Stony Brook University

The Chemical Composition of Sinking Particles and their Vertical Dynamics in the Open Ocean

A Dissertation Presented

by

Jianhong Xue

to

The Graduate School

in Partial fulfillment of the

Requirements

for the Degree of

Doctor of Philosophy

in

Marine and Atmospheric Science

Stony Brook University

August 2008

Stony Brook University
The Graduate School

Jianhong Xue

We, the dissertation committee for the above candidate for the Doctor of Philosophy degree, hereby recommend acceptance of this dissertation.

Robert A. Armstrong, Dissertation Advisor
Associate Professor, School of Marine and Atmospheric Sciences

J. Kirk Cochran, Chairperson of Defense
Professor, School of Marine and Atmospheric Sciences

Cindy Lee
Distinguished Professor, School of Marine and Atmospheric Sciences

Dong-Ping Wang
Professor, School of Marine and Atmospheric Sciences

Adrian B. Burd
Assistant Professor, Department of Marine Sciences, University of Georgia

This dissertation is accepted by the Graduate School

Lawrence Martin
Dean of the Graduate School

Abstract of the Dissertation

The Chemical Composition of Sinking Particles and their Vertical Dynamics in the Open Ocean

by

Jianhong Xue

Doctor of Philosophy

in

Marine and Atmospheric Science

Stony Brook University

2008

Sinking particles are one of the main vehicles for transporting chemical species, including carbon, from surface waters to the deep ocean and sediments. This process reduces the partial pressure of carbon dioxide in the surface mixed layer, and plays an important role in the oceanic carbon cycle. This thesis focuses on the vertical dynamics of sinking particles, and their chemical compositions and degradation trajectories.

A new method, based on fitting Fourier series to time-series data from sediment traps, was developed to estimate settling velocities (SV's) of sinking particles in the open ocean. The results showed that estimation of SV's is more reliable when single-elements, rather than element ratios, are used as tracers. Modal settling velocities estimated using single-tracer fluxes with good temporal resolution, averaged separately for each tracer, are 205 ± 74 m/d. This new estimate is in essential accord with measurements made using Indented Rotating Sphere sediment traps in "setting velocity" mode at the same site.

Three-dimensional Principal Components Analysis (PCA) was applied to data on the organic composition of sinking particles collected in MedFlux. The results showed that constructing a 1-dimensional “degradation index” is oversimplified, and the trajectory from the first 2 or 3 axes is more informative.

Masses were summed with all the major chemical components, including organic matter (OM), opal, $CaCO_3$, and lithogenic minerals, for sinking particles. The summed mass is often less than the mass that is directly measured. This mass deficit is also observed for sinking particles at the Ross Sea, but not for those collected from other US JGOFS deep ocean sites. OM was found to be directly related to mass deficit, especially in shallow water where OM content in particles is high. I hypothesize that water bound to organic molecules or minerals particles, especially in samples from shallow depths, may not be totally dehydrated after oven drying.

Dedication

To my parents, my husband, and my daughter

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Acknowledgements

I express sincere thanks to my advisor, Dr. Rob Armstrong, for his unfaltering guidance and support throughout the past years. It was his encouragement and enthusiasm which aided in the completion of this dissertation. I thank him for encouraging and giving me opportunities to take courses in the departments of mathematics and biology, and to many ocean science meetings. I thank him for his continuous support during the last two years when I was a visiting student in Dr. Eileen Hofmann's lab at Old Dominion University. I especially thank him for teaching me program writing, as well as the right altitude for doing sciences.

Many thanks also go to the other members of my dissertation committee, Drs. Kirk Cochran, Cindy Lee, Dong-Ping Wang, and Adrian Burd, for their helpful comments and insight into this dissertation. I especially thank Dr. Cindy Lee for her financial support within the context of MedFlux project. I also want to thank Cindy for her smile, help and encouragements throughout these years.

I thank many faculty and staff members in MSRC. Dong-Ping Wang guided me for my teaching practicum. Sergio Sañudo-Wilhelmy tutored me as my counselor in my first year, and allowed me to use his lab for ballast mineral measurement. Tim Essington and Markus Schartau gave me much advice on my aggregation and dissaggregation project. Sultan Hameed had a lot of statistical discussion with me on my settling velocity estimation project. Bob Cerrato gave me suggestions on the use of principal component analysis. Dave Hirschberg taught me experiment on radioisotope ^{228}Th and Atomic

Absorption Spectrometry. Many thanks go to Nancy Glover, Carol Dovi, Katerina Panagiotakopoulou, Eileen Goldsmith, Maria Riegert and Christina Fink.

A very special thank to my MedFlux project colleagues, Stuart Wakeham (SKIO), Michael Peterson (UW), Lynn Abramson, Jenni Szlosek, Gillian Stewart, and Aaron Beck…, for sharing data and the discussions on this project, and for doing lab work and going through all of the good and bad time together. I thank Robert Collier (OSU) for offering EqPac titanium data. I would like to thank Dr. Eileen Hofmann and her lab members for the wonderful time I spent in the past two years at the Old Dominion University.

I thank all my friends –Qingzhi and family, Xueju and Li, Jun and Lihong, Xiaolin and Ying, Yan and Jianhua, Deli and Hui, Xiaona, Tiantian, Peng, Jingbo, Lijuan, Hua, Yuan, Haifei, Qianqian, and many others - for sharing time out of my research and for bringing happiness to my family.

Last, but not least, I express my love and gratitude to my family. My parents and my brother understand and support me all the years. My parents and parents-in-law came here to help us take care of Haiying when we needed. I thank my two-year daughter Haiying who has been giving me tremendous fun since she was born. She never complained when we put her into a car-seat and drove 8 hours back and forth between Norfolk and Stony Brook many times. I send my heartfelt thanks and love to my husband Zhanfei – thanks for your unconditional love, support, and encouragement throughout the years. I was so lucky to be together with you in my life!

This work was supported by the NSF Chemical Oceanography program and the International Atomic Energy Agency in Monaco. I thank the crew members of R/V Seward Johnson II and Tethys II.

CHAPTER ONE: Introduction and background

1. Introduction

Sinking particles are one of the major vehicles for transporting carbon and other biologically-associated elements from surface waters to the ocean interior. During transit, the flux of sinking particles is greatly attenuated due to dissolution and decomposition, releasing chemical elements and dissolved organic matter into the surrounding water. A small fraction of particles eventually reaches the sea floor and provides a food source for benthic organisms. This transport process not only removes particulate elements from the surface ocean, but also controls the vertical distribution of the transported elements (Anderson, 2003). Sinking particles overall play an important role in the biogeochemical cycles of many chemical elements in the ocean.

Among all elements that are affected by sinking particles, carbon attracts the most concern because the transport of organic carbon reduces the partial pressure of carbon dioxide (pCO_2) in the surface mixed layer, allowing the ocean to take up more atmospheric CO_2 than would otherwise be possible (Sarmiento and Gruber, 2006); this mechanism has become known as the 'biological pump' (Fig. 1.1). In the euphotic zone (defined by the depth of water that have 0.1% light level of the surface, generally less than 200 m in the open ocean), and in the presence of necessary nutrients, phytoplankton transform dissolved CO_2 into organic forms through photosynthesis:

$$6CO_2 + 6H_2O \Rightarrow C_6H_{12}O_6 + 6O_2$$

This process decreases the pCO_2 in surface water, and thus enhances CO_2 transfer from atmosphere to the surface ocean. However, most of the primary production is converted back to CO_2 and nutrients by zooplankton respiration or bacterial remineralization within the euphotic zone. Only about 10% of primary production, mostly in the form of fecal pellets and phytoplankton aggregates, is exported out of the surface layer (e.g., Eppley and Peterson, 1979; Bacon et al., 1996). On their way to the ocean floor, sinking particles are continuously subject to degradation, and less than 1% of the primary production reaches the sediment. Understanding the chemical and dynamical characteristics of sinking particles is critical to the understanding of the biogeochemical cycle of carbon in the ocean and its affect on atmospheric CO_2 and climate change. Two factors are key to determining the carbon transport efficiency of the biological pump: the settling velocity (SV) of sinking particles, and the degradation rate of POC in sinking particles. Faster sinking rates leave less time for degradation, increasing the efficiency of the biological pump; for the same reason, faster degradation rates reduce the efficiency of the pump.

The collection of sinking particles has relied mainly on the use of sediment traps. "Sinking" particles can also be collected by *in situ* pump systems, where particles collected on the Teflon or Nitex screen (those larger than 53 or 70 μm, depending on the mesh of the screen) are defined as the sinking fraction, and the particles on the glass fiber or microquartz (0.7 or 1.0 μm) are defined as the suspended fraction (Bishop et al., 1984). Since the 1980s, sediment trap data have been collected at many depths and locations around the world ocean (Klaas and Archer, 2002; Honjo et al., 2008).

This thesis presents the studies on the vertical dynamics of sinking particles, the chemical composition of sinking particles, and the degradation trajectories of sinking particles based on their organic biomarkers, using mathematical modeling and statistics. In the following, I summarize the background of these studies, and the questions to be solved in this thesis.

2. Background

2.1. The vertical dynamics of sinking particles in water column

In a simple way**,** sinking particles can be visualized as small spheres. A mathematical description of settling velocity of small spheres was first developed by G. Stokes in 1850, and is now known as Stokes' law (e.g., Mann and Lazier, 1991; Diercks and Asper, 1997; Waite et al., 1997):

$$V = \frac{2gr^2(d_1 - d_2)}{9\mu} \quad , \qquad (1.1)$$

where V = velocity of fall (cm sec^{-1})*;*

g = acceleration of gravity (cm s^{-2});

r = 'equivalent' radius of particle (cm);

d_1 = density of the particle (g cm^{-3});

d_2 = density of ocean water (g cm^{-3}); and

μ = kinematic viscosity of ocean water (dyne s cm^{-2} or g cm^{-1} s^{-1}).

In Eq. (1.1), particle radius (r) and density (d_1), are the keys for determining particle SV, assuming ocean water has constant density (d_2) and viscosity (μ). However, these two parameters for sinking particles are not easily determined because sinking particles do not exist in regular shapes, but as single particles of various shapes, and as aggregates, as flocs and marine snow, and have significant physical and biological transformations in the water column (Eisma, 1986; Fowler and Knauer, 1986; Hill, 1998).

The size and density of sinking particles are subject to continuous modification by many processes, among which aggregation and disaggregation are two major processes (Clegg and Whitfield, 1990; Clegg and Whitfield, 1991; Cochran et al., 1993; Jackson and Burd, 1998). Small particles can aggregate into large particles aided by Brownian diffusion, laminar and turbulent shear, and differential sedimentation. Zooplankton ingest phytoplankton cells or small particles, producing large fecal pellets. Zooplankton feeding can also break up particles by sloppy feeding and swimming, and microbial decomposition can also lead to particle disaggregation. Physical disaggregation occurs if the frictional force or shear difference between the water and particles is sufficient to break the particles apart (Jackson, 1990; Burd and Jackson, 2002). Both the size and density of sinking particles are dynamically variable, and determining the parameters controlling these dynamics is not simple. Many studies have been conducted on particle size spectra (e.g., Sheldon et al., 1972; McCave, 1975; Bishop et al., 1980), but the uncertainty in the size spectra is high due to the large range of variation on the spectra (Jackson and Burd, 1998). A model that reaches beyond Stokes' law is clearly needed to replace it , and to quantify the settling rate spectra or bulk settling rates of sinking particles.

Many SV measurements for specific types of sinking particles, such as for ultraplankton, diatoms, and fecal pellets, have been made; these results are summarized in Table 1.1 (modified from Table 1 in Waniek et al., 2000 and Table 1 in Hill, 1998). These particles were collected in field, but their sinking velocities were measured in a laboratory using cylinders via simple observations (e.g., Smayda, 1969; Small et al., 1979; Bienfang, 1980; Bruland and Silver, 1981). Later it was argued that the SVs estimated in the lab and those from field work do not match well because sinking particles are fragile (Alldredge and Gotschalk, 1989; Hill et al., 1998). There were also measurements for aggregates and marine snow by in situ photography in a variety of environments (Table 1.1); settling rates in these studies are in the range of 48 – 328 m/d (Hill, 1998). Using the information in Table 1.1, Waniek *et al*. (2000) concluded that 100 – 200 m/d is the typical sinking speeds of large particles. However, there are big compositional differences among particles sinking in the water column, the *in situ* measurements of any single aggregate may not represent the average velocity of sinking particles.

A "benchmark" method for obtaining settling rate has been defined and used since 1980's. This method is to compare the distribution of flux and constituents from sediment traps at different depth (Deuser et al., 1981; Honjo, 1982; Billett et al., 1983; Deuser, 1986; Fowler and Knauer, 1986; Honjo, 1996). For example, Deuser et al. (1981) found that variation of bulk constituents and size fractions of sinking particles, collected at the depth of 3200 m in Sargasso Sea, were closely tied to the annual cycle of primary production in the surface water; they concluded that sinking particles fully went through water column within 60 days, which is the trap interval time. Siegel and Deuser (1997) pointed out that sinking speeds range from 50 to 200 m/d based on these earlier studies.

However, most of these traps were opened for at least 30 days, and the resolution is limited for the estimation of sinking velocity.

Using molar ratios of particulate organic carbon (OC), inorganic carbon (IC), biogenic Si, Ca, and Al from time-series sediment taps at US JGOFS ASPS (Arabian Sea Process Study) and EqPac (Equatorial Pacific Process Study), Berelson (2002) recently estimated that the average settling rates of sinking particles were 83 – 331 m/d. These time-series traps were generally open for time intervals of 8.5, 14 or 17 days, much shorter than those from earlier traps (≥ 30 days). For each tracer ratio, Berelson (2002) estimated how much time was required for a pattern of fluxes in shallow traps to reach the deeper trap by linearly regressing the ratios measured in these two series of traps under a range of integer cup delays (0, 1, 2, …). The "best" cup shift was determined as the one that has largest average linear regression coefficient among all of the tracer ratios. Settling velocities were then calculated as (depth difference)/(("best" cup shift)*(cup rotation time)). However, in most cases (14 out of 18 in Berelson's study), the no-shift case (0 delay) showed the best correlation. Berelson then used 0.5 cup shift instead of 0 for the above SV calculation to avoid an infinite result. This approximation might have biased his estimation from real SVs.

In mathematics, a Fourier series decomposes a complex continuous function into the sum of several simple periotic functions, which are all functions of time. Fourier series have been successfully used in physical oceanography to decompose waves by their frequencies (e.g., Jacobs et al., 1993; Massel, 2001). As the fluxes of sinking particles are also approximately periodic with year, season, or month, Fourier series is used in this thesis to fit the flux of sinking particles, and to develop a new method for the

estimation of SV. In this study, the MedFlux time-series sediment traps were deployed in 2003 and 2005, at the DYFAMED site in the northwestern Mediterranean Sea, and had much shorter time intervals (4 – 6 days) for each cup than any studies in US JGOFS (Peterson et al., 2005; Lee et al., submitted). This high-resolution sediment trap data might help give a better estimation of settling rate.

In general, the settling rate for sinking particles, as one of the key parameters for studying particles dynamics in the ocean, is still not clear. In this thesis, we try to develop a new method by fitting a Fourier series with sediment trap data collected during the US JGOFS and MedFlux studies.

2.2. The study of degradation trajectories of sinking particles by principal component analysis

The distribution of particulate organic matter (POM) in the ocean is mainly a function of *in situ* production in the surface ocean (Lee and Cronin, 1984). In the euphotic zone (0-200 m), particles are mainly produced by phytoplankton through photosynthesis. Zooplankton respiration and microbial decomposition are two important processes degrading these particles. Only about 5~10% of the surface primary production is exported out of the euphotic zone. Mesopelagic zone (200-1000 m), or the twilight zone, provides an important connection between the euphotic zone and the deep ocean. In the twilight zone, the particles from the surface ocean are continuously subjected to decomposition by bacteria or grazing by zooplankton. Most of the organic matter is recycled back into inorganic carbon and nutrients. There may be aggregation

and disaggregation between suspended (smaller on size) particles and sinking particles in this layer (Lee et al., 2004). However, little is known about the importance of these two processes in controlling the chemical and physical properties of particles in the midwater column. In the deep ocean (>1000 m), the concentration of POM is very low (less than 1 μM C) and almost constant with depth (Loh and Bauer, 2000), suggesting that the decomposition rate is very low, and POM is very refractory in deep ocean. Wakeham and Lee (1993) pointed out that the chemical composition of sinking particles in deeper water is different from their original sources, the plankton biomass in the upper layer, because of selective degradation for POM. The labile compounds are degraded, and the refractory compounds become more concentrated with depth. For example, amino acids serine (SER) and glycine (GLY) were more enriched in sinking particles with depth (Siezen and Mague, 1978; Lee and Cronin, 1984). The long chain fatty acids such as C_{24} are degraded less readily than short chain fatty acids like C_{16} (Wakeham and Lee, 1993). The organic composition of sinking particles can be used to determine their degradation status.

The organic composition of POM, including amino acids, lipids, and pigments, can be used to indicate the degradation status of sinking particles, assuming that the sinking particles from different depths come from the same sources. In many cases, however, the source of organic matter is subject to seasonal variations (Lee et al., 2000). For example, diatoms are usually the dominant species during the spring bloom, while coccolithophores are the dominant species after blooms; as expected, these two algae have different organic compositions, with diatoms being enriched with glycine (GLY) and threonine (THR), and coccolithophores with aspartic acid (ASP) (Lee et al., 2000).

Therefore, the organic composition of sinking particles may reflect not only the degradation status, but also the source of a particle. How to choose an appropriate statistical method to sort out the information of degradation or source is a challenging task, particularly when dealing with a large data set of organic compounds.

Principal component analysis (PCA) is a statistical technique that can reduce a multi-dimensional data matrix to fewer dimensions, so that the re-expressed dataset represents most of the information from the original dataset. More importantly, the re-expressed dataset can be visualized much more easily and quantitatively. Using PCA on amino acid data with different diagenetic status, from fresh phytoplankton to degraded sedimentary organic matter, Dauwe et al. (1999) showed that the systematic variation of the data matrix of amino acids can be explained by the first principal components (PC), which can be interpreted as a degradation index (DI). Later, Sheridan et al. (2002) applied PCA to organic compositional data of suspended particles from the Equatorial Pacific ocean, to trace their degradation state, and found that the first PC decreased as samples became more degraded. Using compositional data of sinking particles from the Southern Ocean, Ingalls et al. (2003) cautioned that the 1st PC may also reflect the source of the POC, in addition to the degradation degree. Besides giving information on degradation, PCA also has been used to classify biomarkers according to their different sources (marine or terrigenous), to classify differences among samples, or trace the fate of OM in marine environment, etc (e.g., Yunker et al., 1995; Moncheva et al., 2001; Yunker et al., 2005).

Even though PCA is becoming increasingly used in the field of marine organic geochemistry, its principles, limitations, and further applications have been rarely

reported or explored. For example, in most of the PCA applications, only the first and/or second principal components (PCs) have been used for data interpretation. In many cases, however, the third PC explains almost as much of the variance of the data matrix as does the second PC, so that including the third PC may provide much additional valuable information.

In this thesis, we aim to further explore PCA technical details including data standardization, and comparison of 3D and 2D PCA plots for resolving trajectories of degradation.

