Journal of Multivariate Analysis 188 (2022) 104820

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

A 50-year personal journey through time with principal component analysis

Ian Jolliffe

Isle of Wight, UK

ARTICLE INFO

Article history:
Received 19 August 2021
Accepted 19 August 2021
Available online 3 September 2021

AMS 2020 subject classifications:
primary 62H25
secondary 62H99

Keywords:
Principal component analysis

ABSTRACT

Principal component analysis (PCA) is one of the most widely used multivariate techniques. A little more than 50 years ago I first encountered PCA and it has played an important role in my career and beyond, for many years since. I have been persuaded that an account of my 50-year journey through time with PCA would be a suitable topic for inclusion in the Jubilee Issue of JMVA and this is the result.

1. Introduction

Principal component analysis (PCA) is one of the most widely used techniques in multivariate analysis, so it seems appropriate for the Jubilee Issue of JMVA to include an account of its developments over the past 50 years. I have been persuaded to write such an account even though I have never published in JMVA. The theoretical developments which are the mainstream of JMVA hold comparatively little interest for me. Rather I am a retired statistician who liked to get his hands dirty analysing and interpreting data or, as John Tukey memorably put it, 'playing in other people's backyards'. However, I also liked to be aware of at least the basic theory underlying any techniques I used, so my book on PCA, Jolliffe [37,41], – probably the main reason I was invited to contribute to this Issue – includes a mixture of methodology, data, and theory. PCA was historically not a very frequently discussed topic in JMVA. As far as I can ascertain, it was more than 10 years before JMVA first included an article with the phrase "principal component analysis" in its title (Ruymgaart [68]). It is possible that because, at heart, PCA is a simple technique, much of its basic theory had already been published before JMVA was launched. There are, however, various subtleties and modifications to PCA, which are still being explored and expanded, and there is strong evidence of increasing recent interest within JMVA, with 16 articles whose title contains one of the phrases "principal component analysis", "principal components" or "PCA", published in the 5 years 2016–2020. This compares with only 15 in the first 34 years of JMVA's existence. There will, of course, be other articles that involve PCA but have none of these phrases in their title, but there is no unique simple way to search for them. I believe that focussing on the easy-to-trace PCA articles gives a reasonable representation of the relative rarity of PCA-based papers in early issues of JMVA as compared to a significant expansion of interest in recent years. From a practical point of view PCA has also never been more widely used. The various techniques that come under headings such as big data, machine learning, artificial intelligence, etc. often need a dimension-reducing first step, a role that can be played by PCA. The next Section of this paper will define PCA and give a brief description of its history prior to the first issue of JMVA. Subsequent Sections will cover the 50 years of JMVA chronologically, concentrating on developments related to PCA, both in my own research and more widely. The final Section will consist of some closing remarks.

E-mail address: iansandloch@gmail.com.

https://doi.org/10.1016/j.jmva.2021.104820

2. PCA and its pre-1971 history

The idea underlying PCA is straightforward and well-known. There are what might be termed 'population' and 'sample' versions. My interest has always been mostly in the sample version, in which $\boldsymbol{X}$: $n \times p$, is a data matrix consisting of n observations on p variables. In much of what follows, $\boldsymbol{X}$ is column centred, i.e., the ith entry in the jth column is $X_{ij} - \overline{X}_j$, with obvious notation. If p is large and there are substantial correlations between the variables, then the data can often be represented in $q \ll p$ dimensions without losing much information. PCA achieves such a reduction, replacing the original p variables by q linear functions of those variables, where the coefficients in these new linear functions are chosen to successively maximise the sum of variances of the q new variables, subject to constraints. The constraints are that the vectors of coefficients defining the q new variables in terms of the original p variables are mutually orthogonal and have unit length. These constraints also imply that the q new variables are mutually uncorrelated.

In the 'population' version of PCA we have a vector $\boldsymbol{x}$ of p random variables, and seek new variables, which are linear functions $\boldsymbol{\alpha}_1^\top \boldsymbol{x}, \ldots, \boldsymbol{\alpha}_p^\top \boldsymbol{x}$ of $\boldsymbol{x}$, such that the variances of the new variables are successively maximised, subject to constraints, where $^\top$ denotes the transpose of a matrix. Similarly to the 'sample' case these constraints are that the vectors $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_p$ are mutually orthogonal and have unit length, and the new variables, called the principal components, are then mutually uncorrelated.

It turns out that the variances of the principal components are the eigenvalues, in descending order, of the sample (population) covariance matrix for the sample (population) versions, and the coefficients defining the principal components are the corresponding eigenvectors. Frequently PCA is done using correlation matrices instead of covariance matrices. This is equivalent to finding optimal linear functions of the p variables, which have each first been normalised to have unit variance.

The early history of many statistical techniques is often unclear, but in the case of PCA, the key early references are Pearson 1901, [61], and Hotelling 1933, [30]. Hotelling has a similar definition to that given above, whilst Pearson adopts a geometric approach.

