinear dimensionality reduction by local linear embedding. Science 290:2323–2326

Tenenbaum JB, de Silva V, Langford JC (2000) A global geometric framework for nonlinear dimensionality reduction. Science 290:2319–2323

Tibshirani R (1996) Regression shrinkage and selection via the lasso. J R Stat Soc B 58:267–288

van den Dool HM, Saha S, Johansson Å (2000) Empirical orthogonal teleconnections. J Clim 13:1421–1435

Vines SK (2000) Simple principal components. Appl Stat 49: 441–451

Zweirs FW (1999) The detection of climate change. In: von Storch H, Flöser G (eds) Anthropogenic climate change. Springer-Verlag, Berlin, p 161–206

Editorial responsibility: Hans von Storch, Geesthacht, Germany

Submitted: March 14, 2001; Accepted: October 3, 2001
Proofs received from author(s): February 28, 2002