2.3. Quantifying the mass of sinking particles by their major chemical compositions

The major components of sinking particles include organic matter (OM) and minerals. The mineral fraction can be further categorized into $CaCO_3$, biogenic opal ($SiO_2.H_2O$), and lithogenic minerals. OM concentration in sinking particles varies greatly; it can be as high as 50% in particles within the euphotic zone, and decreases to about 10% in the mesopelagic ocean. Accordingly, the mineral content increases with depth, becoming over 90% at depths over 1000 m (Armstrong et al., 2002). The types of minerals in sinking particles are site specific. Sinking particles from most open oceans (e.g., equatorial Pacific, Arabian Sea, and North Atlantic) are enriched with $CaCO_3$, as high as 50-70% in deep waters (Honjo and Manganini, 1993; Honjo et al., 1995; Honjo et al., 1999). But the concentration of opal in sinking particles from the Southern Ocean can reach 80% (Collier et al., 2000; Honjo et al., 2000). In equatorial Pacific and Southern Ocean, there are almost no lithogenic minerals due to the low aeolian input

(Honjo et al., 1995; Honjo et al., 2000). In Arabian Sea or North Atlantic, lithogenic mineral is about 10% in sinking particles (Honjo and Manganini, 1993; Honjo et al., 1999).

The flux and composition of sinking particles have been extensively studied in different oceans (e.g., Lee and Cronin, 1982; Wakeham et al., 1984; Honjo and Manganini, 1993; Honjo et al., 1995; Honjo et al., 1999; Collier et al., 2000; Honjo et al., 2000; Prahl et al., 2000). The measurements of total mass and the major chemical compositions of sinking particles are important, because they not only represent the total amount of the bulk material or a single element being transferred from surface to ocean interior, but also provide information about the relationship of organic vs. inorganic components (Armstrong et al., 2002). The total mass and compositions of sinking particles collected by time-series sediment traps can also be used for benchmark estimates of particle sinking velocity (Chapter 2: Xue and Armstrong, submitted). As directly connected to the preservation of OC in deep oceans, OC is a key parameter to quantify carbon fluxes with depth, and to understand diagenetic processes in the water column (e.g. Wakeham and Lee, 1993; Lee et al., 2000).

Aluminum (or titanium) in sinking particles is often used to estimate the flux of lithogenic minerals in the water column since the only sources of Al and Ti are from aeolian and fluvial transport after the weathering of the continental crust (Taylor and Mclennan, 1995). The Al (or Ti) content in sinking particles is low (e.g., ~0.06% in equatorial Pacific, Honjo et al., 1995), so its measurement might not be accurate enough especially when the sample size is small. Consequently, the total mass minus the mass of OM, $CaCO_3$, and opal is often used to represent lithogenic minerals (e.g., Haake et al.,

1993; Takahashi et al., 2000). However, we found that the sum of OM, $CaCO_3$, opal, and lithogenic minerals (calculated from lithogenic Al and Al/Ti), the calculated total mass, did not always match the measured total mass for sinking particles in MedFlux study, as shown later in this thesis. The difference between the measured and calculated mass is beyond the analytical errors, and mechanistic explanations are needed.

In this thesis, we will compare the calculated total mass with the measured total mass for sinking particles from MedFlux and US JGOFS studies, and discuss the possible mechanisms leading to the observed mass deficit.

3. Research questions

Overall in this thesis, I try to address the following main research questions:

1. How can we estimate the SV for sinking particles using the benchmark method?
2. How does PCA work and how can we explore different aspects of PCA to solve geochemical problems, particularly in the field of particle geochemistry? How do 3D and 2D PCA work to track degradation trajectories for sinking particles?
3. How different is the measured total mass of sinking particles from the total mass calculated by summing their major chemical compositions? How can we explain this systematical mass difference observed?

4. Thesis organization

In this thesis, I will investigate the chemical composition of sinking particles and their vertical dynamics in water column. In this Chapter I, the background of this study on sinking particles is presented, and the research questions are introduced. Chapter II investigates the settling velocities of sinking particles by introducing a new benchmark method based on fitting Fourier series to time-series sediment trap data from MedFlux and JGOFS studies. Chapter III explains the mechanisms and applications of PCA to the geochemistry of particulate matter. I provide both 3D and 2D PCA to investigate degradation trajectories of sinking particles. In chapter IV, I will show that the difference between measured and calculated mass for sinking particles is systematic, and I discuss the possible reasons for this difference. In Chapter V, I summarize the conclusions for this thesis, and suggest future research directions.

References

Alldredge, A. L. and Gotschalk, C. C., 1989. Direct observations of the mass flocculation of diatom blooms - characteristics, settling velocities and formation of diatom aggregates. *Deep-Sea Research* **36**, 159-171.

Anderson, R. F., 2003. Chemical tracers of particle transport. *Treatise on Geochemistry, Holland HD and Turekian KK. Elsevier Ltd. New York. pp 247-273*.

Armstrong, R. A., Lee, C., Hedges, J. I., Honjo, S., and Wakeham, S. G., 2002. A new, mechanistic model for organic carbon fluxes in the ocean based on the quantitative association of POC with ballast minerals. *Deep-Sea Research II* **49**, 219-236.

Bacon, M. P., Cochran, J. K., Hirschberg, D., Hammar, T. R., and Fleer, A. P., 1996. Export flux of carbon at the equator during the EqPac time-series cruises estimated from Th-234 measurements. *Deep-Sea Research II* **43**, 1133-1153.

Berelson, W. M., 2002. Particle settling rates increase with depth in the ocean. *Deep-Sea Research II* **49**, 237-251.

Bienfang, P. K., 1980. Phytoplankton sinking rates in oligotrophic waters off Hawaii, USA. *Marine Biology* **61**, 69-77.

Billett, D. S. M., Lampitt, R. S., Rice, A. L., and Mantoura, R. F. C., 1983. Seasonal sedimentation of Phytoplankton to the deep-sea benthos. *Nature* **302**, 520-522.

Bishop, J. K. B., Collier, R. W., Kettens, D. R., and Edmond, J. M., 1980. The chemistry, biology, and vertical flux of particulate matter from the upper 1500m of the Panama basin. *Deep-Sea Research* **27**, 615-640.

Bishop, J. K. B., Schupack, D., Sherrell, R. M., and Conte, M., 1984. A multiple-unit large-volume in situ filtration system for sampling oceanic particulate matter in mesoscale environments. *Advances in Chemistry Series* 209, 155-175.

Bruland, K. W. and Silver, M. W., 1981. Sinking rates of fecal pellets from gelatinous Zooplankton (Salps, Pteropods, Doliolids). *Marine Biology* **63**, 295-300.

Burd, A. B. and Jackson, G. A., 2002. Modeling steady-state particle size spectra. *Environmental Science & Technology* **36**, 323-327.

Clegg, S. L. and Whitfield, M., 1990. A generalized-model for the scavenging of trace-metals in the open ocean .1. Particle cycling. *Deep-Sea Research* **37**, 809-832.

Clegg, S. L. and Whitfield, M., 1991. A generalized-model for the scavenging of trace-metals in the open ocean .2. Thorium scavenging. *Deep-Sea Research* **38**, 91-120.

Cochran, J. K., Buesseler, K. O., Bacon, M. P., and Livingston, H. D., 1993. Thorium isotopes as indicators of particle dynamics in the upper ocean - results from the JGOFS North-Atlantic Bloom Experiment. *Deep-Sea Research I* **40**, 1569-1595.

Collier, R., Dymond, J., Honjo, S., Manganini, S., Francois, R., and Dunbar, R., 2000. The vertical flux of biogenic and lithogenic material in the Ross Sea: Moored sediment trap observations 1996-1998. *Deep-Sea Research II* **47**, 3491-3520.

Dauwe, B., Middelburg, J. J., Herman, P. M. J., and Heip, C. H. R., 1999. Linking diagenetic alteration of amino acids and bulk organic matter reactivity. *Limnology & Oceanography* **44**, 1809-1814.

de la Rocha, C. L., 2003. The biological pump. *Treatise on Geochemistry, Holland HD and Turekian KK. Elsevier Ltd. New York. pp 83-111*.

Deuser, W. G., 1986. Seasonal and interannual variations in deep-water particle fluxes in the Sargasso Sea and their relation to surface hydrography. *Deep-Sea Research* **33**, 225-246.

Deuser, W. G., Ross, E. H., and Anderson, R. F., 1981. Seasonality in the supply of sediment to the deep Sargasso Sea and implications for the rapid transfer of matter to the deep ocean. *Deep-Sea Research* **28**, 495-505.

Diercks, A. R. and Asper, V. L., 1997. In situ settling speeds of marine snow aggregates below the mixed layer: Black Sea and Gulf of Mexico. *Deep-Sea Research I* **44**, 385-398.

Eisma, D., 1986. Flocculation and de-flocculation of suspended matter in estuaries. *Netherlands Journal of Sea Research* **20**, 183-199.

Eppley, R. W. and Peterson, B. J., 1979. Particulate organic matter flux and planktonic new production in the deep ocean. *Nature* **282**, 677-680.

Fowler, S. W. and Knauer, G. A., 1986. Role of large particles in the transport of elements and organic-compounds through the oceanic water column. *Progress in Oceanography* **16**, 147-194.

Haake, B., Ittekkot, V., Rixen, T., Ramaswamy, V., Nair, R. R., and Curry, W., 1993. Seasonality and interannual variability of particle fluxes to the deep Arabian Sea. *Deep-Sea Research I* **40**, 1323-1344.

Hill, P. S., 1998. Controls on floc size in the sea. *Oceanography* **11**, 13-18.
Hill, P. S., Syvitski, J. P., Cowan, E. A., and Powell, R. D., 1998. In situ observations of floc settling velocities in Glacier Bay, Alaska. *Marine Geology* **145**, 85-94.
Honjo, S., 1982. Seasonality and interaction of biogenic and lithogenic particulate flux at the Panama basin. *Science* **218**, 883-884.
Honjo, S., 1996. Fluxes of particles to the interior of the open ocean. *Particle flux in the ocean. Ittekkot, V. John Wiley & Sons, Chichester ; New York, pp.91-154.*
Honjo, S., Dymond, J., Collier, R., and Manganini, S. J., 1995. Export production of particles to the interior of the equatorial Pacific-Ocean during the 1992 EqPac Experiment. *Deep-Sea Research II* **42**, 831-870.
Honjo, S., Dymond, J., Prell, W., and Ittekkot, V., 1999. Monsoon-controlled export fluxes to the interior of the Arabian Sea. *Deep-Sea Research II* **46**, 1859-1902.
Honjo, S., Francois, R., Manganini, S., Dymond, J., and Collier, R., 2000. Particle fluxes to the interior of the Southern Ocean in the Western Pacific sector along 170 degrees W. *Deep-Sea Research II* **47**, 3521-3548.
Honjo, S. and Manganini, S. J., 1993. Annual biogenic particle fluxes to the interior of the north-Atlantic ocean - studied at 34-degrees-N 21-degrees-W and 48-degrees-N 21-degrees-W. *Deep-Sea Research II* **40**, 587-607.
Honjo, S., Manganini, S. J., Krishfield, R. A., and Francois, R., 2008. Particulate organic carbon fluxes to the ocean interior and factors controlling the biological pump: A synthesis of global sediment trap programs since 1983. *Progress in Oceanography* **76**, 217-285.
Ingalls, A. E., Lee, C., Wakeham, S. G., and Hedges, J. I., 2003. The role of biominerals in the sinking flux and preservation of amino acids in the Southern Ocean along 170 degrees W. *Deep-Sea Research II* **50**, 713-738.
Jackson, G. A., 1990. A model of the formation of marine algal flocs by physical coagulation processes. *Deep-Sea Research part A* **37**, 1197-1211.
Jackson, G. A. and Burd, A. B., 1998. Aggregation in the marine environment. *Environmental Science & Technology* **32**, 2805-2814.
Jacobs, G. A., Emery, W. J., and Born, G. H., 1993. Rossby Waves in the Pacific-ocean extracted from geosat altimeter data. *Journal of Physical Oceanography* **23**, 1155-1175.
Klaas, C. and Archer, D. E., 2002. Association of sinking organic matter with various types of mineral ballast in the deep sea: Implications for the rain ratio. *Global Biogeochemistry Cycle* **16**, 1-16.
Lee, C. and Cronin, C., 1982. The Vertical flux of particulate organic nitrogen in the sea - decomposition of amino-acids in the Peru upwelling area and the equatorial Atlantic. *Journal of Marine Research* **40**, 227-251.
Lee, C. and Cronin, C., 1984. Particulate Amino-Acids in the Sea - Effects of primary productivity and biological decomposition. *Journal of Marine Research* **42**, 1075-1097.
Lee, C., Peterson, M. L., Wakeham, S. G., Armstrong, R. A., Cochran, J. K., Miquel, J. C., Fowler, S., Hirschberg, D., Beck, A., and Xue, J. Particulate organic matter and ballast fluxes measured using Time-Seires and Settling Velocity sediment traps in the northwestern Mediterranean Sea. *Deep-Sea Research II (*submitted)

Lee, C., Wakeham, S. G., and Arnosti, C., 2004. Particulate organic matter in the sea: The composition conundrum. *Ambio* **33**, 565-575.

Lee, C., Wakeham, S. G., and Hedges, J. I., 2000. Composition and flux of particulate amino acids and chloropigments in equatorial Pacific seawater and sediments. *Deep-Sea Research I* **47**, 1535-1568.

Loh, A. N. and Bauer, J. E., 2000. Distribution, partitioning and fluxes of dissolved and particulate organic C, N and P in the eastern North Pacific and Southern Oceans. *Deep-Sea Research I* **47**, 2287-2316.

Mann, K. H. and Lazier, J. R. N., 1991. *Dynamics of marine ecosystems : biological-physical interactions in the oceans*. Blackwell Scientific Publications, Boston.

Massel, S. R., 2001. Wavelet analysis for processing of ocean surface wave records. *Ocean Engineering* **28**, 957-987.

McCave, I. N., 1975. Vertical flux of particles in ocean. *Deep-Sea Research* **22**, 491-502.

Moncheva, S., Gotsis-Skretas, O., Pagou, K., and Krastev, A., 2001. Phytoplankton blooms in Black Sea and Mediterranean coastal ecosystems subjected to anthropogenic eutrophication: Similarities and differences. *Estuarine Coastal and Shelf Science* **53**, 281-295.

Peterson, M. L., Wakeham, S. G., Lee, C., Askea, M. A., and Miquel, J. C., 2005. Novel techniques for collection of sinking particles in the ocean and determining their settling rates. *Limnology & Oceanograpgy-Method* **3**, 520-532.

Prahl, F. G., Dymond, J., and Sparrow, M. A., 2000. Annual biomarker record for export production in the central Arabian Sea. *Deep-Sea Research II* **47**, 1581-1604.

Sarmiento, J. L. and Gruber, N., 2006. *Ocean biogeochemical dynamics*. Princeton University Press, Princeton, N.J.

Sheldon, R. W., Sutcliff.Wh, and Prakash, A., 1972. Size distribution of particles in ocean. *Limnology & Oceanography* **17**, 327-340.

Sheridan, C. C., Lee, C., Wakeham, S. G., and Bishop, J. K. B., 2002. Suspended particle organic composition and cycling in surface and midwaters of the equatorial Pacific Ocean. *Deep-Sea Research I* **49**, 1983-2008.

Siegel, D. A. and Deuser, W. G., 1997. Trajectories of sinking particles in the Sargasso Sea: modeling of statistical funnels above deep-ocean sediment traps. *Deep-Sea Research I* **44**, 1519-1541.

Siezen, R. J. and Mague, T. H., 1978. Amino-acids in suspended particulate matter from oceanic and coastal waters of Pacific. *Marine Chemistry* **6**, 215-231.

Small, L. F., Fowler, S. W., and Unlu, M. Y., 1979. Sinking rates of natural copepod fecal pellets. *Marine Biology* **51**, 233-241.

Smayda, T. J., 1969. Some measurements of sinking rate of fecal pellets. *Limnology & Oceanography* **14**, 621-625.

Takahashi, K., Fujitani, N., Yanada, M., and Maita, Y., 2000. Ling-term biogenic particle fluxes in the Bering Sea and the central subarctic Pacific Ocean, 1990-1995. *Deep-Sea Research I* **47**, 1723-1759.

Taylor, S. R. and Mclennan, S. M., 1995. The geochemical evolution of the continental-crust. *Reviews of Geophysics* **33**, 241-265.

Waite, A., Fisher, A., Thompson, P. A., and Harrison, P. J., 1997. Sinking rate versus cell volume relationships illuminate sinking rate control mechanisms in marine diatoms. *Marine Ecology-Progress Series* **157**, 97-108.

Wakeham, S. G. and Lee, C., 1993. Production, transport, and alteration of particulate organic matter in the marine water column. In: Macko, M. E. a. S. (Ed.), *Organic Geochemistry*.

Wakeham, S. G., Lee, C., Farrington, J. W., and Gagosian, R. B., 1984. Biogeochemistry of particulate organic-matter in the oceans - results from sediment trap experiments. *Deep-Sea Research* **31**, 509-528.

Waniek, J., Koeve, W., and Prien, R. D., 2000. Trajectories of sinking particles and the catchment areas above sediment traps in the northeast Atlantic. *Journal of Marine Research* **58**, 983-1006.

Xue, J. and Armstrong, R. A., submitted. An improved "benchmark" method for estimating particle settling velocities from time-series sediment trap fluxes. *Deep-Sea Research II*.

Yunker, M. B., Belicka, L. L., Harvey, H. R., and Macdonald, R. W., 2005. Tracing the inputs and fate of marine and terrigenous organic matter in Arctic Ocean sediments: A multivariate analysis of lipid biomarkers. *Deep-Sea Research II* **52**, 3478-3508.

Yunker, M. B., Macdonald, R. W., Veltkamp, D. J., and Cretney, W. J., 1995. Terrestrial and marine biomarkers in a seasonally ice-covered Arctic estuary - integration of multivariate and biomarker approaches. *Marine Chemistry* **49**, 1-50.

Particle type		Sinking speed m/day		References
Phytoplankton	Ultraplankton	<0.1-2	*	Bienfang (1980)
	Net plankton	<10-20	*	Smayda (1970)
	Diatoms	10->70	*	Passow (1990)
	Diatoms	100-150		Brillett et al. (1983)
	Diatom frustule	71-100		Honjo & Manganini (1993)
	Algal aggregates	40-150	*	Smetacek (1985)
	Marine snow	1-368	*	Alldredge & Silver (1988)
	Marine snow	112	+	Alldredge & Gotschalk (1989)
	Aggregate	48-244	+	van Leussen & Cornelisse (1993)
	Mud aggregates	133	+	ten Brinke (1994)
	Aggregate	86-259	+	Dyer et al. (1996)
	Aggregate	190	+	Hill et al. (1998)
	Aggregate	86-328	+	Sternberg et al. (1999)
Protozoans	Foraminifera	30-4800	*	Kuenen (1950)
	Radiolarians	350	*	Kuenen (1950)
	Acantharia	600	*	Antia et al. (1993)
Zooplankton	Amphipoda	875	*	Smayda (1970)
	Chaetognata	435	*	Smayda (1970)
	Heteropoda	1400	*	Vinogradov (1961)
	Pteropoda	760-2270	*	Smayda (1970)
Fecal material	Nauplii & copepiodids	5-28	*	Paffenhoffer & Knowles (1979)
	Crustacae	20-150	*	Small et al. (1979)
	Salps	450-2700	*	Bruland & Silver (1981)
	Aggregate	100	*	Fowler & Knauer (1986)
	Fecal pellets	36-376		Smayda (1969)
Sinking particles		>100		Deuser (1986)
		50-200		Siegel & Deuser (1997)
		83-331		Berelson (2002)

* references see Waniek et al. (2000);
+ references see Hill (1998).

Table 1.1 Sinking speeds (m/d) for different kinds of particles.