Work on PCA during the 30 years after Hotelling [30] was mainly theoretical, although Hotelling [31] discussed accelerated methods for computing PCs. The main virtue of PCA is that it can greatly reduce the dimension of a large data set whilst discarding relatively little information, but before the advent of electronic computers it was not feasible to use it except on data sets with very few variables. Theoretical results connected to PCA were discussed by Girshick [25,26], Anderson [4] and Gower [27], while Rao [64] is remarkable for its introduction of a large number of new ideas concerning uses, interpretations and extensions to PCA.

Two of the first articles to give practical applications of PCA were Craddock [19] and Jeffers [33]. Craddock's [19] example from meteorology had 11 variables and Jeffers's [33] examples from forestry and ecology had 13 and 19 variables, respectively. These were at the limit of the size of data set for which PCA could be done in minutes or hours, rather than days or weeks, on the computers available at the time. These two papers were to be important to me.

Although the number of variables in these examples was small, interpretation of the PCs, which were linear combinations of 10–20 variables could be difficult. I started my Ph.D. in 1967, and the subject suggested by my supervisor, John Scott, was to investigate methods for reducing the number of variables in a way that simplified interpretation, whilst minimising any loss of information. John Scott had worked with John Jeffers, so I used Jeffers's examples in my thesis. Later John Jeffers was the External Examiner for my Ph.D. thesis. I do not think that I was aware of Craddock [19] at that time, but a considerable proportion of the work I did later in my career on PCA, and indeed in my wider research, was for data of similar type to that described by Craddock [19]. Such data will be discussed later.

3. 1971–1980

When JMVA was launched in 1971 I was in my second year as a Lecturer in Statistics, busy preparing teaching material for undergraduate and M.Sc. courses, and writing up two papers on the contents of my Ph.D. thesis (Jolliffe [34,35]). The thesis tackled the problem of finding and removing what it called 'redundant' variables, in the context of PCA. Redundant variables were defined as variables whose removal gave PCs that were easier to interpret than those using all the variables, without sacrificing much information. A number of so-called 'rejection methods' were suggested for identifying and rejecting/removing potentially redundant variables, building in part on ideas from Beale et al. [5]. Artificial data were generated, in which the identity of the redundant variables was known, and the various rejection methods were tested on each data set. The rejection methods were also tested on a number of real data sets. The increase in computer speed in the previous five years is reflected by the inclusion of a data set with 57 variables. The performance of the various rejection methods on the artificial, Jolliffe [34], and real, Jolliffe [35], data sets was reported and recommendations made. What still amazes me is that the first of these papers is my second most-cited paper out of roughly 100 published (with the second paper in sixth place). At the time of writing, Google Scholar gives 1231 citations for the first paper, with an approximate increase of 100 per year in recent years — though how many of the authors who cite it have actually read it is an open question.

It has already been noted that JMVA had no articles with PCA in their titles in its first 10 years. I was also quiet on the PCA front during that period. The only other multivariate paper that I published was on cluster analysis, but as part of my teaching I was writing booklets on various topics in multivariate analysis to distribute to students. Notes that I made at the time show that I was considering writing a book on multivariate analysis as a whole, with these booklets as a starting point, concentrating on the descriptive, rather than theoretical, aspects of the field.

4. 1981–1990

My book on PCA (Jolliffe [37,41]) is almost certainly my best known publication, with 41903 citations, according to Google Scholar at the time of writing. As noted above, in the late 1970s I considered writing a more general book on multivariate analysis, and submitted proposals and draft chapters to potential publishers. None of them was interested — among other things I was told that there were already authors working on books in the area. I then considered something more specialised — in September 1981 I wrote 'My present feeling is that a monograph on 'Principal Components' might be a better bet at present, since the general Multivariate field seems to be increasingly overcrowded.'

So in late 1981 I sent a proposal to Springer to which they reacted positively, though they obviously wanted further detail and sample chapters. Over the next two and a half years there was an iterative process, in which I wrote a chapter, sent it to Springer, received comments/reviews, revised the existing chapters and wrote a new one, until, with 6 chapters completed, Springer offered me a contract in spring 1984. There followed a very busy 18 months, and I was continually finding extra relevant references, but the book was completed by the end of 1985 and published in early May 1986. At the time it was the only book with extensive coverage of PCA, though two followed in the next few years, namely Jackson [32] and Preisendorfer and Mobley [62]. Jackson [32] is aimed at a similar audience to that of Jolliffe [37] and I am fortunate that my book appeared first. Preisendorfer and Mobley [62] is very different. It concentrates on PCA and related techniques in the context of meteorology and oceanography, an area that became of great interest to me. Its notation differs substantially from that used in the mainstream statistical literature, but brings together a remarkable amount of relevant material and novel ideas.

I was a member of the Multivariate Study Group of the Royal Statistical Society between 1983 and 1987, but apart from writing the book, most of my research time in the early 1980s was spent on topics other than PCA, especially cluster analysis. One exception was a short note, demonstrating that in principal component regression it was not necessarily the large variance PCs that were the best predictors [36]. This 4-page note has turned out to be my fourth most cited paper.