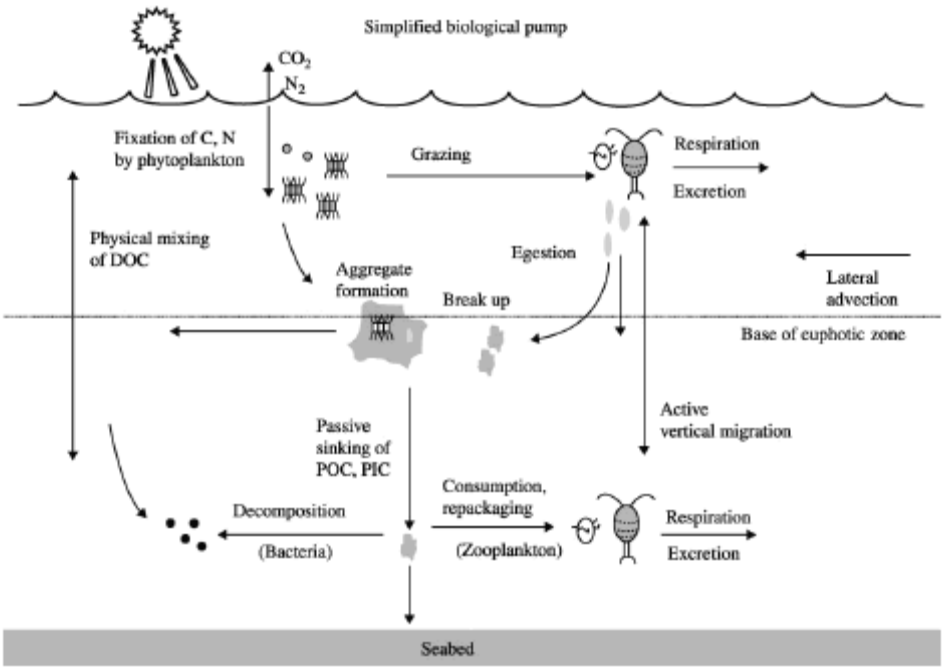


Figure 1.1 Simplified biological pump (De La Rocha, 2003).

CHAPTER TWO: An improved "benchmark" method for estimating particle settling velocities from time-series sediment trap fluxes

Abstract

A new method, based on fitting Fourier series to time-series (TS) data from sediment traps, has been developed to estimate the settling velocities (SV's) of sinking particles in the open ocean. This new method was applied to data from MedFlux, as well as from the US JGOFS NABE, EqPac, and ASPS studies. Fluxes of mass and of four chemical tracers, as well as the molar ratios of the latter, were plotted on logarithmic scales; Fourier series were then fit to these data. In each case we determined the most likely sinking velocity using a likelihood-based nonlinear fitting algorithm. Variation among estimates using single tracers was significantly less than variation using tracer ratios; we therefore concluded that estimates based on single tracers are to be preferred to estimates based on tracer ratios. Our results also showed no obvious differences among SV's estimated using different single tracers. The best estimate of settling velocity using single tracer fluxes with good temporal resolution (i.e, for sites with cup rotation times $\leq$ 8.5 days) is 205 m/d, with standard deviation 74 m/day. For MedFlux data alone (which has a resolution of 4-6 d), the estimate is 220 $\pm$ 65 m/d. This latter value is within 10% of the estimate of average sinking velocity (242 $\pm$ 31 m/d) made using MedFlux IRS traps in "sinking velocity" mode.

1. Introduction

Sinking particles are major vehicles for transporting organic carbon from surface waters to the deep ocean. This transport reduces the partial pressure of carbon dioxide (the pCO_2) in the surface mixed layer, allowing the ocean to take up more atmospheric CO_2 than would otherwise be possible (Sarmiento and Gruber, 2006). Because it determines the residence time of sinking particles in the water column, settling rate is one of the major factors that determine the depth at which organic carbon is remineralized. Therefore, accurate estimation of settling velocity is critical for understanding mechanistically the role of the ocean in determining atmospheric CO_2.

Recently it has been conjectured that the flux of organic carbon is determined by its quantitative association with ballast minerals (opal for diatoms; calcite for coccolithophorids and foraminifera; aragonite for pteropods; and dust; see Armstrong et al., 2002; François et al., 2002; Klaas and Archer, 2002). In the MedFlux program we have been seeking to identify mechanisms that determine observed patterns, so that we can understand and predict their variability in space and time. A central focus of our research has developed around the use of Indented Rotating Sphere (IRS) sediment traps in "settling velocity" (SV) mode (Peterson et al., 2005, and submitted; see also Armstrong et al., submitted; Lee et al., submitted). These traps sort particles into settling velocity classes, enabling flux estimates and chemical analyses of particles as functions of settling velocity. In an analysis of mass fluxes during the eight deployments of these settling velocity traps, Armstrong et al. (submitted) estimated *modal* settling velocities (that is, settling velocities with the highest mass fluxes per settling velocity interval) of

"fast settling" particles to be 353 m/d, with a standard deviation of 76 m/d, while the *average* velocities of the “fast sinking” fraction is much lower: 242 ± 31 m/d. These values are somewhat (but not radically) higher than the “canonical” range 80-200 m/d (Honjo, 1996; Siegel and Deuser, 1997; see also Armstrong et al., submitted).

The method of choice for assessing settling velocity using time-series trap data is to compare the time of arrival of a sediment peak at one depth to its time of arrival at a deeper depth; this approach has become known as the “benchmark” method (Deuser et al., 1981; Honjo, 1996). In the present case, a “benchmark” estimate of settling velocities is made possible by the fact that IRS traps can also be deployed in “time series” (TS) mode (Peterson et al., 2005). In TS mode, the IRS valve is rotated several times per day, and the sediment on the IRS valve is deposited in whichever cup is open at the time of rotation. We analyzed MedFlux time-series data statistically to give an estimate of sinking velocity at the DYFAMED site that was independent of the estimates from “settling velocity” (SV) mode. In addition, we reanalyzed data from the North Atlantic Bloom Experiment (NABE) (Honjo and Manganini, 1993) and from the US JGOFS Equatorial Pacific study (EqPac) and Arabian Sea Process Study (ASPS) used by Berelson (2002). This analysis was performed using a new, more powerful implementation of the benchmark method.

Using time-series trap data from the US JGOFS EqPac and ASPS studies, Berelson (2002) estimated particle settling velocities to be 83 – 331 m/day. Berelson used ratios of organic carbon (OC), biogenic Si (BioSi), Ca, and Al fluxes from different depths in the water column, arguing that ratios may have less measurement error than do measurements of individual tracers, and so would be more useful for assessing sinking

velocity. While the first statement may be true, the second – that ratios are better for assessing sinking velocity – is not, as will be shown below.

For each tracer ratio at each site, Berelson (2002) regressed the ratio measured in cup *n* in a deep trap on the same ratio measured in cup *m* in a shallower trap; his goal was to estimate how much time (in terms of cup rotation intervals) it took for a pattern of fluxes in the shallower trap to reach the depth of the deeper trap. Berelson performed separate regressions for a range of integer cup delays $n - m = \{0,1,2\ldots\}$ between upper and lower traps. The "best" cup shift, which was used in calculating the settling velocity of particles from the upper depth to the lower depth, was determined (roughly) as the shift having the largest average regression coefficient across tracer ratios. Settling velocities were then calculated as (depth difference)/(("best" cup shift)*(cup rotation time)). However, in most cases (14 of 18 in Berelson's study), the no-shift case $n - m = 0$ showed the best correlation. To avoid the possibility of infinite settling velocity (= depth difference/$(n - m)$), Berelson made the assumption that when the best shift was $n - m = 0$, he would use a cup shift of 0.5 cups. This assumption inflates the transit time between depths, and so biases Berelson's results towards lower settling velocities.

This problem notwithstanding, Berelson then compared settling velocities between "shallow" trap pairs to settling velocities between "deep" trap pairs. This comparison was made separately for EqPac and ASPS. However, within each of these data sets, he did not specifically compare deep and shallow estimates from the same site; e.g., for the EqPac comparison he used shallow settling velocities from the Equator, 5N. and 5S, while the deep velocities were from the Equator, 5N, and 12S. He concluded that settling velocity tended to increase with depth.

There are three problems with this conclusion. First, since most of the cup shifts were 0.5 cups, his sinking velocity estimates depend more on the spacing between traps than on their time offsets. Second, the composition of settling material, and hence its settling velocity, may depend as much or more on site than it does on depth; it would therefore be highly desirable to make depth comparisons only within sites. Third, in sites with traps at three depths, data from the middle trap were used to estimate time offsets between both the upper pair of traps and the lower pair of traps; any error in the middle trap thus shows up in both upper SV and the lower SV, causing them to be negatively correlated.

To avoid these problems, we devised a method whereby three Fourier series, differing only in their absolute timing, were fit simultaneously to data from shallow, middle, and deep traps at each site where data from three depths were available. (See Materials and Methods for details.) The estimated timing differences were then used in settling velocity calculations. In our method, timing differences can be any real number, and are not restricted to integer multiples of cup rotation time.

Here we apply this new method to data from MedFlux, from the US JGOFS NABE, and from the studies used by Berelson (2002). We also present results on whether measurements of single elements or ratios of elements are better indicators of settling velocity. We compare the estimated SV's to those found by Berelson (2002) and to estimates from MedFlux SV traps. Finally, we comment on the minimum timing between successive time-series cups needed to achieve satisfactory results.

2. Materials and methods

2.1. Sample collection

Particles in the MedFlux study were collected using IRS (Indented Rotating Sphere) time-series sediment traps (Peterson et al., 2005) at the DYFAMED site (43°25'N, 7°52'E) in the northwestern Mediterranean Sea (http://alpha1.msrc.sunysb.edu/MedFlux/). Separate collections were made from March to May in 2003, from May to June in 2003, and from March to April in 2005. IRS traps have 11 collection cups, plus a twelfth position that is open during trap deployment and retrieval. During deployment, each cup was open for a programmed interval of 4 to 6 days. Trap depths and deployment dates for each deployment are listed in Tables 2.1 and 2.2. Total mass and masses of ballast elements (inorganic carbon (IC), biogenic silica (BioSi), and aluminum (Al)) and organic carbon (OC) were measured after collection. Total Si, not BioSi, was measured in March to April 2005.

We also analyzed data from the US JGOFS North Atlantic Bloom Experiment (NABE), equatorial Pacific (EqPac), and Arabian Sea (ASPS) studies. These data were collected by Honjo and others (Honjo and Manganini, 1993; Honjo et al., 1995; Honjo et al., 1999) between 1989 and 1995; they are available on the US JGOFS website (http://usjgofs.whoi.edu/). Generally, time-series cups in NABE, EqPac, and ASPS were open for 14, 17, and 8.5 days, respectively. Many of these traps have continuous data for an entire year. For NABE 34N, fluxes at the shallow trap for the first half-year were much lower than those for middle and deep traps; many were close to zero, perhaps due

to some technical problem during sampling. We therefore considered only data from the second half-year in our study.

2.2. Statistical Methods

We fit Fourier series expansions to the time-series data described in the previous section. We used Fourier series only to characterize the temporal patterns occurring at successive depths, so that these patterns could be aligned and their time-offsets estimated; we did not use these fits to estimate the "power" associated with each mode, or any other potential use of Fourier series.

Usually data from a triplet of traps, (a shallow trap, a mid-depth trap, and a deep trap), deployed during the same time period, were analyzed together. All data (either for a single tracer, or for ratios between two tracers) were pretreated by transforming onto a logarithmic scale, followed by subtracting the mean of TS sample for each depth before analysis, leaving only the temporal pattern. These three sets of pretreated time series data were then fit simultaneously to the following harmonic curves:

$$\begin{aligned}
\widehat{F}_s(t_j) &= \sum_{k=1}^{m} a_k \cos(k\omega(t_j - \phi_{sk})) \\
\widehat{F}_m(t_j) &= p_{sm} \cdot [\sum_{k=1}^{m} a_k \cos(k\omega(t_j - \phi_{sk} - \frac{\Delta z_{sm}}{SV_{sm}}))] \\
\widehat{F}_d(t_j) &= p_{sd} \cdot [\sum_{k=1}^{m} a_k \cos(k\omega(t_j - \phi_{sk} - \frac{\Delta z_{sm}}{SV_{sm}} - \frac{\Delta z_{md}}{SV_{md}}))] \quad .
\end{aligned} \tag{2.1}$$

In Eq. (2.1), $\hat{F}_s(t_j)$, $\hat{F}_m(t_j)$, and $\hat{F}_d(t_j)$ are model predictions of pretreated time series data at shallow (*s*), middle (*m*), and deep (*d*) traps, respectively, at the time of closing t_j of the *j*th collection cup. In equation (2.1), a_k is the amplitude and ϕ_{sk} is the phase shift for the *k*th Fourier component for the shallow layer; ω is the frequency (d^{-1}) of the first Fourier component at $k = 1$; Δz_{sm} and Δz_{md} are depth differences between shallow and middle, and middle and deep traps, respectively; and SV_{sm} and SV_{md} are the corresponding mean settling velocities (SV's) for particles in these two depth intervals. The ratios $\Delta z_{sm} / SV_{sm}$ and $\Delta z_{md} / SV_{md}$ are therefore the estimated delays (or offsets) of the fitted pattern between shallow and mid-depth traps, and between mid-depth and deep traps, respectively. Losses of material with depth are reflected in the parameters p_{sm} and p_{sd}. At the MedFlux DYFAMED site and at EqPac-12S, only two traps were deployed; therefore only the shallow and middle equations were considered in Eq. (2.1).

Since particle fluxes have well-defined annual cycles, we used one year as the fundamental period if the length of data set was equal to or more than one year. Otherwise, we chose the fundamental period to be the closest integer divisor of 366 days. For example, the fundamental period for a 61-day time series was taken to be 61 (= 366/6) days, and the fundamental period for a 170-day series was 183 (=366/2) days.

Maximum likelihood is a well-known statistical method for finding the most likely values for a set of parameters. The logarithm of the likelihood, log(likelihood), is usually used instead of likelihood itself because likelihood is often a very small value and it can be interpreted more easily with a logarithm scale (Edwards, 1992; Hilborn and Mangel, 1997; Burnham and Anderson, 1998; Armstrong et al., 2002). In this study, data

points were fit by maximizing log(likelihood) values $\log(L)$ using the following equation:

$$\log(L) = -n\left[\log(\sigma) + \frac{1}{2}\log(2\pi)\right] - \sum_{j=1}^{n} \frac{(Y_j - \hat{Y}_j)^2}{2\sigma^2} \quad . \tag{2.2}$$

Eq. (2.2) is based on a normal error distribution for the difference between the observed and predicted values: the Y_j are observed shallow, middle, or deep layer data points, the $\hat{Y}_j$ are predicted values at these points, *n* is the number of data points in the time series, and σ is the standard deviation for the difference between Y_j *and* $\hat{Y}_j$. The value of *m*, the number of Fourier components, was usually between 5 and 7; adding additional Fourier components increased $\log(L)$ very little. Estimated settling velocities SV_{sm} and SV_{md} were constrained to be in the range 10 – 1000 m/day.

Berelson (2002) used linear-scale molar ratios for SV estimation. Here we used log-transformed ratios to remove the asymmetry between A/B and B/A: fits using A/B may give different results than fits using B/A, whereas log(A/B) = log(A) – log(B) = –log(B/A), so that a fit of log(A/B) yields a temporal pattern that is identical (except for sign) to a fit of log(B/A). For the single-tracer fits, we used Inorganic Carbon (IC) rather than Ca, which was used by Berelson (2002) to represent $CaCO_3$, assuming that all IC is in $CaCO_3$. For comparing the use of single tracers to the use of ratios, we used only the three ratios OC/Al, OC/BioSi, and OC/IC because the other ratios are not independent. We also performed fits on log-transformed data using mass fluxes and fluxes of the four elemental tracers discussed above.

3. Results

3.1. Log-transformed (single tracers) or log(tracer ratios)?

3.1.1. Insights from MedFlux March – May 2003

Settling velocities (SV's) were estimated from log-transformed data on mass fluxes and fluxes of four chemical tracers, and on three different elemental molar ratios (OC/Al, OC/BioSi, and OC/IC), for time-series sediment traps data from the MedFlux DYFAMED site in March to May 2003 (Table 2.1, first row). Estimates of settling velocity based on molar ratios had a higher mean settling velocity (393 m/d) and a much larger coefficient of variation (*cv* = 0.82) than estimates from single tracers (mean 216 m/d and *cv* = 0.22). To explore the basis for these results, we plotted (Fig. 2.1) time series of fluxes, plus model fits, for both single tracers and their molar ratios at shallow (238 m) and middle depths (771 m). The top five panels in Fig. 2.1 show results using fluxes of mass and of the four chemical tracers. These figures show that different single tracers yield slightly different estimates of settling velocity. However, these differences are minor (range: 144 – 276 m/d), so that use of one tracer in preference to another should be made on scientific, as opposed to statistical, grounds.

The bottom 3 panels in Fig. 2.1 show that fits using molar ratios can produce SV estimates that vary widely (range: 181 – 761 m/d). There is almost no visible cup delay

between upper and lower layers using OC/IC as the tracer; but the other two molar ratios (OC/Al and OC/BioSi) show obvious shifts between these two depths.

This difference may be explained as follows: use of molar ratios combines temporal patterns from two tracers; for example, log(OC/Al) = log(OC) – log(Al) depends on both OC and Al. If the two tracers in a ratio have contrasting temporal patterns, the resulting overall pattern will be enhanced; but if they have similar patterns, the signal will be reduced. This difference may help explain the large coefficient of variation in the fits using tracer ratios, and argues for choosing single elements for SV estimation. The using of molar ratios by Berelson (2002) for assessing sinking velocity may not an appropriate method.

3.1.2. Analysis of the complete data set

To investigate quantitatively how well single tracers and tracer ratios work for the complete data set, we estimated SV's at all the open-ocean sites (4 couplets and 11 triplets of traps; see Table 2.1) using the same tracers. Data from EqPac 5S were excluded because the sediment traps at this station were too close to one another in depth. Estimated SV's are shown in Table 2.1, along with the summary statistics {mean, standard deviation *sd*, and coefficient of variation *cv* = *sd* / mean}, both for SV's estimated using single tracers and SV's estimated using tracer ratios.

As shown in Table 2.1, most (21 of 26) of the *cv*'s estimated using elemental molar ratios are higher than those obtained using single tracers. A nonparametric one-sided binomial test was applied to this ratio (*cv*) shows $(cv)_{\text{single tracer}} < (cv)_{\text{ratios}}$ to be

significant at $p < 0.001$. We conclude that use of single elements gives more consistent results across tracers; single elements were therefore used for the SV estimation in the rest of this paper.

3.2. Particle SV's estimated using log(single tracers)

Table 2.2 gives time offsets ($\Delta z_{sm} / SV_{sm}$ or $\Delta z_{md} / SV_{md}$) for sinking particles from 4 couplets and 11 triplets of time-series sediment traps, and the best-fit settling velocities estimated from them, for total mass and for the four chemical tracers, respectively. Examination of these results showed that not all estimates were of equal quality. For the MedFlux DYFAMED site, estimated cup shifts were between 0.87 and 10.5 days, with an average of 4.59 days; estimated delays were therefore comparable to the DYFAMED cup rotation time (4 – 6 days). In contrast, many of the deployments at NABE and EqPac had much longer cup rotation schedules; the estimated delay was often only a fraction of a cup.

To assess the importance of this resolution issue for our estimates of settling velocity, we divided the data set into three groups: the first group included only MedFlux data (15 estimates); the second group included all estimates made using data from traps with rotation times $\leq$ 8.5 d (75 estimates, including those from MedFlux); and a third group with rotation times $\geq$ 14 d (55 estimates). (Estimates from EqPac 5S, were again excluded.)

The cumulative distribution functions (cdf) for these three groups are shown in Fig. 2.2. First, the cdf from MedFlux data is S-shaped, which is typical of a unimodal

distribution; this unimodality is confirmed in Fig. 2.3. In contrast, neither of the other two groups is ‘S’ shaped: both curves have a relatively flat region between 400 and 600 m/d, which means that few SV’s fall in this range. The histograms (Fig. 2.3) for SV’s of these two groups show clearly that both groups have two modes, centered at ~175 and 800 m/d for the high-resolution data and at ~100 and 850 m/d for the low-resolution data. The distribution for the total data set is also bimodal, with modes at ~150 and 875 m/d.

We interpret these results as implying that the MedFlux results, being of the highest resolution, are unimodal because their resolution is sufficient to allow the Fourier method to work well in almost all cases. In contrast, at many other sites the data are not sufficiently resolved to allow the method to pick out a pattern. In these cases, the fact that we imposed a maximum SV of 1000 m/d has caused results to pile up in an apparent second mode, which we deem to be an artifact.

To eliminate this artifact, we analyzed the data in Fig. 2.3 by excluding any SV > 400 m/d, which visually appears as a natural break point in Fig. 2.3. The means and standard deviations of the remaining data (also highlighted in gray color in Table 2.2 for settling velocities) are listed in Table 2.3. The average settling velocity increases with increasing resolution (from 126 m/d for low-resolution data, to 205 m/d for high-resolution data, to 220 m/d for MedFlux data), while the standard deviation and coefficient of variation decrease. We conclude that the relatively slow sinking speeds of classical studies (summarized in Honjo, 1996) are at least partly due to the low resolution of the data available at that time.

3.3. Do sinking velocities increase or decrease with depth?

In Table 2.4 we compare cup offsets and inferred sinking velocities for all data in our study. As noted in the Introduction, Berelson (2002) concluded that sinking velocity increases with depth. His conclusion is based on a comparison of SV's estimated at "shallow" and "deep" trap pairs; this comparison was made separately for ASPS and EqPac. In each of these regions, "shallow" SV's were pooled and compared to pooled "deep" SV's in the same region; no attempt was made to compare shallow and deep traps from the same site. Berelson's conclusion (his figure 5) was that settling velocity increases with depth. Comparing the error bars in his figure, it is clear that this trend was statistically significant at EqPac but not at ASPS.

Berelson's estimates of sinking velocities (Table 2.4; results taken from his Table 2.1, but excluding the result from EqPac 5S) support his conclusion: in 6 out of 8 cases, the SV between the deeper trap pair was greater than the SV between the shallower trap pair. However, Table 2.4 also contains SV's estimated using our Fourier technique, which allows an analysis based on shallow and deep trap pairs at the same site. Considering only the data from EqPac and ASPS used by Berelson (2002), our estimates indicate that the deep SV was greater than the shallow SV at only 3 of 9 sites; when the NABE sites are added, deep SV's were greater at only 5 of 11 sites. Our results stand in contradiction to those of Berelson (2002). However, since none of these results (including Berelson's) is statistically significant (exact binomial tests), we conclude that these data by themselves do not provide strong evidence for or against an increase of sinking velocity with depth.

Table 2.4 also contains a comparison of cup shifts estimated in the present paper with those reported in Berelson (2002). For EqPac and ASPS, Berelson reported 4 cup-shifts of 1.0 cup and 14 cup-shifts of 0.5 cup. Our procedure produced cup shifts between 0.10 and 1.10. For the 18 cup shifts that can be compared between studies, 12 of ours were shorter than those in Berelson (2002), and the other 6 were slightly longer than Berelson's.

4. Discussion

We estimated sinking velocities between 26 pairs of traps, in each case making 5 estimates using different tracers. In each case we determined the most likely value for the velocity at which signals are propagating down the water column. We then computed the averages of these velocities across tracers. Since estimates using different tracers are not independent, we cannot test whether results using different tracers differ significantly in a statistical sense. However, note that among the 15 couplets of high-resolution traps (3 from MedFlux and 12 from ASPS, indicated in bold in Table 2.2), settling velocities for all tracers were <400 m/d in 7 couplets, all SV's were >400 m/d in 3 couplets, and SV's were mixed in 5 couplets. More interesting, when aluminum is excluded as a tracer, all SV's are <400 m/d in 10 couplets, all are >400 m/d in 4 couplets, and estimates are mixed in only one couplet. These results suggest that even if differences among tracers do exist, they are usually unimportant (at least if Al is excluded as a tracer). They also suggest that use of Al in tracer ratios (Berelson, 2002) may lead to problems in SV

estimation, perhaps because of the “excess aluminum” problem (see Lee et al., submitted; and Chapter 4 of this dissertation).

In the present study, we estimated the average SV to be slightly more than 200 m/day. Assuming that two traps are deployed 1000 m apart, it would take about 5 days for sinking particles to settle from the upper trap to the lower trap. To have an offset comparable with the cup rotation time, sampling would need to be done with 5-day or better resolution time for each cup. For the DYFAMED site, average cup shifts for each deployment interval were between 0.38 and 1.83, with a grand average of 0.89 cup, so that the estimated shift was comparable to the resolution of the data at DYFAMED (4 to 6 days). An interesting study would be to simulate sinking of a spectrum of particles, simulate sampling by a depth series of sediment traps, then see what resolution would be needed to capture the time delays in the patterns. However, to produce reliable results such a study would require use of a sophisticated model of particle interaction. Since the data on which such a model might be produced are only now becoming available (Armstrong et al., submitted; Trull et al., in press), such a study would at this point be premature.

Finally, one of the distinct advantages of our method is that it can be used to estimate SV’s between 2 or more pairs of traps at the same time; it therefore provides a method for investigating whether sinking velocity increases with depth that is free from the statistical correlation problems outlined in the Introduction. In Table 2.5 we have listed all 11 triplet traps, with the difference for log(likelihood) between two cases: in case A we allowed two SV’s, one between the upper and middle traps and one between the middle and lower traps; in case B we fitted only a single SV for the whole water

column. To justify an extra parameter (2 SV's instead of 1), the log(likelihood) in case B must be at least 2 log(likelihood) points above that for case A (Hilborn and Mangel, 1997). This criterion was met for all tracers at only one station (NABE 34N), and only for a few tracers at other two station (ASPS 2SW and ASPS 5). Curiously, both the NABE 34N and the ASPS 5 sites were "low resolution" (Table 2.2), with ≥ 14 d cup rotation times. We again conclude that there is no strong evidence for increased sinking velocities with depth.

References

Armstrong, R.A., Lee, C., Hedges, J.I., Honjo, S., and Wakeham, S.G., 2002. A new, mechanistic model for organic carbon fluxes in the ocean based on the quantitative association of POC with ballast minerals. *Deep-Sea Research II* **49**, 219-236.

Armstrong. R.A., Peterson, M.L., Lee, C., and Wakeham, S.G. Settling velocity spectra and the ballast ratio hypothesis. *Deep-Sea Research II (submitted)*.

Berelson, W.M., 2002. Particle settling rates increase with depth in the ocean. *Deep-Sea Research II* **49**, 237-251.

Burnham, K.P., and Anderson, D.R., 1998. Model selection and inference: a practical information-theoretic approach. Springer, New York.

Deuser, W.G., Rost, E.H., Anderson, R.F., 1981. Seasonality in the supply of sediment to the deep Sargasso Sea and implications for the rapid transfer of matter to the deep ocean. *Deep-Sea Research* **28**, 495-505.

Edwards, A. W. F., 1992. *Likelihood*. Johns Hopkins University Press, Baltimore, MD, USA.

François, R., Honjo, S., Krishfield, R., and Manganini, S., 2002. Factors controlling the flux of organic carbon to the bathypelagic zone of the ocean. *Global Biogeochemical Cycles* **16**:GB1087.

Hilborn, R., and Mangel, M., 1997. The ecological detective: confronting models with data. Princeton University Press, Princeton, NJ, USA.

Honjo, S., 1996. Fluxes of particles to the interior of the open oceanp. p. 91-154 in Particle flux in the ocean, Ittekkot, V., Schäfer, P., Honjo, S., and Depetris., P.J., eds. Wiley, New York.

Honjo, S., Dymond, J., Collier, R., and Manganini, S.J., 1995. Export production of particles to the interior of the equatorial Pacific Ocean during the 1992 Eqpac experiment. *Deep-Sea Research II* **42**, 831-870.

Honjo, S., Dymond, J., Prell, W., and Ittekkot, V., 1999. Monsoon-controlled export fluxes to the interior of the Arabian Sea. *Deep-Sea Research II* **46**, 1859-1902.

Honjo, S. and Manganini, S. J., 1993. Annual biogenic particle fluxes to the interior of the North Atlantic Ocean studied at 34° N 21°W and 48°N 21°W. *Deep-Sea Research II* **40**, 587-607.

Klaas, C., and Archer, D. E., 2002. Association of sinking organic matter with various types of mineral ballast in the deep sea: Implications for the rain ratio. *Global Biogeochemical Cycles* **16**, 1-16.

Lee, C., Peterson, M.L., Wakeham, S.G., Armstrong, R.A., Cochran, J.K., Miquel, J.-C., Fowler, S., Hirschberg, D., Beck, A., Xue, J., Particulate organic matter and ballast fluxes measured using in time-series and settling velocity sediment traps in the northwestern Mediterranean Sea. *Deep-Sea Research II (submitted).*

Peterson, M.L., Wakeham, S.G., Lee, C., Askea, M., and Miquel, J. C., 2005. Novel techniques for collection of sinking particles in the ocean and determining their settling rates. *Limnology & Oceanography: Methods* **3**, 520-532.

Peterson, M.L., Fabres, J., Wakeham, S.G., Lee, C., Miquel. J.-C., Sampling the vertical particle flux in the upper water column using a large diameter free-drifting NetTrap adapted to an Indented Rotating Sphere sediment trap. *Deep-Sea Research II (submitted).*

Sarmiento, J.L., and Gruber, N., 2006. *Ocean biogeochemical dynamics*. Princeton Univ. Press, Princeton, NJ, USA.

Siegel, D. A. and Deuser, W. G., 1997. Trajectories of sinking particles in the Sargasso Sea: modeling of statistical funnels above deep-ocean sediment traps. *Deep-Sea Research I* **44**, 1519-1541.

Trull, T.W., Bray, S.G., Buesseler, K.O., Lamborg, C.H., Manganini, S., Moy, C., Valdes, J. In-situ measurement of mesopelagic particle sinking rates and the control of carbon transfer to the ocean interior during the Vertical Flux in the Global Ocean (VERTIGO) voyages in the North Pacific. *Deep-Sea Research II* (in press).

Station	deployment interval	Depth (m) upper	lower	SV (m/d) by log(tracer) mass	OC	IC	BioSi	Al	*mean*	*sd*	*cv*	SV(m/d) by log(tracer ratio) OC/Al	OC/BioSi	OC/IC	*mean*	*sd*	*cv*
MedFlux	3/06/03-5/06/03	238	771	210	229	224	276	144	*216*	*48*	*0.22*	235	181	761	*393*	*321*	*0.82*
	5/14/03-6/30/03	117	1918	172	231	293	219	179	*219*	*49*	*0.22*	115	320	507	*314*	*196*	*0.62*
	3/4/05-4/28/05	313	924	169	180	165	389	702	*321*	*233*	***0.73***	212	283	49	*181*	*120*	***0.66***
NABE-48N	4/3/89-4/2/90	1110	2109	47	116	59	40	45	*61*	*31*	*0.51*	787	19	680	*495*	*415*	*0.84*
		2109	3734	145	252	176	52	61	*137*	*83*	*0.61*	825	49	778	*550*	*435*	*0.79*
34N	10/16/89-4/2/90	1248	1894	111	175	115	102	82	*117*	*35*	*0.3*	577	828	249	*551*	*290*	*0.53*
		1894	4391	615	571	350	688	987	*642*	*230*	***0.36***	498	527	711	*579*	*116*	***0.2***
EqPac-Equator	2/02/92-1/07/93	880	2284	930	878	883	780	851	*864*	*55*	*0.06*	28	29	61	*39*	*19*	*0.48*
		2284	3618	233	54	609	118	294	*262*	*216*	*0.83*	688	804	12	*501*	*428*	*0.85*
12S	2/02/92-1/07/93	1292	3594	142	879	109	47	53	*246*	*356*	***1.45***	79	151	386	*206*	*161*	***0.78***
5N	2/02/92-1/07/93	1200	2100	131	113	154	122	121	*128*	*16*	*0.12*	43	378	744	*388*	*351*	*0.9*
		2100	3800	859	829	950	859	717	*843*	*83*	*0.1*	27	914	43	*328*	*507*	*1.55*
ASPS-2NE	11/13/94-4/26/95	903	1974	252	818	719	782	682	*650*	*229*	*0.35*	84	965	931	*660*	*499*	*0.76*
		1974	3141	245	291	263	241	305	*269*	*28*	*0.11*	858	810	530	*733*	*177*	*0.24*
3NE	12/2/94-2/8/95	858	1857	833	657	719	908	768	*777*	*98*	*0.13*	348	249	699	*432*	*236*	*0.55*
		1857	2871	780	852	542	922	780	*775*	*143*	*0.18*	719	10	786	*505*	*430*	*0.85*
4NE	11/13/94-4/26/95	821	2229	190	167	139	232	749	*296*	*256*	***0.87***	258	891	741	*630*	*331*	***0.52***
		2229	3478	142	194	148	134	53	*134*	*51*	*0.38*	251	886	223	*453*	*375*	*0.83*
2SW	5/25/95-9/09/95	924	1996	153	275	166	188	71	*170*	*73*	*0.43*	949	531	270	*583*	*343*	*0.59*
		1996	3159	98	76	82	116	736	*222*	*288*	***1.3***	214	843	89	*382*	*404*	***1.06***
3SW	6/28/95-9/17/95	888	1882	297	262	365	271	290	*297*	*41*	*0.14*	103	450	127	*226*	*194*	*0.86*
		1882	2991	186	205	161	293	186	*206*	*51*	*0.25*	760	284	645	*563*	*248*	*0.44*
4SW	6/11/95-8/23/95	807	2215	880	787	880	833	674	*811*	*86*	*0.11*	805	204	127	*379*	*371*	*0.98*
		2215	3489	584	462	754	582	224	*521*	*196*	*0.38*	45	13	564	*207*	*309*	*1.49*
5	12/6/94-4/30/95	800	2363	145	90	191	134	94	*130*	*41*	*0.32*	965	44	930	*646*	*522*	*0.81*
		2363	3915	499	848	877	800	658	*736*	*157*	*0.21*	228	125	792	*382*	*359*	*0.94*

Table 2.1 Settling velocities (SV's) estimated using either single tracers or tracer ratios. Stations with triplets of traps (shallow, middle, deep) are highlighted in gray. Data from EqPac 5S were excluded because the sediment traps at this station were deployed too closely in depth. Summary statistics (mean, standard deviation (*sd*), and coefficient of variation *cv* = *sd* / mean) are listed in each row. In 21 of 26 cases, the *cv* for single tracers was less than the *cv* for ratios; this proportion is significant at $p < 0.001$ (exact binomial test).

Station	deployment interval	Depth (m) upper	Depth (m) lower	cup rotation period (d)	offset (d) log(mass)	offset (d) log(OC)	offset (d) log(IC)	offset (d) log(BioSi)	offset (d) log(Al)	Settling velocity (m/day) log(mass)	Settling velocity (m/day) log(OC)	Settling velocity (m/day) log(IC)	Settling velocity (m/day) log(BioSi)	Settling velocity (m/day) log(Al)
MedFlux	**3/06/03-5/06/03**	**238**	**771**	**5.5**	**2.54**	**2.33**	**2.38**	**1.93**	**3.71**	**210**	**229**	**224**	**276**	**144**
	5/14/03-6/30/03	**117**	**1918**	**4.5**	**10.5**	**7.8**	**6.14**	**8.21**	**10.08**	**172**	**231**	**293**	**219**	**179**
	3/4/05-4/28/05	**313**	**924**	**5**	**3.62**	**3.4**	**3.7**	**1.57**	**0.87**	**169**	**180**	**165**	**389**	**702**
NABE-48N	4/3/89-4/2/90	1110	2109	14	21.4	8.61	16.9	24.9	22.1	47	116	59	40	45
		2109	3734	14	11.2	6.45	9.21	31.3	26.6	145	252	176	52	61
34N	10/16/89-4/2/90	1248	1894	14	5.84	3.69	5.6	6.33	7.91	111	175	115	102	82
		1894	4391	14	4.06	4.37	7.13	3.63	2.53	615	571	350	688	987
EqPac-Equator	2/02/92-1/07/93	880	2284	17	1.51	1.6	1.59	1.8	1.65	930	878	883	780	851
		2284	3618	17	5.72	24.9	2.19	11.3	4.53	233	54	609	118	294
12S	2/02/92-1/07/93	1292	3594	17	16.2	2.62	21.2	49.2	43.7	142	879	109	47	53
5N	2/02/92-1/07/93	1200	2100	17	6.85	7.96	5.86	7.38	7.45	131	113	154	122	121
		2100	3800	17	1.98	2.05	1.79	1.98	2.37	859	829	950	859	717
ASPS-2NE	**11/13/94-4/26/95**	**903**	**1974**	**8.5**	**4.25**	**1.31**	**1.49**	**1.37**	**1.57**	**252**	**818**	**719**	**782**	**682**
		1974	**3141**	**8.5**	**4.76**	**4.01**	**4.44**	**4.84**	**3.82**	**245**	**291**	**263**	**241**	**305**
3NE	**12/2/94-2/8/95**	**858**	**1857**	**8.5**	**1.2**	**1.52**	**1.39**	**1.1**	**1.3**	**833**	**657**	**719**	**908**	**768**
		1857	**2871**	**8.5**	**1.3**	**1.19**	**1.87**	**1.1**	**1.3**	**780**	**852**	**542**	**922**	**780**
4NE	**11/13/94-4/26/95**	**821**	**2229**	**8.5**	**7.41**	**8.42**	**10.1**	**6.07**	**1.88**	**190**	**167**	**139**	**232**	**749**
		2229	**3478**	**8.5**	**8.81**	**6.43**	**8.44**	**9.35**	**23.6**	**142**	**194**	**148**	**134**	**53**
2SW	**5/25/95-9/09/95**	**924**	**1996**	**8.5**	**7.02**	**3.9**	**6.46**	**5.7**	**15.2**	**153**	**275**	**166**	**188**	**71**
		1996	**3159**	**8.5**	**11.9**	**15.4**	**14.2**	**10**	**1.58**	**98**	**76**	**82**	**116**	**736**
3SW	**6/28/95-9/17/95**	**888**	**1882**	**8.5**	**3.35**	**3.8**	**2.72**	**3.67**	**3.43**	**297**	**262**	**365**	**271**	**290**
		1882	**2991**	**8.5**	**5.95**	**5.4**	**6.89**	**3.79**	**5.95**	**186**	**205**	**161**	**293**	**186**
4SW	**6/11/95-8/23/95**	**807**	**2215**	**8.5**	**1.6**	**1.79**	**1.6**	**1.69**	**2.09**	**880**	**787**	**880**	**833**	**674**
		2215	**3489**	**8.5**	**2.18**	**2.76**	**1.69**	**2.19**	**5.7**	**584**	**462**	**754**	**582**	**224**
5	12/6/94-4/30/95	800	2363	17	10.8	17.4	8.2	11.7	16.7	145	90	191	134	94
		2363	3915	17	3.11	1.83	1.77	1.94	2.36	499	848	877	800	658

Table 2.2 Time offsets (days) estimated for sinking particles from 4 couplets and 11 triplets of time-series sediment traps, and their estimated SV's, from single tracers (total mass and four chemical constituents). Stations with triplets of traps are again highlighted in gray. Data are grouped into two classes by cup rotation times: cups rotating in $\leq$ 8.5 days in boldface, and cups rotating in $\geq$ 14 days. Settling velocities less than 400 m/d are highlighted in gray.