My ambition as a teenager was a career in meteorology. However my preference of statistics over fluid mechanics was deemed by the UK Met. Office as incompatible with such an aspiration at that time, helping to steer me towards a statistical Ph.D. and career. I never lost my interest in weather and climate and gradually increased my involvement with the subjects to the extent that I could probably be best described as a statistical climatologist in the later years of my career.

As noted earlier, Craddock [19] gave an early example of PCA in meteorology. In the typical meteorological example the data matrix $\boldsymbol{X}$: $n \times p$, consists of (centred) values of a meteorological variable (sea level pressure, surface temperature, etc.) measured at n times and p spatial locations. The principal components are time series and the vectors of their coefficients or loadings form patterns in space. In the weather and climate literature these patterns are often referred to as 'empirical orthogonal functions (EOFs)' and PCA is called EOF analysis. Interpretations of the EOFs are sought in order to understand the main sources (modes) of variation in the atmosphere of the measured variables. For example, when dealing with sea level pressure in the North Atlantic, the so-called North Atlantic Oscillation (NAO) is often a dominant component and is roughly interpreted as measuring the strength of pressure gradient between The Azores and Iceland.

With different measured variables and different areas of the globe, and also with principal components (PCs) beyond the first two or three highest variance components, the EOFs may be more difficult to interpret. To improve interpretability it is possible to borrow the idea of rotation from factor analysis. Suppose that the $(n \times m)$ matrix $\boldsymbol{A}_m$ has as its columns $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_m$, the coefficients (loadings) of the first m (highest variance) PCs. Let $\boldsymbol{B}_m = \boldsymbol{A}_m\boldsymbol{T}$, where $\boldsymbol{T}$ is an $(m \times m)$ orthogonal matrix. The columns of $\boldsymbol{B}_m$ give loadings of m rotated components, which together account for the same amount of variation in the data as the m unrotated components, but with $\boldsymbol{T}$ chosen to simplify the loadings — make them all large or small with few intermediate values. There are various complications, especially what criterion is chosen to measure and optimise the simplicity of $\boldsymbol{B}_m$. In practice, that choice may be less important than the choice of m, and the normalisation constraints on $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_m$ also have a role to play, as does the question of whether the rotated PCs should be uncorrelated or their vectors of loadings orthogonal — both hold for unrotated PCs, but not after rotation. Richman [65] discussed rotation of PCs at some length, and in a lengthy rejoinder (Richman [66]) to a comment by Jolliffe [38].

Rotation of PCs, in atmospheric science and elsewhere, was usually performed on the first (highest variance) PCs, which often had widely different variances. Jolliffe [39] noted that the variances of the rotated PCs were more evenly spread than those of the unrotated PCs and hence may lose sight of the most dominant sources of the variation in the data. Jolliffe [39] argued that rotation was especially useful for groups of PCs with nearly equal variances, and this need not be restricted to the first m components. Rotation with different normalisation constraints, and implications for whether the components are uncorrelated or the loadings orthogonal, is explored further in Jolliffe [40].

In October 1979 I attended the IRIA (Data Science and Informatics) Second International Symposium in Versailles, where I presented a paper on cluster analysis. This symposium opened my eyes to the large amount of research in multivariate analysis, and particularly in principal component analysis, which was being done outside the English-speaking research community. The symposium, and two further French conferences later in the decade, inspired me to broaden my search when accumulating material for my book. I became particularly interested in the related topics of sensitivity and influence in PCA. Sensitivity has been defined in a number of ways with influence one of those. Influential observations in PCA are those that when removed from a data set lead to a large change in the PCs and/or their variances. It can be defined empirically or investigated theoretically Critchley [20], Tanaka [71], Benasseni [7]. One of my early Ph.D. students (Patricia Pack, née Calder) submitted her thesis in 1986 (Calder [18]) on influence for a number of multivariate techniques, including PCA. A case study for a correlation matrix based PCA is given by Pack et al. [60]. As well as influence, sensitivity encompasses, among other things, varying weights of observations, Tanaka and Tarumi [72], changes in PCs if only a subset of variables is used, Krzanowski [49], and perturbations to observations, Benasseni [6].

There were only five articles in JMVA in its first 20 years, with "principal component analysis", "principal components" or "PCA" in their title. The first of these, Ruymgaart [68] dealt with a robust version of PCA, but only in the bivariate case. Boente's [9] 'robustified' version of PCA is more general. These two articles include asymptotic distributional theory as do Dauxois et al. [21] and Waternaux [78]. Bru [12] proposes a stochastic differential equation approach to PCA. Of these, only Ruymgaart [68] was referenced in Jolliffe [37], though Waternaux made it into Jolliffe [41].