Data category	number of trap pairs	number of SV's < 400 m/d	mean	*sd*	*cv*
All data	26	83	172	83	0.48
Low resolution (rotation interval>=14d)	11	34	126	73	0.58
High resolution (rotation interval<=8.5d)	15	49	205	74	0.36
MedFlux	3	14	220	65	0.29

Table 2.3 Numbers of trap pairs and summary statistics, arranged by data-resolution class, for sites and tracers where the SV was estimated to be less than 400 m/d (these are highlighted in gray in the last 5 columns of Table 2.2). See text for further details.

Station	deployment interval	Depth (m)		distance	cup rotation period	this paper		Berelson(2002)	
		upper	lower	(m)	(d)	SV(m/d)*	offset cup#	SV(m/d)	offset cup#
MedFlux	3/06/03-5/06/03	238	771	533	5.5	216	**0.45**	-	-
	5/14/03-6/30/03	117	1918	1801	4.5	219	**1.83**	-	-
	3/4/05-4/28/05	313	924	611	5	321	**0.38**	-	-
NABE-48N	4/3/89-4/2/90	1110	2109	999	14	61	**1.16**	-	-
		2109	3734	1625	14	137	**0.85**	-	-
34N	10/16/89-4/2/90	1248	1894	646	14	117	**0.39**	-	-
		1894	4391	2497	14	642	**0.28**	-	-
EqPac-Equator	2/02/92-1/07/93	880	2284	1404	17	864	**0.1**	83	**1**
		2284	3618	1334	17	262	**0.3**	157	**0.5**
12S	2/02/92-1/07/93	1292	3594	2302	17	246	**0.55**	271	**0.5**
5N	2/02/92-1/07/93	1200	2100	900	17	128	**0.41**	106	**0.5**
		2100	3800	1700	17	843	**0.12**	200	**0.5**
ASPS-2NE	11/13/94-4/26/95	903	1974	1071	8.5	650	**0.19**	252	**0.5**
		1974	3141	1167	8.5	269	**0.51**	137	**1**
3NE	11/13/94-4/26/95	858	1857	999	8.5	777	**0.15**	235	**0.5**
		1857	2871	1014	8.5	775	**0.15**	261	**0.5**
4NE	11/13/94-4/26/95	821	2229	1408	8.5	296	**0.56**	331	**0.5**
		2229	3478	1249	8.5	134	**1.1**	294	**0.5**
2SW	5/25/95-9/09/95	924	1996	1072	8.5	170	**0.74**	252	**0.5**
		1996	3159	1163	8.5	222	**0.62**	274	**0.5**
3SW	5/25/95-12/15/95	888	1882	994	8.5	297	**0.39**	117	**1**
		1882	2991	1109	8.5	206	**0.63**	261	**0.5**
4SW	5/25/95-12/15/95	807	2215	1408	8.5	811	**0.2**	-	-
		2215	3489	1274	8.5	521	**0.29**	300	**0.5**
5	12/6/94-12/15/95	800	2363	1563	17	130	**0.7**	92	**1**
		2363	3915	1552	17	736	**0.12**	183	**0.5**

Table 2.4 Average settling velocities and cup shifts estimated using single tracers (total mass and 4 critical elements) are compared to results from Berelson (2002). The vertical distances between upper and lower depths, and the number of days each cup was open, are also noted.

Station	deployment	Depth (m)			log(Likelihood) difference between Case A & Case B				
	interval	shallow	middle	deep	log(mass)	log(OC)	log(IC)	log(BioSi)	log(Al)
NABE-48N	**4/3/89-4/2/90**	**1110**	**2109**	**3734**	1.11	0.08	0.86	0.24	0.13
34N	**10/16/89-4/2/90**	**1248**	**1894**	**4391**	4.52	3.06	3.62	3.51	8.53
EqPac-Equator	**2/02/92-1/07/93**	**880**	**2284**	**3618**	0.11	0.66	0.05	0.02	0.07
5N	**2/02/92-1/07/93**	**1200**	**2100**	**3800**	0.61	0.63	0.22	0.98	0.29
ASPS-2NE	**11/13/94-4/26/95**	**903**	**1974**	**3141**	0.01	1.41	1.54	1.81	0.72
3NE	**11/13/94-4/26/95**	**858**	**1857**	**2871**	0.28	0.27	0	0.23	0.2
4NE	**11/13/94-4/26/95**	**821**	**2229**	**3478**	0.21	0.05	0	0.29	0.21
2SW	**5/25/95-9/09/95**	**924**	**1996**	**3159**	0.76	5.99	1.88	0.87	1.02
3SW	**5/25/95-12/15/95**	**888**	**1882**	**2991**	0.05	0.1	0.36	0.02	0.48
4SW	**5/25/95-12/15/95**	**807**	**2215**	**3489**	0.26	0.02	0.11	0.12	0.46
5	**12/6/94-12/15/95**	**800**	**2363**	**3915**	2.13	2.7	1.26	1.19	1.93

Table 2.5 Eleven triplets of time-series sediment traps analyzed by Berelson (2002) were reanalyzed to assess whether deeper particles sink faster than shallower particles. Data from EqPac 5S were again excluded. Case A: Sinking particles were allowed to have separate SV's for shallow/middle and middle/deep trap pairs. Case B: Sinking particles were constrained to have only a single SV for shallow/middle and middle/deep trap pairs. Likelihood differences must differ by at least two log(likelihood) points to justify an additional parameter (see text); only the NABE 34N site, and a few tracers at ASPS sites 2SW and 5, meet this criterion, and many sites have miniscule differences between fits made with one vs. two sinking velocities. This test provides little evidence that deeper particles sink faster.

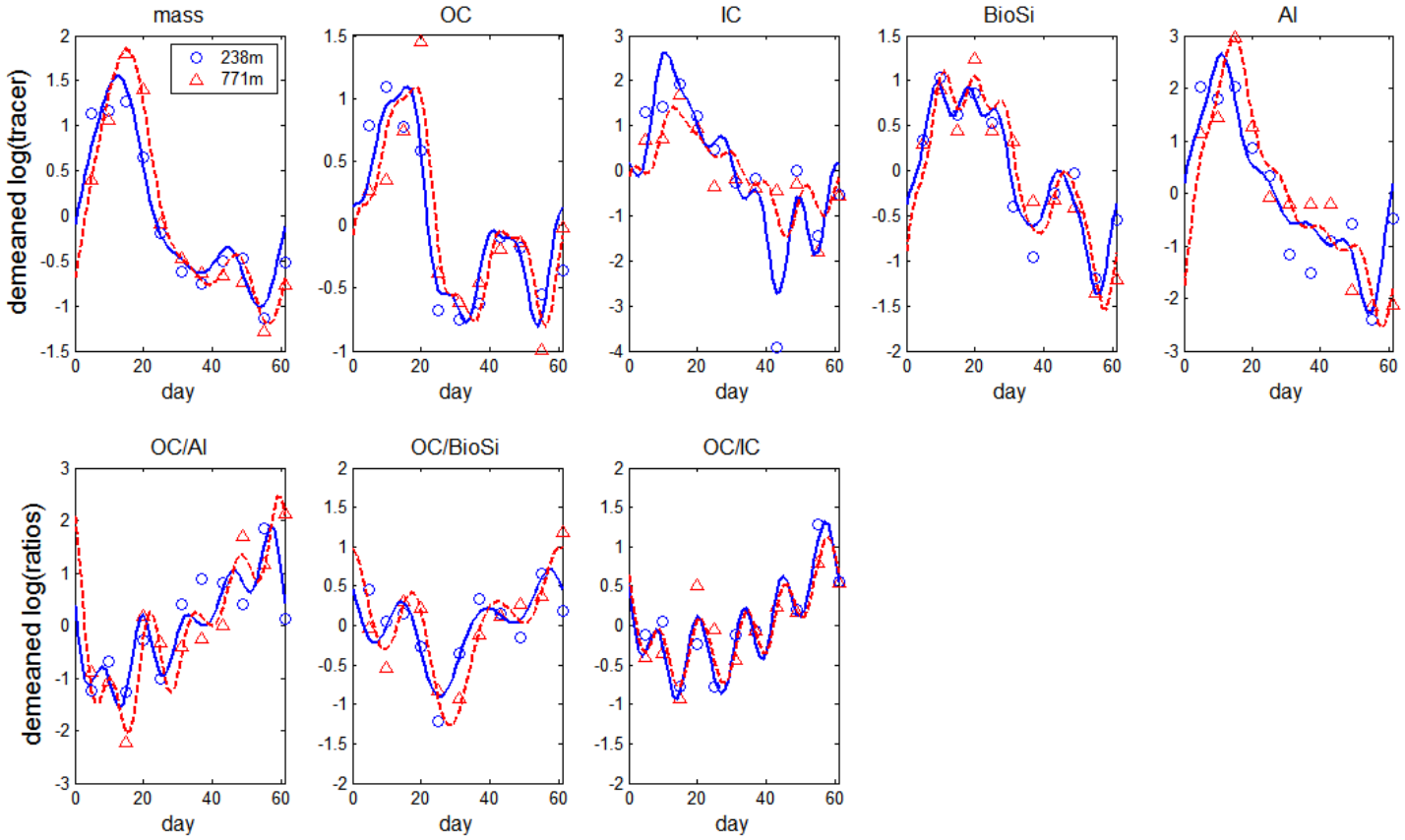


Figure 2.1 Fits of the model to demeaned log-transformed single tracers (top 5 subplots) and tracer ratios (bottom 3 subplots). Data are from MedFlux time-series sediment traps at 238m (upper depth) and 771m (lower depth) at the Mediterranean DYFAMED site, from March to May 2003; they were pretreated by transforming onto log scales, followed by subtracting the mean of each trap record before analysis.

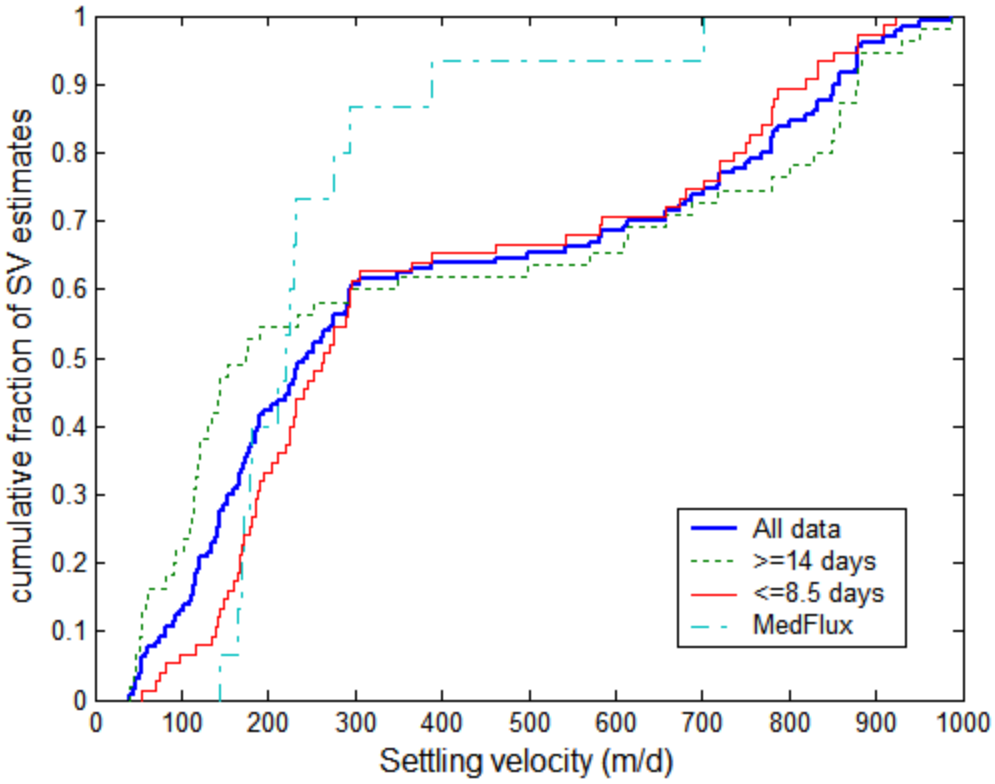


Figure 2.2 Cumulative distribution function (cdf) curves for the SV's for: (a) MedFlux data only; (b) data with cup rotation times ≤ 8.5 days (including MedFlux data); (c) data with cup rotation times ≥ 14 days; (d) all data in Table 2. The MedFlux data produce an S-shaped distribution, indicative of a unimodal distribution, whereas the other data sets produce cdf's indicative of bimodal distributions; see also Fig. 2.3. See text for further details.

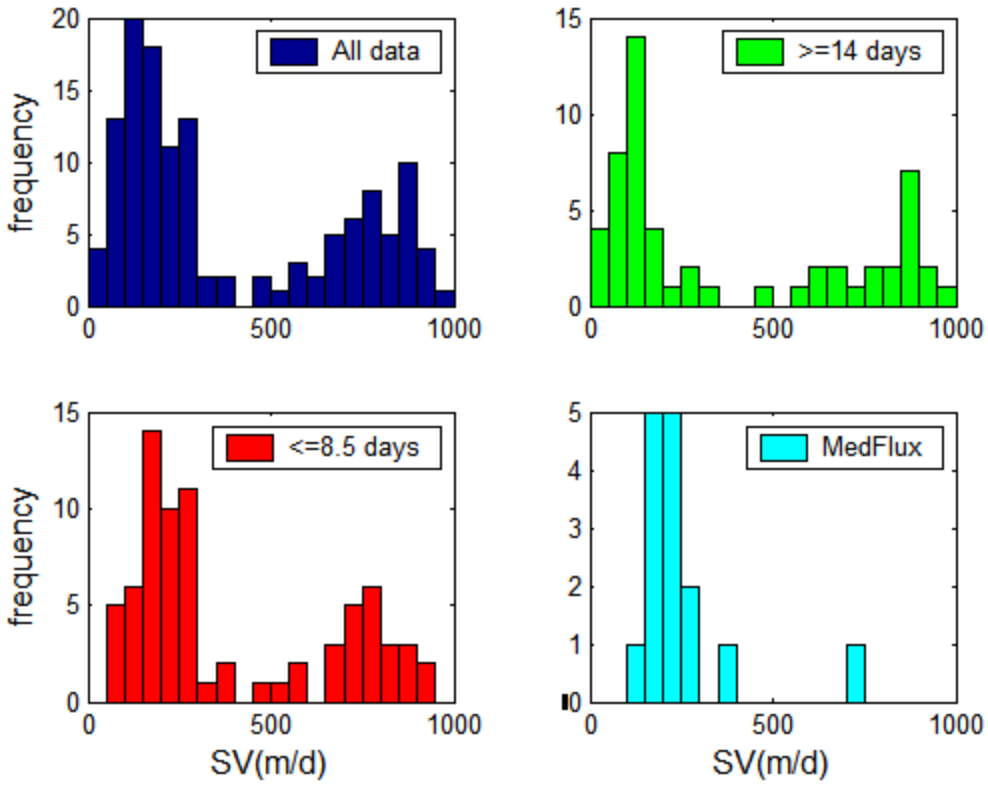


Figure 2.3 Histograms of SV estimates for the four data sets shown in Fig. 2.2.

CHAPTER THREE: Using principal components analysis (PCA) along with cluster analysis to study the organic geochemistry of sinking particles in the ocean

Abstract

Principal components analysis (PCA) is a statistical tool that can be used to re-express large multivariate datasets in a way that the first few dimensions account for the majority of the variance in the data. Although the application of PCA in marine organic geochemistry is becoming more and more popular, the capabilities of PCA have not been systematically explored. In this chapter, we illustrate the use of PCA through a series of examples that arose while investigating the geochemistry of sinking particles during the MedFlux project. Our results show that attempting to construct a 1-dimensional "degradation index" is in most cases oversimplified, and that viewing the degradation trajectories using the first 2 or 3 principle components (PC's) is more informative. Use of three - dimensional PCA is indicated when the variance explained by the third eigenvalue is comparable in magnitude to that explained by the second eigenvalue. We also discuss the use of scree plots and cluster analysis in helping decide whether a 1D, 2D, or 3D PCA most efficiently portrays the essential information in the dataset.

1. Introduction

In the field of marine organic geochemistry, more and more geochemical data are being collected due to the development of modern instrumentation. For example, different lipids compounds, including both fatty acids and neutral lipids, are analyzed and become available by the use of gas chromatography-mass spectrometry (GC-MS). Methods to detect patterns and reduce noise in huge databases are therefore becoming more important to geochemists. An appropriate statistical technique is often crucial to sift the "story" out of a puzzling dataset, or to visualize a large dataset. Principal Component Analysis (PCA) is a statistical technique that can be used to capture much of the information in a large multi-dimension data matrix in fewer dimensions, so that the re-expressed dataset can be visualized easily and quantitatively.

Previous applications of PCA in marine geochemistry fall into three categories. First, PCA has been used to classify biomarkers according to source (i.e., terrigenous or marine) and to trace the fate of organic matter (OM) in the oceanic environment (Yunker et al., 1995; Goni et al., 2000; Zimmerman and Canuel, 2001; Dittmar, 2004; Yunker et al., 2005; Hu et al., 2006). Second, the similarity and differences among samples have been investigated according to their variables, and grouping information also has been obtained through PCA (Moncheva et al., 2001; Gonzalez-Silvera et al., 2004; Schrimm et al., 2004); Third, the relationship among variables has been tracked, and summary statistics (i.e., "degradation indices") have been developed using PCA (Mariano et al., 1996; Dauwe and Middelburg, 1998; Dauwe et al., 1999; Sheridan et al., 2002; Yamashita and Tanoue, 2003; Boehme et al., 2004).

A second statistical method, cluster analysis, can also be used to explore the relationships among geochemical samples. However, cluster analysis only gives sample cluster information without giving any information of the variables within each group. Using PCA in concert with cluster analysis gives much more information than is possible using either method alone, as we will illustrate in the following examples.

PCA has been used in geochemistry for many different purposes; however, there has not been a systematic description of its principles, coupled with examples of how to apply this technique to geochemical datasets. For example, most of the applications of PCA in geochemistry have assumed that the first 1 or 2 dimensions of a PC are adequate for capturing the essential structure, so that the conclusions of these studies are based on the first one or two principal components (PCs). In particular, the 3rd PC often explains as much variance as the 2nd PC, yet its use is rarely considered.

In this chapter, we start by explaining how PCA works. Several geochemical datasets, mainly organic compositions of sinking particles from the MedFlux project, will then be discussed as examples of PCA application. We present most of our results with both 2D and 3D structure. The use of scree plots and cluster analysis in deciding whether a 3D PC structure is necessary and helpful in a given situation will be discussed.

2. Method

2.1. Intuition on PCA

PCA is a mathematical technique to re-express a dataset with a new coordinate, especially for those data of high dimension, so that the first few dimensions explain as much as possible of the variance. Here we use an artificial dataset to explain the principles in a more intuitive way. This dataset, with two variables X_1 and X_2, is plotted in Fig. 3.1a. A correlation, which is always between -1 and 1, is a single number that represents the relationship between two variables. If the correlation is positive, the relationship between two variables is positive; otherwise, the relationship is negative. The correlation matrix for these two variables is:

$$\begin{array}{ccc} & X_1 & X_2 \\ X_1 & 1.00 & 0.54 \\ X_2 & 0.54 & 1.00 \end{array},$$

which means the correlation for X_1 and X_2 is 0.54. The goal of PCA is to re-express the dataset with another two dimensions Z_1 and Z_2 in a new coordinate so that both new dimensions are linear combinations of X_1 and X_2:

$$\begin{aligned} Z_1 &= u_{11}X_1 + u_{12}X_2 \\ Z_2 &= u_{21}X_1 + u_{22}X_2 \end{aligned} \tag{3.1}$$

$$\text{Or} \qquad \begin{bmatrix} Z_1 & Z_2 \end{bmatrix} = \begin{bmatrix} X_1 & X_2 \end{bmatrix} \begin{bmatrix} u_{11} & u_{21} \\ u_{12} & u_{22} \end{bmatrix} . \tag{3.2}$$

With the proper choice of u_{11}, u_{12}, u_{21}, and u_{22}, the axis Z_1 (the first principal component) will express maximum variance of data, the second dimension Z_2 (the second principal component) exhibits the second maximum variance, and these two dimensions will be orthogonal (at right angles; "orthogonal" is the n-dimensional analog of "perpendicular") to each other. $\begin{bmatrix} u_{11} & u_{21} \\ u_{12} & u_{22} \end{bmatrix}$ is a matrix of eigenvectors (or characteristic vector) with

orthonomal character (orthogonal for two columns, and each column has length 1; see Appendix A for more explanations).

The variances of the original data points in the X_1 and X_2 directions are 12.6 and 5.4, respectively (Fig. 3.1a). The linear combinations of X_1 and X_2 are simply projections of all data points onto new axes Z_1 and Z_2. As shown in Fig. 3.1b, after transformation the largest variance is 15.8, increasing the variance among dataset from 70% on X_1 axis to 88% of the total variance on axis Z_1 (the first PC). However, PC1 does not account for all of the variance among dataset; it only gives information along Z_1. The left variance (2.2) is on the second PC, which contains all of the information not included in PC1. If there are more than two variables on original dataset, PC2 will have the largest variance among the left PC's. There is no correlation between Z_1 and Z_2, as shown by the correlation matrix:

	Z_1	Z_2
Z_1	1.00	0.00
Z_2	0.00	1.00

This is just a simple illustration of the principles of PCA. In reality, there are usually more than two variables in a given dataset. In these more general cases, PCA is used to find the first few dimensions that represent the maximum variances for data points among a multidimensional space. More mathematical details on the mechanisms of PCA can be found in Appendix A.

2.2. Questions regarding the performing of PCA

In order to perform a PCA analysis, several key factors must be considered, including how to standardize the original geochemical data and to decide how many principal components should be considered. For example, there are many ways of standardizing the data, which may lead to different conclusions. Therefore, choosing an appropriate standardization is a key to apply the PCA and make appropriate interpretation. Here we consider a dataset from MedFlux to illustrate the performing of PCA.

Particles in this MedFlux study were collected using IRS time-series sediment traps at 117m at the DYFAMED site in the northwestern Mediterranean Sea from March to May 2003 (http://www.msrc.sunysb.edu/MedFlux/) (Lee et al., submitted). These traps have 11 collection cups, which were open sequentially for preset time intervals during trap deployment. X is the dataset for 11 samples with measurements of 17 amino acids (mole percentage in composition, data listed in Table 3.1) in each cup.

2.2.1. How should the original data be standardized?

In the raw dataset X, the values of each variable are often differ in magnitude, or even in units. For example, the mole%'s and variances of the amino acids ASP, GLU, SER, GLY, and ALA are about one order of magnitude larger than those of BALA, GABA, and MET. Without standardization, the dominant amino acids like ASP would account for a much larger percentage of the variance in the linear combination, the re-expressed dataset Y, than the minor amino acids. In many cases, the minor components are often the key, more unambiguous biomarkers that for indicating geochemical processes For example, the amino acids GABA and BALA are clear indicators of

bacterial degradation, although they account less than 1% of the amino acids in sinking particles (Dauwe et al., 1999; Sheridan et al., 2002). If the scientific question is best attacked by conferring the same weight to all of the measurements (as in often the case in geochemical studies), X should be standardized before performing the PCA. The most common and simplest way to standardize the data, assuming all measurements are equally important, is, for each variable, to subtract its mean and to divide by its standard deviation:

$$T_j = \frac{X_j - \overline{X}_j}{\sigma_j} \quad , \tag{3.3}$$

where X_j is the raw data of *j*th variable for all samples, $\overline{X}_j$ is the mean of *j*th variable over all samples, σ_j is the standard deviation of *j*th variable for all samples, and T_j is the transformed data of *j*th variable for all samples. This transformation converts each variable in the dataset into a new variable with zero mean and unit standard deviation. As a result, all variables have the same weight. It should be noted that the standardization may amplify the noise brought by some minor variables whose values have relatively larger analytical errors.

There are other types of standardization techniques such as mean normalization, maximum normalization, range normalization, autoscaling, minimum-maximum transformation, and half-range and central value transformation in chemistry (Moreda-Pineiro et al., 2001). For example, the minimum-maximum transformation maps the minimum value of each variable into 0 and the maximum value into 1 so that all sample data range between 0 and 1. No matter which standardization technique is chosen, the

key is to re-assign a score for each variable in each sample so that the standardized dataset is better for solving a specific question.

It is important to choose the most appropriate way of standardizing the data pre-treatment method based on the specific dataset and the specific question to be solved. For example, in the case that will be discussed later, BALA and GABA are important bacterial degradation indicators, even though their mole percentages are much smaller than those of other major amino acids. Standardizing data using Eq. 3.3 can increase the relative weight of these two amino acids among all of the variables, and is very helpful for investigating degradation pathways in sinking particles.

2.2.2. How many principle components (PCs) should be retained in a PCA?

The total number of PCs extracted is the same as the number of variables in a PCA analysis. In a successful PCA the first few PCs explain most of the variance in the data. As such, the original dataset is simplified and reduced to fewer dimensions.

There are several ways to determine how many PC's to retain (Lattin et al., 2003). One simple method is to plot the distribution of eigenvalues. In this method (Fig. 3.2), eigenvalues are ordered from largest to smallest. The resulting figure is known as a scree plot, and the eigenvalues are the variances explained by the corresponding PC's. In Fig. 3.2, which shows the scree plot for the amino acid database (Table 3.1), there is a significant decline in explained variance between the third and the fourth eigenvalues. This region is therefore, the 'elbow' for this curve; only the PCs before the elbow are included (Cattell, 1966; Lattin et al., 2003).

For the amino acid dataset, the first three PC's together explain 87% of the variance (46%, 23%, and 18%, respectively). Retention of the first 3 PC's is suggested by the location of the knee in Fig. 3.2; the resulting 3D plot is shown in Fig. 3.3a, including Figs. 3a-1 & 3a-2 (This plot and all the left 3D plots in this chapter are all viewed from two different angles.). For comparison, a plot retaining only the first 2 PC's is shown in Fig. 3.3b. Clearly, these 11 cups are well-separated in the 3D plots. Sample cups TS 09-12 in the 2D structure, however, are not separated well from sample cups TS 06-09.

The grouping character for the 11-cup samples by PCA could also be examined by cluster analysis (Fig. 3.3c), another statistical method that can be used to visualize the relationships among cups. The advantage of cluster analysis is that its groupings consider all of the dataset variance compared to the 60-90% represented by the first few PCs in PCA analysis. With cluster analysis, cups close to each other are more highly correlated than those that are not. Cluster analysis shows that these 11 cups are separated into three groups: TS02-05, TS06-09, and TS10-12, consistent with the 3D PCA result. In contrast, the 2D plot the group TS10-12 mistakenly appears to lie *inside* group TS06-09. Thus the addition of the third PC in this case not only more accurate in conveying grouping information; it also is needed to convey the basic structure of the dataset.

2.2.3. How many variables should be retained for PCA?

Redundant variables have been discussed to discard unnecessary variables without the loss of extra variation, especially when the environmental dataset contains

large number of variables (King and Jackson, 1999). In geochemistry studies, some variables do not show much distinction among data. The adding and deleting of them from dataset does not affect the variations among samples. These types of variables could be disregard before the performing of PCA to simplify result. A useful example by deleting some unnecessary variables will be given in case 3 in the application part.

The eigenvector is the loadings of each variable on the new axes. For example, the first column of eigenvector represents the loadings of each variable on the first PC. These loadings also represent the correlations of each variable with corresponding PC's (Lattin et al., 2003). If the loading of a variable is very low, it means that the variable has low correlation relationship with the corresponding PC. For the amino acid dataset, the first three columns of eigenvector are listed in Table 3.2 with a 3D plot shown their loadings on each PC in Fig. 3.3a. The first PC (PC1), for example, exhibits good correlations with amino acids ALA, TYR, GABA, HIS, and VAL, which are all have high positive loadings above 0.28 on the x axis. The first PC, also clearly negative correlated with SER, which has negative loadings -0.33 on PC1. However, BALA does not show good correlation with the first PC as its loading is very low (0.04) on PC1. The second PC (PC2), has high correlation above 0.29 with ILE, ARG, LEU, PHE, LYS, as well as ASP (-0.41), but BALA, ALA, and TYR (correlation below 0.05). The third PC (PC3) has high correlation with GLU and THR (above 0.30), as well as GLY and BALA (negative, lower as -0.46). But it does not show much correlation with LYS, ALA, and TYR (absolute correlation below 0.06). Those variables that have higher loadings on each PC contribute more to the variance among data.

Using the student t test, correlation between two sample groups is significant if correlation coefficient $r > 0.28$ (80% significance) for sample number $n=11$. With this criterion, the variable that has correlation coefficient with any remained PCs above 0.28 should be kept for PCA. In this example, amino acids except MET will be remained for 80% significance. Without the non-significant correlated amino acids MET, the new PCA result is shown in Fig. 3.4, which is very similar to the one with all amino acids information. The grouping information does not changed, but the variance explained by the first three PC's increases slightly from 87% to 89%. Therefore, choosing significantly correlated amino acids for PCA can help increase relative variance explained by the remained PCs. It is very useful in many other cases when the number of variables is very large and some of the variables are not significantly distinct among samples. This is the first step to reduce the dimension of dataset without losing much information among samples.

There are some other questions regarding the operation of PCA. For example, there is still interest in finding a most approximate degradation direction among samples; through we have found in this study that the degradation trajectories do not simply follow a straight curve. There is question on how to rotate the axes to get a single degradation direction and to obtain all of the loadings of samples on that direction. This question is beyond the interesting of this study, but could be found in Appendix B in case of interesting.

3. Applications

3.1. Case 1: using 3D plots from PCA to visualize degradation trajectories of sinking particles with different settling velocities

Sinking particles are one of the main vehicles to transform OC from surface to deep ocean. The settling rate of sinking particles and the degradation rate of them are the two factors that determine how deep OC be transported into ocean, or how long CO_2 can be stored in the ocean. Degradation of sinking particles in water column has been well studied (Wakeham and Lee, 1993). However, it is still not very clear how different degradation is for particles with different settling rate. Sinking particles were collected using a NetTrap from 200 m in May 2003 at DYFAMED site in western Mediterranean Sea. These particles were separated by elutriation into 4 classes (A, B, C and D) with different settling velocity: >230 (A), 115-230(B), 58-115(C), and <58(D) m/day. To investigate their degradation trajectory with time, the four classes of particles were incubated for 5 days. Around 70 components (amino acids, pigments, lipids and carbohydrates) were measured for these four classes (ABCD) of particles at each time point: 0, 12, 24, 48, and 120 hours (Goutx et al., 2007; Lee et al., submitted). We tried to differentiate the degradation patterns among the four classes of sinking particles. PCA is an efficient statistical method to integrate this large dataset, and can help to sieve out the degradation patterns.

After the data are standardized with Eq. 3.3, the PCA results showed that the first few PCs explained 22%, 21%, 14%, 9%, 6%, 5%, 5%... of the total variance, respectively. Theoretically, the fourth eigenvalue is more like the elbow of the scree plot (Fig. 3.5a), and the first four PCs will explain 66% of the total variance. The 3D PCA result,

however, will be presented instead for visualization since 4D figure is harder to be shown by picture. Fig. 3.5b is a 3D plot including the first three PCs and their loading information. For the variables, it is known that GLU and ASP, chlorophyll, pyrophide, and fucoxanthin represent relatively fresh phytoplankton, either diatom or coccolithpores (Sheridan et al., 2002). In contrast, the amino acids BALA, GABA, the pigments phaeophytin, the deoxysugars fucose and rhamnose, and lipid metabolites represent a more degraded state. As Fig. 3.5b shown, the degradation direction is therefore, approximately from left to right on PC1, from backward to forward on PC2, and from upper to lower direction on PC3. With this 3D plot, particles of different settling velocities have different composition initially; the 3D coordinates for A and B are (-0.20 0.42 0.05) and (-0.36 -0.16 0.05), and they are (-0.26 -0.23 0.49) and (0.24 0.27 0.38) for C and D at time 0. Slower sinking particles (C and D) are more bacterial and zooplankton related due to the biomarker GABA along PC1. During the incubation, however, all the four classes go from initial stage to the more degraded direction (shown by arrows in Fig. 3.5b) with time, and become more similar in composition than they are in initial states. After 5 days, classes C (0.49 -0.03 -0.25) and D (0.46 -0.08 -0.16) are almost identical, A (-0.01 0.33 -0.23) and B (-0.11 -0.26 -0.30) both have tendency to merge to that point but might take longer time than 5 days. Moreover, slower sinking particles degraded more rapidly than faster sinking particles. In comparison, the 2-dimentional PCA (Fig. 3.5c) could not separate sinking particles B from C, as they are separated in a 3-D PCA. Clearly, the 3-D PCA is more consistent with results from cluster analysis (Fig. 3.5d). In contrast, the 1-dimensional PC gives little details on the degradation trajectories of sinking particles with different SV. Over all, with the original

dataset of 20×70 matrix, the 3D PCA makes it possible to directly visualize these degradation patterns of the sinking particles with different settling rates.

3.2. Case 2: using PCA to sort out the effects of mercuric chloride and protease inhibitors on the degradation of particles

Mercuric chloride *($HgCl_2$)* and protease inhibitors *(PI)* are often used to inhibit bacterial or enzymatic activity in environment samples. These inhibitors are necessary in preserving the sinking particles, which have to be collected in several months' periods using sediment traps, to make sure that enough amounts of materials are collected. However, how these inhibitors affect the composition of sinking particles is not known. Tens of organic compounds are usually measured to characterize the composition of particles. To evaluate the effect of inhibitors, PCA provides an efficient way of combining all the organic compound together, to discern the compositional changes.

In the experiment, we used a diatom culture of *Thalassiosira pseudonana* to represent marine particles, and $HgCl_2$ and PI were added to the diatom cultures. The culture was incubated for 33 days (Liu et al., 2006). A series of variables were measured, including amino acids, pigments, lipids, C and N at time zero and time 33 days. Ctrl-0, PI-0, and HG-0 were samples with control, PI-treated, $HgCl_2$-treated at time zero. Ctrl-33, PI-33, HG-33 and PI+HG-33 were samples with control, PI-treated, $HgCl_2$-treated, both PI- and $HgCl_2$-treated after 33 days. PCA was applied on the whole dataset to test how the particle composition changes between inhibitor treated and the control samples, and how the individual organic components had been preserved during the incubation.

PCA results showed that the second eigenvalue is the elbow in scree plot (Fig. 3.6a), and the first two PCs explain 43% and 25%, in total, 68% of the variance. The 2D plot clearly separate these 7 samples into three groups (Fig. 3.6c): 1) Ctrl-0, PI-0, and HG-0; 2) Ctrl-33 and PI-33; 3) HG-33 and PI+HG-33. The degradation direction is along with PC1, with the fresh on the right side to the degraded on the left side. The first group (projections 4.5, 3.3, and 3.2 on x axis) is on the right side along x axis, containing most organic compounds representing fresh diatomic matter, such as GLU, SER, and 24-methyl-5,24(28)-dien-3B-ol. There is no significant difference between samples with PI-treated, HG-treated and controlled one at the initial status, as expected. The second group (projections -3.8 and -7.0 on x axis), on the left side along PC1, has more bacterial degraded product, GABA. Within this group, PI-treated sample is similar in organic composition with the control sample, suggesting that PI does not inhibit the degradation. The third group (projections 1.4 and -1.7 on x axis) is located between group 1 and 2, with HG-33 more closed to group 1 along PC1, suggesting that HG did not fully inhibit the degradation of OM. Adding PI into HG-treated sample does not help much either. However, group 3 (projections -5.4 and -3.9 on y axis, see Fig. 3.6c) is also separated from group 1 (projections 2.2, 1.2, and 2.3 on y axis) and group 2 (projections 1.7 and 2.0 on y axis) along PC2, indicating Hg can also lead to specific compositional change of OM. The HG-treated samples are more enriched with amino acids THR, TYR, ASP, ARG, and fatty acids 16:0, 14:0. A significant fraction of Chl *a* was transformed to its allomer in the HG-treated samples.

This grouping character could also be seen by cluster analysis (Fig. 3.6d) results, the adding of PC3 (12% variance) with 3D structure (Fig. 3.6b) could increase total

variance to 80% in report. The grouping information, however, is very similar with that from 2D results, and is also consistent with cluster analysis. As the second eigenvalue is the elbow for scree plot, the third eigenvalue (5.0) is not comparable with the second one (10.3), and the first two PCs have already explained 68% of the variance, 2D plots for PCA are good enough for grouping these samples.

The effects of different inhibitors are clearly differentiated by the PCA plot in a semi-quantitative way. The results suggest that PI is not a good inhibitor at all, and Hg can cause some degree of degradation by chemical oxidation. In addition, the PCA plot allows one to quickly find out the inhibitor effect on specific organic compound. For example, 18:4, 20:5 were not preserved well while 16:0, 14:0 became enriched under the effect of Hg (Figs. 3.6b and 3.6c).

3.3. Case 3: using PCA to test association of natural radioisotopes with organic compounds in sinking particles

The natural radionuclides, daughter and parent pairs of $^{210}Po/^{210}Pb$ and $^{234}Th/^{238}U$, have been used to trace the fluxes of sinking particles in the ocean. This is based on the disequilibrium between the daughter and parent radioisotopes in the surface ocean due to their different particle activities and half-lives (Rama and Goldberg, 1961). In general, ^{210}Po and ^{234}Th have higher particle affinity than ^{210}Pb and ^{238}U, respectively. In addition to the deficiency of ^{234}Th or ^{210}Po relative to their parents, we also have to know the ratios of Th or Po to carbon. These ratios can be obtained from particles collected by either *in situ* pumps or sediment traps (Buesseler et al., 2006). Unfortunately, these ratios

vary significantly with particle types or size. One key factor in controlling Th/C or Po/C ratios is how these isotopes associate with the organic compounds, and the degrades of this association are unclear. In this study, we take advantage of all the chemical compound and isotope information in the sediment trap particles collected during the MedFlux project, and use PCA to find their correlations.