5. 1991–2004

I have chosen to diverge from segmenting by decade for the final 30 years of my story, because the year 2004 marks a threshold in two respects. Firstly, it was the year in which I took early retirement. Although I continued to be active in research, I (very) slowly cut back, and the majority of what I did was in statistical climatology or forecast verification, with relatively little on PCA. The second threshold concerns the degree of interest in PCA in JMVA. There were again few papers on PCA in JMVA in the 14-year period 1991–2004, specifically only ten with "principal component analysis", "principal components" or "PCA" in their title. However in the following six years there were ten more such papers, with 27 in the final ten years of JMVA's 50-year existence.

A variety of topics was covered in 1991–2004, including functional PCA (Ocãna et al. [59]), relationships between factor analysis and PCA (Schneeweiss and Mathes [69]), additive principal components (El Faouzi and Sarda [23]), principal component selection (Qian et al. [63]), common principal components (Neuenschwander and Flury [58], Gu and Fung [28]) and robust PCA (Kamiya and Eguchi [47]) with more general theoretical contributions from Boudou and Dauxois [11], Dumbgen [22] and Bilodeau and Duchesne [8].

The years 1991–2004 were probably the most productive for me in terms of PCA-related research outwith the book. This was helped immensely by the fact that three of my eleven research students in that decade produced theses that explored various aspects of PCA: Jorge Cadima [13], Noriah Al-Kandari [1] and Mudassir Uddin [76]. My Ph.D. thesis explored variable selection and interpretation in PCA based on the correlation matrix. Noriah Al-Kandari's thesis also did so, but for a wider range of variable selection criteria and artificial (as well as real) data sets — Al-Kandari and Jolliffe [3]. It also considered the case in which PCA was based on the covariance matrix, rather than the correlation (Al-Kandari and Jolliffe [2]).

Simplification and interpretation in PCA and elsewhere are key themes running through a substantial proportion of my research. Uddin [76] tackled these objectives in two different ways. In PCA, the variance of successive PCs is maximised subjective to orthogonality and normalisation constraints. This may then be followed by a second step, for instance by rotation, whose objective is to simplify the interpretation of the components. It is possible to combine the two steps into a simplified component technique (SCoT). If $\boldsymbol{c}_1^\top \boldsymbol{x}$ is the first simplified component, it is obtained by maximising $V(\boldsymbol{c}_1^\top \boldsymbol{x}) + \phi S(\boldsymbol{c}_1)$, subject to a normalisation constraint, where $V(\boldsymbol{c}_1^\top \boldsymbol{x})$ is the variance of $\boldsymbol{c}_1^\top \boldsymbol{x}$, $S(\boldsymbol{c}_1)$ is a simplification criterion such as that used in varimax rotation, and ϕ is the weight given to the simplicity criterion relative to the variance criterion. The second simplified component $\boldsymbol{c}_2^\top \boldsymbol{x}$ is obtained by maximising $V(\boldsymbol{c}_2^\top \boldsymbol{x})+\phi S(\boldsymbol{c}_2)$, subject to a normalisation and orthogonality constraints, and so on for subsequent components. Various properties, advantages, pitfalls and computation for *SCoT* are examined (Uddin [76], Jolliffe and Uddin [44]).

The other simplification technique investigated by Uddin [76] borrowed the idea of the LASSO (least absolute shrinkage and selection operator) from linear regression (Tibshirani [74]). In the PCA context, the so-called simplified component LASSO (SCoTLASS) technique maximises the usual variance criterion, but subject to a constraint that places an upper bound on the sum of absolute values of the coefficients of a simplified component. Efficient implementation of the technique was a non-trivial problem, and Nickolay Trendafilov's expertise in this area was invaluable, with the resulting paper (Jolliffe et al. [43]) becoming my third most cited with further computational improvements in [75]. My continuing (and increasing) interest in statistical climatology is reflected by Hannachi et al. [29] and Jolliffe et al. [45]. In the latter, SCoT and SCoTLASS are compared with another method for constructing simplified versions of PCs due to Vines [77].

The title of Jorge Cadima's thesis [13], 'Topics in descriptive principal component analysis', indicates that it does not confine itself to a single research aim. Rather, it is a treasure trove of algebraic and geometric results concerning PCA. It has chapters on 'Interpreting PCs', 'PCA and the geometry of matrix spaces', 'PCA and linear transformations of the data', 'Projections' and 'A scale-invariant component analysis'. I have intermittently worked with Jorge over the years since 1992 on topics related to, and leading on from, his thesis. Some of this work has been published and some has not. My abilities in algebra and geometry are extremely modest compared to Jorge's, so the 'hard' parts in our joint publications have invariably been his work.

Both of Cadima and Jolliffe [14,16] deal with interpretation of PCs and the former is also concerned with variable selection, topics dear to my heart over many years. Compared to previous work in the area Cadima and Jolliffe [16] formalised criteria for deciding which subspaces of variables best approximated the full set of variables or their first few PCs. Cadima and Jolliffe [14] discussed the inadvisability of solely using the size of loadings in a component to interpret that component, and suggested alternatives. Cadima and Jolliffe [15] suggested a new modification of PCA that can be used to separate morphological variation during the growth of organisms into size and shape components, and compared it with previous methods.

If the matrix $\boldsymbol{X}$: $n \times p$, is a column-centred data matrix, then $\boldsymbol{S} = \boldsymbol{X}^{\top}\boldsymbol{X}/n$ is the covariance matrix of the data, or the correlation matrix if the p variables have been standardised to have unit variance. The eigenvectors and eigenvalues of $\boldsymbol{S}$ give the loadings and variances, respectively, of the principal components. In some circumstances it is useful to carry out a similar analysis, but with the columns of $\boldsymbol{X}$ uncentred, leading to uncentred PCA. Uncentred PCA has been used in ecology (ter Braak [73]) and elsewhere. It projects observations onto the best fitting plane through the origin, rather than through the centroid of the data. Cadima and Jolliffe [17] explore the relationships between standard column-centred PCA and uncentred PCA, both in theoretical detail and in practice.

Other forms of centring have also been suggested and used, for example row-centring or double-centring by both rows and columns. An unusual form of centring, sometimes referred to as short-segment centring or decentring, used by Mann et al. [54], contributed to a lengthy and heated (some might use a stronger adjective) discussion around the so-called 'hockey stick' graph found by Mann et al. [54]. The graph, plotting global temperature over the past few centuries showed a slowly declining picture until the final century, when there was a steep upward trend, mimicking the shape of an ice-hockey stick, and suggesting strong evidence of climate change. Climate change sceptics criticised Mann et al. [54]'s analysis on a number of grounds, including the non-standard form of PCA used — see McIntyre and McKitrick [55] for example; also Mann [53]. Interest in PCA and EOFs for explaining patterns in atmospheric science, both spatial and temporal, continues to this day — see for example Risbey et al. [67].

During 1991–2004 I had various strands of research unrelated to PCA also in progress, but in the background the idea of a second edition of the PCA book was beginning to form. Notes that I made at the time state in November 1991 that 'a second edition is a long way off', but in February 1994 I noted that I 'continue to record new references with a second edition in mind'. It was in September 1998 that I wrote to Springer suggesting a second edition. They were interested and a detailed proposal was submitted in August 1999. I had been promised my first (and only) sabbatical year in 2000–2001, and my plan was to do the bulk of the book writing in that period.

Back in 1970 my thesis was handwritten, and then typed by a secretary. Ordinary mortals did not have the skill, or access to the 'golf balls' used for mathematical text, to type it themselves. When the first edition was written in the early 1980s, it was still common practice for most teaching and research documents to initially be handwritten and then typed (on paper) by one of the departmental secretaries. In the case of the first edition, most of it was expertly done by Mavis Swain. For the second edition things had changed. Springer sent me a LaTeX version of the first edition which I then used as the starting point to create the second edition electronically. The plan to do the bulk of the book writing in my sabbatical year was largely successful, with periods of time spent at institutions in Melbourne, Toulouse and Lisbon, although I worked ridiculously antisocial hours in the months following the sabbatical in order to meet the target publication date of 2002.

6. 2005–2020

I was offered the opportunity of early retirement in 2004 and gratefully accepted. By good fortune I had already set up a 6 month Discipline Bridging Award to work in the Meteorology Department at the University of Reading. The funding body kindly allowed this to go ahead, though on a 12 month part-time, rather than 6-month full-time, basis. At the end of the 12 months I transferred to a Visiting post at Reading and in 2007, when my main research collaborator at Reading, David Stephenson, moved to a Chair at the University of Exeter, I also transferred to a Visiting post at Exeter, a post that eventually came to an end in 2019. Following the publication of the second edition of the book I continued to collect additional references for a possible third edition, but came to realise that it was too big a project to contemplate doing on my own at my time of life. Some exploration was made in 2015 of bringing in a co-author, both through Springer and personal contact, but there was no enthusiasm and the idea was abandoned in 2016.

Following my move to Reading and onward to Exeter I was firmly embedded and active within the Statistical Climatology, and specifically Forecast Verification, community, with very little time spent on PCA. It therefore came as something of a surprise when, in late 2014, I was invited to provide a review article on PCA for a special issue of the Philosophical Transactions of the Royal Society Series A on Adaptive Data Analysis. I was honoured to be asked, but felt that, as I had not been keeping a close watch on recent research in PCA, I would like to enlist the help of a more research-active co-author. This was agreed and my ideal collaborator, Jorge Cadima, was signed up. The paper Jolliffe and Cadima [42], giving a review of, and recent developments involving, PCA was duly written, and appeared in 2016. It has proved to be astonishingly popular; it was the most-read article from the journal in 2020 and, at the time of writing, according to Altmetric, it has been downloaded a total of 220704 times since it first appeared online. It has become my most-cited paper.