Time series sediment trap samples (14 cups in total) were collected at the DYFAMED site in the western Mediterranean Sea from March to June 2003 (Stewart et al., 2007). Natural radioisotopes ^{210}Po, ^{210}Pb, ^{234}Th, ^{228}Th, along with organic components (amino acids, lipids, and pigments), organic carbon, nitrogen, tracer metals (Fe, Zn, Ti), and minerals were measured. We use PCA to transform the data into the first few dimensions, which explains most of the variance among samples. The relationship between radioisotopes and different components will be investigated using the first few PCs.

In the original dataset, there are 86 variables for each of the 14 samples. PCA was performed after the data were standardized. In the preliminary runs, those insignificantly correlated variables are deleted for final PC result; only 34 variables are remained. The variances that could be explained by the first few PCs are 37%, 19%, 10%, 7%, 6%, 6%....etc. Although the second eigenvalue is more like the elbow from the scree plot (Fig. 3.7a), the first two PCs only gives 56% (lower than 60%) variance, which would increase to 66% by adding the PC3. In the 3D structure, samples are grouped into three (Fig. 3.7b). Samples collected in March (cup 1-4) were grouped together, enriched in diatom-related fecal pellets (28D5,22; 28D5, 24(28);SER, pyro-a, phytol, C20:1). Samples collected in April (cup 5-9) are in the second group, enriched in

bacterial degradation products (BALA, GABA, BR15/17). The remaining cups (10-14) collected in May and early June are in the third group, mainly fresh phytoplankton-derived (Chla, OC, TN, Fuco). The first PC could be seen as zooplankton degradation index, and the second PC is more about bacterial degradation index. The overall degradation index direction is approximately running from Chla to BALA, as shown in Fig. 3.7b. In the 2D structure, the grouping information is somewhat different with that from 3D, as cups 10 and 11 obviously go with the second group (Fig. 3.7c). Cluster analysis offers a way to check the grouping information, which is more consistent with the PCA results from 3D structure (Fig. 3.7d). 3D structure is therefore preferred even when the second eigenvalue is the elbow for scree plot but the first two PCs explain less than 60% variance.

The radionuclides ^{210}Po and ^{210}Pb are more abundant in samples 10-14, which are dominant by fresh phytoplankton-derived material. The OC, TN, Chl-a, fuco, Zn and amino acid MET, have very closed loading with these two radioisotopes and are more likely to be associated with them. The isotopes, ^{234}Th and ^{228}Th, are loaded close to both the phytoplankton-derived and bacterially-degraded groups. Overall, Th is more correlated with degraded materials than with ^{210}Po and ^{210}Pb. The possible explanation is that ^{210}Po can be taken up by phytoplankton cells, rather than simple absorption onto the cell surface, so that Po is associated with labile organic material. In contrast, ^{210}Pb is abundant in sample 10-14, mostly from the input of Saharan dust in May 2003 event; higher ^{234}Th and ^{228}Th in more degraded material may reflect the high surface to volume ratio of small particles, and ^{228}Th is probably more associated with degraded material than ^{234}Th as ^{228}Th has more positive loadings on both PC1 and PC2(Stewart et al., 2007).

With the aid of PCA, it is much easier to sift out the possible association mechanisms between radio isotopes and organic compounds, especially for a formable large dataset.

3.4. Case 4: using PCA to compare amino acid compositions in time series (TS) and settling velocity (SV) sediment trap samples

The newly invented SV traps can be used to separate particles based on their *in-situ* sinking velocities (Peterson et al., 2005). It is important to understand the compositional differences among particles with different SV's, if we are to decipher what factors control the SV's. Particles were collected with IRS time-series sediment traps at the DYFAMED site (200 m depth) in the western Mediterranean Sea from March to May, 2003. Settling velocity traps (there were two SV traps, and they were duplicated) and TS traps were deployed concurrently; these two types of traps collected the same particles, but with difference mode of processing the particles (Peterson et al., 2005). Seventeen types of amino acids were compared for both TS and SV samples, in order to investigate the correlation between amino acids composition and settling velocity, and how the TS and SV samples are correlated. Those cups that have similar sources for sinking particles in either TS or SV traps are assumed to have similar amino acids allocate distributions, which could be revealed by PCA.

PCA results show that there is no obvious elbow for the first 5 components in the scree plot (Fig. 3.8a). In order to visualize the data more easily, we take the first three PCs, which explained 32%, 22% and 17% (71% in total) of the total variance among samples, and make a 3D plot using these three PCs (Fig. 3.8b). The TS and SV samples

are positioned by their amino acids allocated information. Based on the degradation indicators of BALA and GABA, and the relative fresh amino acids of SER, GLY, THR and ASP, we define the degradation direction as from left to right side on PC1, and from negative to positive values on PC2. There is no obvious degradation information on PC3. Particles in TS02-06 are less degraded than those in later cups TS07-12. Particles in SV104, 204, 203, 205 are fresher than those in the other SV cups. TS samples and SV samples are basically separated on PC3, with TS series on the top of SV series. TS02-06, SV104, SV203, SV203 and SV205 have similar composition, which may be due to SV cup 04 collected the same particles as TS02-06. This grouping information is consistent with the results from cluster analysis (Fig. 3.8d). Since the flux of sinking particles in TS 02-06 accounts for 80% of the flux during the whole sampling period, and SV04 accounts for 32% of the total flux, only part of the particles that collected by TS 02-06 go into SV04, with the settling velocity in the range of 196-490 m/d. The rest of the particles collected during TS02-06 are separated into other SV traps. Particles in TS07-12 are grouped together, but are different on composition with any of the SV cup samples. The particles collected by SV mode traps are, therefore, more likely reallocated during the separation of them by different SV's. The two series SV trap samples (i.e. SV102-SV112 and SV203-SV212) are similar on their amino acids compositions for each corresponding cups as expected, and there is no significant distinction between these two series. Most of the SV particles are grouped together except SV102 103, 104, and 204. The first two PCs, which explained 54% variance in total, are explained with a 2D PCA plot (Fig. 3.8c) as well. An overlap does not exist in 3D PCA plot (Fig. 3.8b) and cluster analysis (Fig. 3.8d) between TS 11-12 and SV cups 111-112, but is found in 2D figure. The adding of

PC3 in this specific case is therefore, giving more accurate information than just using the first two PCs.

4. Conclusions and future work

This chapter has described the basic principles of PCA, and demonstrated how to use PCA to investigate geochemical problems with the focus on the sinking particles in the ocean. Our results clearly show that PCA is a powerful tool to seek relationships among many variables and to group the samples, when dealing with large dataset. Our results also show that constructing each 1-dimensional PC for "degradation index" is oversimplified, and the trajectory from the first 2 or 3 axes is more informative.

PCA, however, is only a semi-quantitative statistical approach. We group the samples close on loadings together, yet we often need to quantify the similarities among samples in the same group. Combining cluster analysis and PCA in a quantitative way may be a solution that needs to be explored. Organic matter including sinking particles is usually differentiated by source and degradation. While PCA provides a feasible approach to sort these factors out as this study showed, these two are often tangled together. These how to better separate these two factors and determine the real degradation index warrants further investigation. Moreover, it may be possible to construct a universal OM database from different environments including rivers, lakes, estuaries, costal oceans, and to map the OM compositions to their sample locations. With the database, it may be possible to tell the sources and degradation status for an unknown sample, using PCA.

References

Boehme, J., Coble, P., Conmy, R., and Stovall-Leonard, A., 2004. Examining CDOM fluorescence variability using principal component analysis: seasonal and regional modeling of three-dimensional fluorescence in the Gulf of Mexico. *Marine Chemistry* **89**, 3-14.

Buesseler, K. O., Benitez-Nelson, C. R., Moran, S. B., Burd, A., Charette, M., Cochran, J. K., Coppola, L., Fisher, N. S., Fowler, S. W., Gardner, W., Guo, L. D., Gustafsson, O., Lamborg, C., Masque, P., Miquel, J. C., Passow, U., Santschi, P. H., Savoye, N., Stewart, G., and Trull, T., 2006. An assessment of particulate organic carbon to thorium-234 ratios in the ocean and their impact on the application of Th-234 as a POC flux proxy. *Marine Chemistry* **100**, 213-233.

Cattell, R. B., 1966. The scree test for the number of factors. *Multivariate Behavioral Research* **1**, 245-276.

Dauwe, B. and Middelburg, J. J., 1998. Amino acids and hexosamines as indicators of organic matter degradation state in North Sea sediments. *Limnology & Oceanography* **43**, 782-798.

Dauwe, B., Middelburg, J. J., Herman, P. M. J., and Heip, C. H. R., 1999. Linking diagenetic alteration of amino acids and bulk organic matter reactivity. *Limnology & Oceanography* **44**, 1809-1814.

Dittmar, T., 2004. Evidence for terrigenous dissolved organic nitrogen in the Arctic deep sea. *Limnology & Oceanography* **49**, 148-156.

Goni, M. A., Yunker, M. B., Macdonald, R. W., and Eglinton, T. I., 2000. Distribution and sources of organic biomarkers in arctic sediments from the Mackenzie River and Beaufort Shelf. *Marine Chemistry* **71**, 23-51.

Gonzalez-Silvera, A., Santamaria-Del-Angela, E., Garcia, V. M. T., Garcia, C. A. E., Millan-Nunez, R., and Muller-Karger, F., 2004. Biogeographical regions of the tropical and subtropical Atlantic Ocean off South America: classification based on pigment (CZCS) and chlorophyll-a (SeaWiFS) variability. *Continental Shelf Research* **24**, 983-1000.

Goutx, M., Wakeham, S. G., Lee, C., Duflos, M., Guigue, C., Liu, Z. F., Moriceau, B., Sempere, R., Tedetti, M., and Xue, J. H., 2007. Composition and degradation of marine particles with different settling velocities in the northwestern Mediterranean Sea. *Limnology & Oceanography* **52**, 1645-1664.

Hu, J. F., Zhang, H. B., and Peng, P. A., 2006. Fatty acid composition of surface sediments in the subtropical Pearl River estuary and adjacent shelf, Southern China. *Estuarine Coastal and Shelf Science* **66**, 346-356.

King, J. R. and Jackson, D. A., 1999. Variable selection in large environmental data sets using principal components analysis. *Environmetrics* **10**, 67-77.

Lattin, J. M., Carroll, J. D., Green, P. E., and Green, P. E., 2003. *Analyzing multivariate data*. Thomson Brooks/Cole, Pacific Grove, CA.

Lee, C., Peterson, M. L., Wakeham, S. G., Armstrong, R. A., Cochran, J. K., Miquel, J. C., Fowler, S., Hirschberg, D., Beck, A., and Xue, J. Particulate organic matter and ballast fluxes measured using Time-Series and Settling Velocity sediment traps in the northwestern Sea. *Deep-Sea Research II* (submitted).

Liu, Z. F., Lee, C., and Wakeham, S. G., 2006. Effects of mercuric chloride and protease inhibitors on degradation of particulate organic matter from the diatom Thalassiosira pseudonana. *Organic Geochemistry* **37**, 1003-1018.

Mariano, A. J., Hitchcock, G. L., Ashjian, C. J., Olson, D. B., Rossby, T., Ryan, E., and Smith, S. L., 1996. Principal component analysis of biological and physical variability in a Gulf Stream meander crest. *Deep-Sea Research I* **43**, 1531-1565.

Moncheva, S., Gotsis-Skretas, O., Pagou, K., and Krastev, A., 2001. Phytoplankton blooms in Black Sea and Mediterranean coastal ecosystems subjected to anthropogenic eutrophication: Similarities and differences. *Estuarine Coastal and Shelf Science* **53**, 281-295.

Moreda-Pineiro, A., Marcos, A., Fisher, A., and Hill, S. J., 2001. Evaluation of the effect of data pre-treatment procedures on classical pattern recognition and principal components analysis: a case study for the geographical classification of tea. *Journal of Environmental Monitoring* **3**, 352-360.

Peterson, M. L., Wakeham, S. G., Lee, C., Askea, M. A., and Miquel, J. C., 2005. Novel techniques for collection of sinking particles in the ocean and determining their settling rates. *Limnology & Oceanography-Methods* **3**, 520-532.

Rama, K. M. and Goldberg, E. D., 1961. Lead-210 in natural waters. *Science* **134**, 98-99.

Schrimm, M., Buscail, R., and Adjeroud, M., 2004. Spatial variability of the biogeochemical composition of surface sediments in an insular coral reef ecosystem: Moorea, French Polynesia. *Estuarine Coastal and Shelf Science* **60**, 515-528.

Sheridan, C. C., Lee, C., Wakeham, S. G., and Bishop, J. K. B., 2002. Suspended particle organic composition and cycling in surface and midwaters of the equatorial Pacific Ocean. *Deep-Sea Research I* **49**, 1983-2008.

Stewart, G., Cochran, J. K., Xue, J. H., Lee, C., Wakeham, S. G., Armstrong, R. A., Masque, P., and Miquel, J. C., 2007. Exploring the connection between Po-210 and organic matter in the northwestern Mediterranean. *Deep-Sea Research I* **54**, 415-427.

Wakeham, S. G. and Lee, C., 1993. Production, transport, and alteration of particulate organic matter in the marine water column. In: Macko, M. E. a. S. (Ed.), *Organic Geochemistry*.

Yamashita, Y. and Tanoue, E., 2003. Distribution and alteration of amino acids in bulk DOM along a transect from bay to oceanic waters. *Marine Chemistry* **82**, 145-160.

Yunker, M. B., Belicka, L. L., Harvey, H. R., and Macdonald, R. W., 2005. Tracing the inputs and fate of marine and terrigenous organic matter in Arctic Ocean sediments: A multivariate analysis of lipid biomarkers. *Deep-Sea Research II* **52**, 3478-3508.

Yunker, M. B., Macdonald, R. W., Veltkamp, D. J., and Cretney, W. J., 1995. Terrestrial and marine biomarkers in a seasonally ice-covered arctic estuary - integration of multivariate and biomarker approaches. *Marine Chemistry* **49**, 1-50.

Zimmerman, A. R. and Canuel, E. A., 2001. Bulk organic matter and lipid biomarker composition of Chesapeake Bay surficial sediments as indicators of environmental processes. *Estuarine Coastal and Shelf Science* **53**, 319-341.

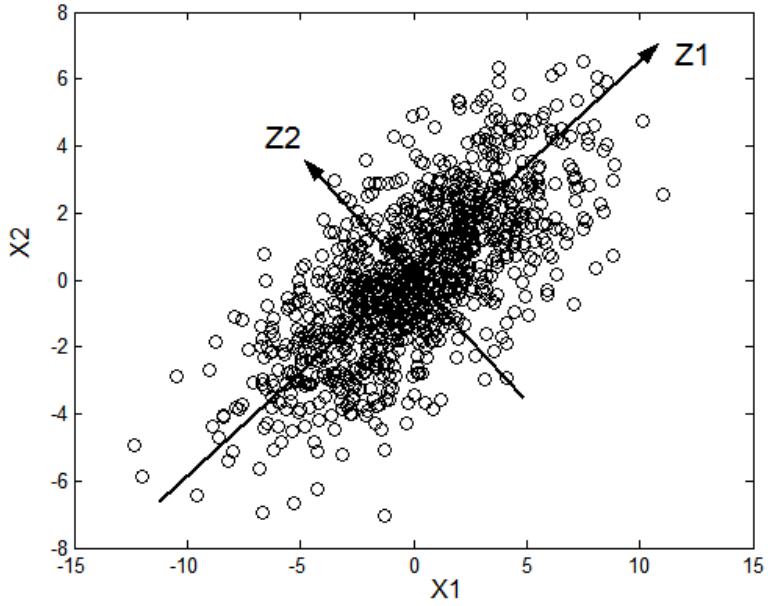
%	ASP	GLU	HIS	SER	ARG	GLY	THR	BALA	ALA	TYR	GABA	MET	VAL	PHE	ILE	LEU	LYS
TS-02	7.86	7.58	1.46	9.87	5.53	16.90	7.73	0.52	11.28	3.70	0.56	0.53	5.79	4.12	3.88	6.86	5.83
TS-03	9.89	8.56	1.59	10.43	4.69	16.11	7.28	0.38	11.47	3.80	0.42	0.55	5.48	3.83	3.52	6.52	5.50
TS-04	8.59	8.22	1.68	10.04	4.38	16.82	7.46	0.45	12.31	4.25	0.48	0.44	5.97	3.65	3.42	6.27	5.59
TS-05	10.88	9.81	1.42	10.32	4.03	15.37	7.57	0.56	11.94	3.96	0.43	0.43	5.65	3.48	3.13	5.98	5.05
TS-06	9.07	10.07	1.62	8.03	4.74	16.74	7.24	0.71	12.30	4.25	0.67	0.41	5.77	3.44	3.14	5.87	5.95
TS-07	7.06	8.35	1.87	7.41	4.88	16.79	7.46	0.66	13.12	4.65	0.70	0.41	6.78	3.66	3.61	6.37	6.22
TS-08	8.84	10.41	1.92	6.76	4.38	15.31	7.41	0.53	13.31	4.70	0.74	0.29	6.75	3.35	3.35	5.98	5.96
TS-09	7.67	10.46	1.99	7.23	4.46	15.05	7.57	0.58	13.45	4.60	0.80	0.31	6.55	3.37	3.28	6.11	6.54
TS-10	8.26	10.73	1.74	7.37	4.53	14.80	7.64	0.62	13.00	4.50	0.82	0.61	5.97	3.48	3.25	6.27	6.41
TS-11	9.27	11.24	1.66	6.72	4.39	13.75	7.94	0.29	13.03	4.62	0.83	0.37	6.31	3.56	3.47	6.38	6.19
TS-12	9.30	11.49	1.71	6.93	4.44	13.68	8.04	0.31	12.53	4.38	0.71	0.58	6.13	3.63	3.51	6.36	6.26

Table 3.1 The 17 amino acids mole% information for sinking particles collected by 11 cups in time-series sediment traps in DYFAMED site, March to May 2003.

	PC1	PC2	PC3
ASP	-0.108	-0.411	0.253
GLU	0.262	-0.160	0.304
HIS	0.292	0.137	-0.169
SER	-0.326	-0.136	-0.088
ARG	-0.160	0.393	-0.135
GLY	-0.199	0.070	-0.456
THR	0.069	0.166	0.474
BALA	0.043	-0.033	-0.462
ALA	0.348	0.036	-0.064
TYR	0.343	0.054	-0.042
GABA	0.308	0.181	0.079
MET	-0.184	0.095	0.252
VAL	0.282	0.194	-0.126
PHE	-0.287	0.291	0.087
ILE	-0.150	0.425	0.089
LEU	-0.193	0.380	0.184
LYS	0.258	0.296	0.043

Table 3.2 The loadings of variables on the first three PC's for amino acids data collected in DYFAMED site, March to May 2003. These loadings represent the correlation relationship of each variable with corresponding PC's.