The 'recent developments' part of the Phil. Trans. article focussed on four of the many modifications and adaptations of PCA that have been developed in numerous disciplines, for various data types. Those topics were Functional PCA, Simplified PCs, Robust PCA and Symbolic Data PCA. Of these, the first two have featured extensively in the recent expansion of PCA-related topics in JMVA. Functional PCA appears in at least nine JMVA articles since 2005 including, most recently, Song and Li [70]), Bongiorno and Goia [10], Li et al. [52], Lakraj and Ruymgaart [50], and Kalogridis and van Aelst [46]. The last of these has a robust version of functional PCA, but there is little else on robustness in the period 2005–2021, and nothing that I found on PCA for symbolic data.

Simplified PCs, referred to as 'sparse PCA', appear in at least seven JMVA papers in the same period, for example, Merola and Chen [56], Fang et al. [24], Kim and Wang [48]. As noted by the first of these, a large number of methods for 'sparse' PCA have been proposed, with the most popular probably being that due to Zou et al. [79]. Apart from 'sparse' PCA and functional PCA no other topics stand out in the recent expansion of interest in PCA in JMVA. Rather, a variety of topics is covered in which PCA can play a role, ranging from clustering (Nakayama et al. [57]) to geodesics (Lazar and Lin [51]).

7. Final remarks

Principal component analysis has been around for over a century, but it was only in the decade before JMVA was first published that computing capabilities had developed sufficiently for it to be used routinely on data sets of a non-trivial size. It is, at heart, a simple technique but, as I discovered in writing the two editions of my book on the subject, it has an amazing range of modifications and adaptations that can be used in variety of disciplines. I was fortunate that the initial expansion of interest in PCA coincided with the start of my academic career and I was able to focus a substantial part of my research activities during my pre-retirement era in that niche. I was also lucky to be the first to publish a substantial book on PCA, thereby making me 'visible' to other researchers, conference organisers, and so on. That and the existence of some excellent research students helped further my career.

JMVA was established at the time that I was writing up my first research papers on PCA, but for the first 34 years of its existence it published relatively few articles on the subject. This has changed in the last 16 years, with a big expansion in the number of PCA-related articles appearing in JMVA. Why is this so? It seems that PCA is currently a hot topic in many disciplines, for many types of data, as evidenced by the reaction to Jolliffe and Cadima [42]. Speculating about the reasons why, it seems likely that the expansion of 'big data' with its overlapping disciplines of data science, data analytics, data mining, artificial intelligence, machine learning etc. (and of course more traditional statistical analysis) is a major reason. Most 'big data' need some initial dimension reduction, and PCA is a classic tool for doing this. Add to this the fact that it is useful in a relatively large number of disciplines where data are created, and that different types of data may need new varieties of PCA and hence new research, then it is unsurprising that PCA is of great interest, and will continue to be so. As for myself, I keep thinking that I have finished with academic-related activities but somehow they never quite go away. I am very grateful to Dietrich von Rosen for his persistence in persuading me to again venture out of retirement to write the present paper, and for his help in enabling me to do so.

References

[1] N. Al-Kandari, Variable selection and interpretation in principal component analysis (Ph.D. thesis), University of Aberdeen, 1998, Unpublished.
[2] N. Al-Kandari, I.T. Jolliffe, Variable selection and interpretation of covariance principal components, Comm. Statist. Simulation Comput. 30 (2001) 339–354.
[3] N. Al-Kandari, I.T. Jolliffe, Variable selection and interpretation in correlation principal components, Environmetrics 16 (2005) 659–672.
[4] T.W. Anderson, Asymptotic theory for principal component analysis, Ann. Math. Stat. 34 (1963) 122–148.
[5] E.M.L. Beale, M.G. Kendall, D.W. Mann, The discarding of variables in multivariate analysis, Biometrika 54 (1967) 357–366.
[6] J. Benasseni, Sensitivity of principal component analysis to data perturbation, in: Fifth International Symposium: Data Analysis and Informatics Tome, Vol. 1, 1987, pp. 249–256.
[7] J. Benasseni, Sensitivity coefficients for the subspaces spanned by the principal components, Commun. Stat. – Theory Methods 19 (1990) 2021–2034.
[8] M. Bilodeau, P. Duchesne, Principal component analysis from the multivariate familial correlation matrix, J. Multivariate Anal. 82 (2002) 457–470.
[9] G. Boente, Asymptotic theory for robust principal components, J. Multivariate Anal. 21 (1987) 67–78.
[10] E.G. Bongiorno, A. Goia, Describing the concentration of income populations by functional principal component analysis on Lorenz curves, J. Multivariate Anal. 170 (2019) 10–24.
[11] A. Boudou, J. Dauxois, Principal component analysis for a stationary random function defined on a locally compact Abelian group, J. Multivariate Anal. 51 (1994) 1–16.
[12] M.-F. Bru, Diffusions of perturbed principal components, J. Multivariate Anal. 29 (1989) 127–136.
[13] J.F.C.L Cadima, Topics in descriptive principal component analysis (Ph.D. thesis), University of Kent at Canterbury, 1992, Unpublished.
[14] J.F.C.L Cadima, I.T. Jolliffe, Loadings and correlations in the interpretation of principal components, J. Appl. Stat. 22 (1995) 203–214.
[15] J.F.C.L Cadima, I.T. Jolliffe, Size and shape-related principal component analysis, Biometrics 52 (1996) 710–716.
[16] J.F.C.L Cadima, I.T. Jolliffe, Variable selection and the interpretation of principal subspaces, J. Agric. Biol. Environ. Stat. 6 (2001) 62–79.
[17] J.F.C.L Cadima, I.T. Jolliffe, On relationships between uncentred and column centred principal component analysis, Pak. J. Statist. 25 (2009) 473–503.
[18] P. Calder, Influence functions in multivariate analysis (Ph.D. thesis), University of Kent at Canterbury, 1986, Unpublished.
[19] J.M. Craddock, A meteorological application of principal component analysis, Statistician 15 (1965) 143–156.
[20] F. Critchley, Influence in principal components analysis, Biometrika 72 (1985) 627–636.
[21] J. Dauxois, A. Pousse, Y. Romain, Asymptotic theory for the principal component analysis of a vector random function: Some applications to statistical inference, J. Multivariate Anal. 12 (1982) 136–154.
[22] L. Dumbgen, Likelihood ratio teats for principal components, J. Multivariate Anal. 52 (1995) 245–258.
[23] N.E. El Faouzi, P. Sarda, Rates of convergence for spline estimates of additive principal components, J. Multivariate Anal. 68 (1999) 120–137.
[24] K. Fang, X. Fan, Q. Zhang, S. Ma, Integrative sparse principal component analysis, J. Multivariate Anal. 166 (2018) 1–16.
[25] M.A. Girshick, Principal components, J. Amer. Statist. Assoc. 31 (1936) 519–528.
[26] M.A. Girshick, On the sampling theory of roots of determinantal equations, Ann. Math. Stat. 10 (1939) 203–224.
[27] J.C. Gower, Some distance properties of latent root and vector methods used in multivariate analysis, Biometrika 53 (1966) 325–338.
[28] H. Gu, W.K. Fung, Influence diagnostics in common principal component analysis, J. Multivariate Anal. 79 (2001) 275–294.
[29] A. Hannachi, I.T. Jolliffe, D.B. Stephenson, N. Trendafilov, In search of simple structures in climate: Simplifying EOFs, Int. J. Climatol. 26 (2006) 7–28.

[30] H. Hotelling, Analysis of a complex of statistical variables into principal components, J. Educ. Psychol. 24 (1933) 417–441, 498–520.
[31] H. Hotelling, Simplified computation of principal components, Psychometrika 1 (1936) 27–35.
[32] J.E. Jackson, A User's Guide to Principal Components, Wiley, New York, 1991.
[33] J.N.R. Jeffers, Two case studies in the application of principal component analysis, Appl. Statist. 16 (1967) 225–236.
[34] I.T. Jolliffe, Discarding variables in a principal component analysis I: Artificial data, Appl. Statist. 21 (1972) 160–173.
[35] I.T. Jolliffe, Discarding variables in a principal component analysis II: Real data, Appl. Statist. 22 (1973) 21–31.
[36] I.T. Jolliffe, A note on the use of principal components in regression, Appl. Statist. 31 (1982) 300–303.
[37] I.T. Jolliffe, Principal Component Analysis, Springer, New York, 1986.
[38] I.T. Jolliffe, Rotation of principal components: Some comments, J. Climatol. 7 (1987) 507–510.
[39] I.T. Jolliffe, Rotation of ill-defined principal components, Appl. Statist. 38 (1989) 139–147.
[40] I.T. Jolliffe, Rotation of principal components: choice of normalization constraints, J. Appl. Stat. 22 (1995) 29–35.
[41] I.T. Jolliffe, Principal Component Analysis, second ed., Springer, New York, 2002.
[42] I.T. Jolliffe, J. Cadima, Principal component analysis: a review and recent developments, Phil. Trans. Roy. Soc. A – Math. Phys. Eng. Sci. 374 (2065) (2016) 20150202.
[43] I.T. Jolliffe, N. Trendafilov, M. Uddin, A modified principal component technique based on the LASSO, J. Comput. Graph. Statist. 12 (2003) 531–547.
[44] I.T. Jolliffe, M. Uddin, The simplified component technique - an alternative to rotated principal components, J. Comput. Graph. Statist. 9 (2000) 689–710.
[45] I.T. Jolliffe, M. Uddin, K. Vines, Simplified EOFs – three alternatives to rotation, Clim. Res. 20 (2002) 271–279.
[46] I. Kalogridis, S. van Aelst, Robust functional regression based on principal components, J. Multivariate Anal. 173 (2019) 393–415.
[47] H. Kamiya, S. Eguchi, A class of robust principal component vectors, J. Multivariate Anal. 77 (2001) 239–269.
[48] D. Kim, Y. Wang, Sparse PCA-based on high-dimensional Itô processes with measurement errors, J. Multivariate Anal. 152 (2016) 172–189.
[49] W.J. Krzanowski, Selection of variables to preserve multivariate structure, using principal components, Appl. Statist. 36 (1987) 22–33.
[50] G.P. Lakraj, F. Ruymgaart, Some asymptotic theory for Silverman's smoothed functional principal components in an abstract Hilbert space, J. Multivariate Anal. 155 (2017) 122–132.
[51] D. Lazar, L. Lin, Scale and curvature effects in principal geodesic analysis, J. Multivariate Anal. 153 (2017) 64–82.
[52] Y. Li, C. Huang, W.K. Härdle, Spatial functional principal component analysis with applications to brain image data, J. Multivariate Anal. 170 (2018) 263–274.
[53] M.E. Mann, The Hockey Stick and the Climate Wars, Columbia University Press, 2012, Which city.
[54] M.E. Mann, R.S. Bradley, M.K. Hughes, Global-scale temperature patterns and climate forcing over the past six centuries, Nature 392 (6678) (1998) 779–787, Corrigendum 430 (6995) (2004) 105.
[55] S. McIntyre, R. McKitrick, Hockey sticks, principal components and spurious significance, Geophys. Res. Lett. 32 (2005) L03710.
[56] G.M. Merola, G. Chen, Projection sparse principal component analysis: An efficient least squares method, J. Multivariate Anal. 173 (2019) 366–382.
[57] Y. Nakayama, K. Yata, M. Aoshima, Clustering by principal component analysis with Gaussian kernel in high-dimension, low-sample-size settings, J. Multivariate Anal. 185 (2021) 104779.
[58] B.E. Neuenschwander, B.D. Flury, Common principal components for dependent random vectors, J. Multivariate Anal. 75 (2000) 163–183.
[59] F.A. Ocãna, A.M. Aguilera, M.J. Valderrama, Functional principal components analysis by choice of norm, J. Multivariate Anal. 71 (1999) 262–276.
[60] P. Pack, I.T. Jolliffe, B.J.T. Morgan, Influential observations in principal component analysis: A case study, J. Appl. Stat. 15 (1988) 39–52.
[61] K. Pearson, On lines and planes of closest fit to systems of points in space, Phil. Mag. 2 (6) (1901) 559–572.
[62] R.W. Preisendorfer, C.D. Mobley, Principal Component Analysis in Meteorology and Oceanography, Elsevier, Amsterdam, 1988.
[63] G.Q. Qian, G. Gabor, R.P. Gupta, Principal component selection by the criterion of the minimum mean difference of complexity, J. Multivariate Anal. 49 (1994) 55–75.
[64] C.R. Rao, The use and interpretation of principal component analysis in applied research, Sankhyā A 26 (1964) 329–358.
[65] M.B. Richman, Rotation of principal components, J. Climatol. 6 (1986) 293–335.
[66] M.B. Richman, Rotation of principal components: A reply, J. Climatol. 7 (1987) 511–520.
[67] J.S. Risbey, D.P. Monselesan, A.S. Black, T.S. Moore, D. Richardson, D.T. Squire, C.R. Tozer, The identification of long-lived Southern Hemisphere flow events using archetypes and principal components, Mon. Weather Rev. 149 (2021) 1987–2010.
[68] F.H. Ruymgaart, A robust principal component analysis, J. Multivariate Anal. 11 (1981) 485–497.
[69] H. Schneeweiss, H. Mathes, Factor Analysis and principal components, J. Multivariate Anal. 55 (1995) 105–124.
[70] J. Song, B. Li, Nonlinear and additive principal component analysis for functional data, J. Multivariate Anal. 181 (2021) 104675.
[71] Y. Tanaka, Sensitivity analysis in principal component analysis: Influence on the subspace spanned by principal components, Commun. Stat. – Theory Methods 17 (1988) 3157–3175.
[72] Y. Tanaka, T. Tarumi, A numerical investigation of sensitivity analysis in multivariate methods, in: Fifth International Symposium: Data Analysis and Informatics Tome, Vol. 1, 1987, pp. 237–247.
[73] C.J.F. ter Braak, Principal component biplots and alpha and beta diversity, Ecology 64 (1983) 454–462.
[74] R.J. Tibshirani, Regression shrinkage and selection via the Lasso, J. Roy. Statist. Soc. B 58 (1996) 267–288.
[75] N.T. Trendafilov, I.T. Jolliffe, Projected gradient approach to the numerical solution of the SCoTLASS, Comput. Statist. Data Anal. 50 (2006) 242–253.
[76] M. Uddin, Interpretation of results from simplified principal components (Ph.D. thesis), University of Aberdeen, 1999, Unpublished.
[77] S.K. Vines, Simple principal components, Appl. Statist. 49 (2000) 441–451.
[78] C.M. Waternaux, Principal components in the nonnormal case: The test of equality of Q roots, J. Multivariate Anal. 14 (1984) 323–335.
[79] H. Zou, T. Hastie, R.J. Tibshirani, Sparse principal component analysis, J. Comput. Graph. Statist. 15 (2006) 265–286.