(a)



(b)

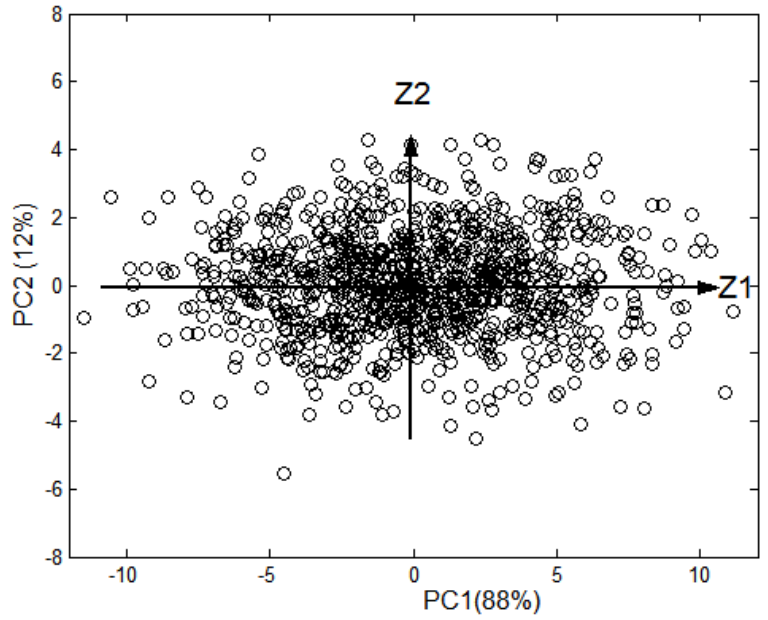


Figure 3.1 (a) The plot of an artificial dataset, which has two variables X_1 and X_2, for the explanation of PCA mechanisms. The maximum variance of dataset is along with Z_1, and the second maximum variance is along with Z_2. (b) The axes have been rotated to have Z_1 and Z_2 as the new axes; the variances among data along with both Z_1 and Z_2 are along listed.

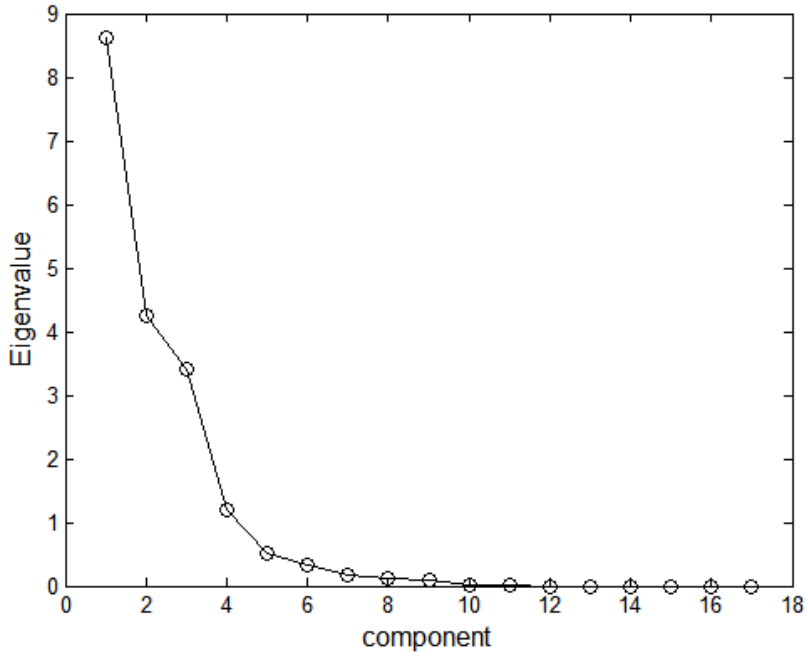
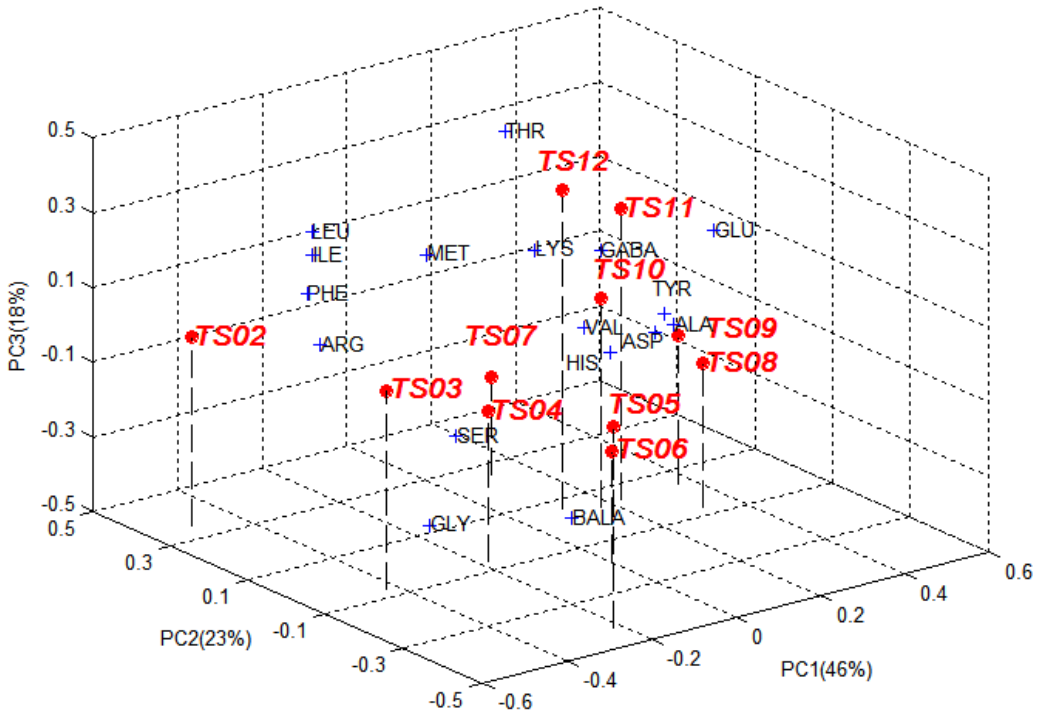
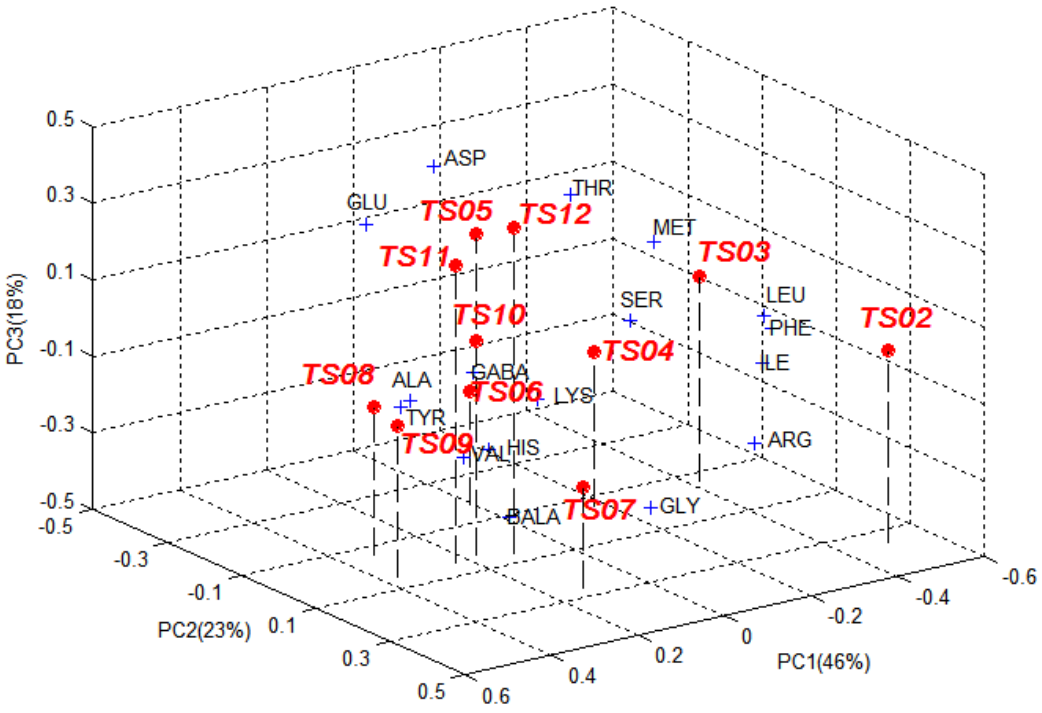


Figure 3.2 The scree plot of eigenvalues (i.e. variances explained by each principal component) for amino acids dataset of time series sediment trap samples in March to May 2003, DYFAMED site Mediterranean Sea. There is a significant decline in explained variance between the third and the fourth eigenvalues. This region is therefore, the 'elbow' for this curve; and the forth eigenvalue is not comparable with the third one.

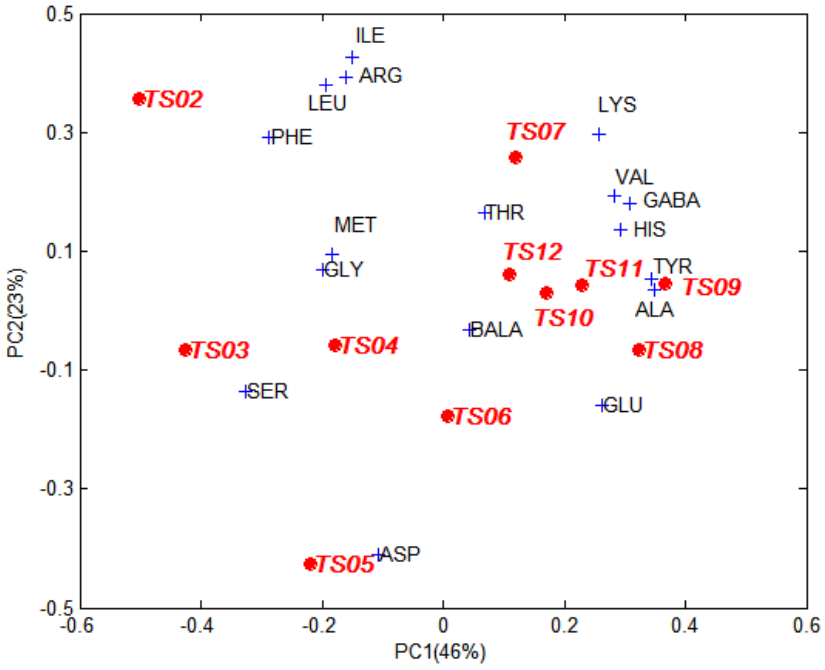
(a-1)


THR
TS12
TS11
LEU
ILE
MET
LYS
GABA
GLU
TS10
TYR
PHE
VAL
ALA
TS09
TS02
ARG
TS07
ASP
TS08
HIS
TS03
TS04
TS05
SER
TS06
GLY
BALA
PC3(18%)
PC2(23%)
PC1(46%)

(a-2)


ASP
THR
GLU
TS05
TS12
MET
TS11
TS03
SER
LEU
TS10
PHE
TS02
TS04
TS08
ALA
GABA
LE
TS06
LYS
TYR
ARG
TS09
VAL
HIS
BALA
TS07
GLY
PC3(18%)
PC2(23%)
PC1(46%)

(b)



(c)

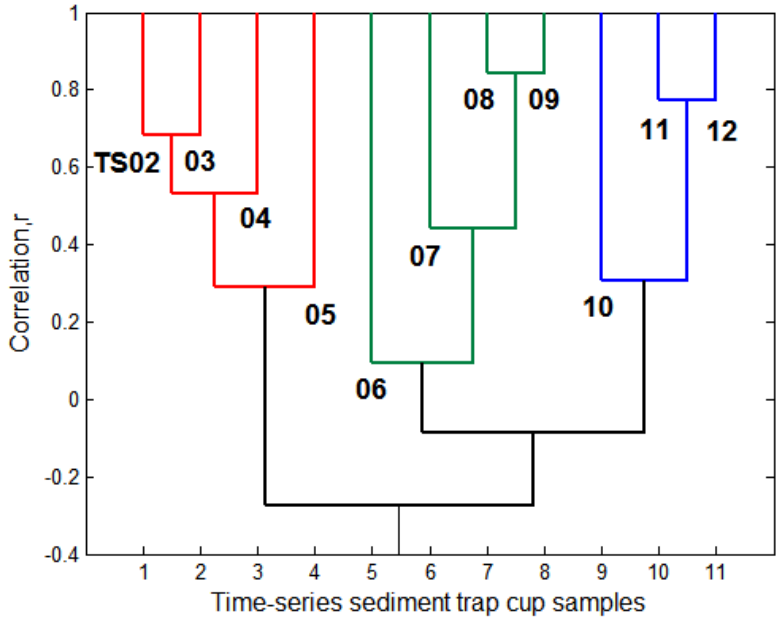
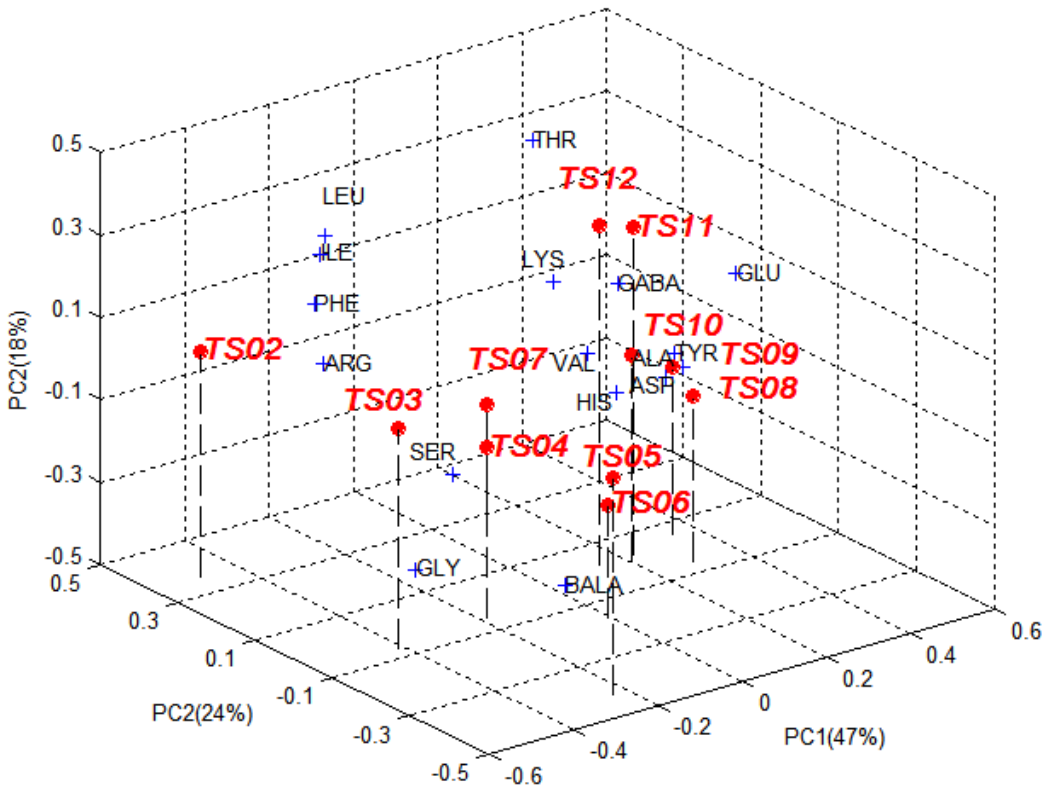


Figure 3.3 The plots of (a) the first 3 principal components (with two view angles, a-1 & a-2); (b) the first 2 principal components, (c) cluster analysis, for amino acid dataset of time series sediment trap samples in March to May 2003, DYFAMED site Mediterranean Sea. Eleven cups were grouped into three: 02-05, 06-09, and 10-12, consistent with the observation from 3D PCA shown by Fig. 3.3(a).

(1)


THR
TS12
LEU
TS11
ILE
LYS
GABA
GLU
PHE
TS10
TS02
ARG
TS07
VAL
TYR
TS09
ALA
TS03
ASP
TS08
HIS
SER
TS04
TS05
TS06
GLY
BALA
PC2(18%)
PC2(24%)
PC1(47%)
0.5
0.3
0.1
-0.1
-0.3
-0.5
0.6
0.4
0.2
0
-0.2
-0.4
-0.6

(2)

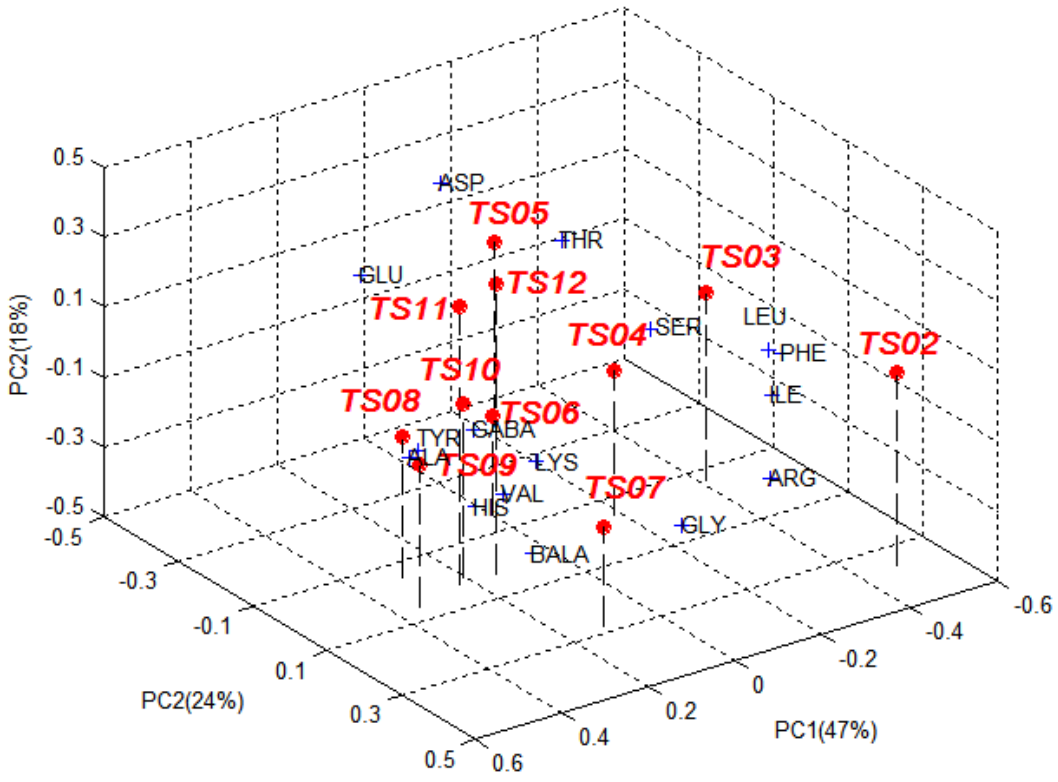
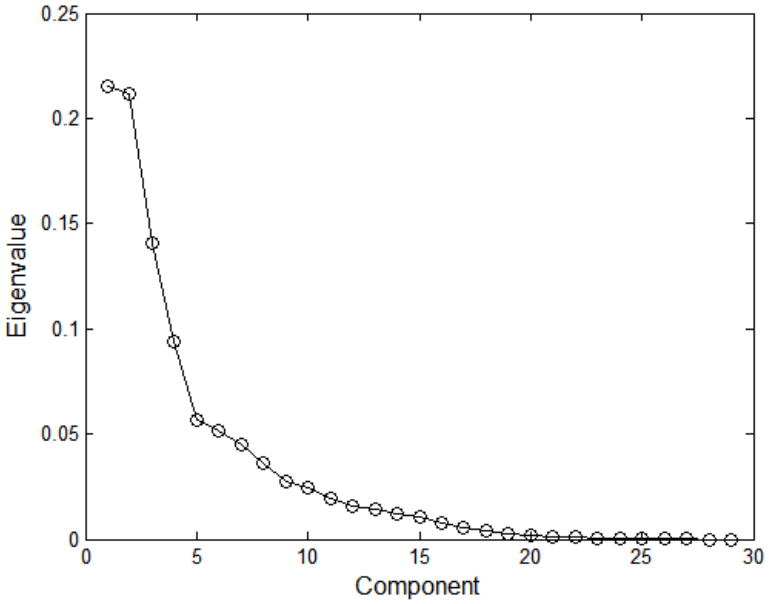
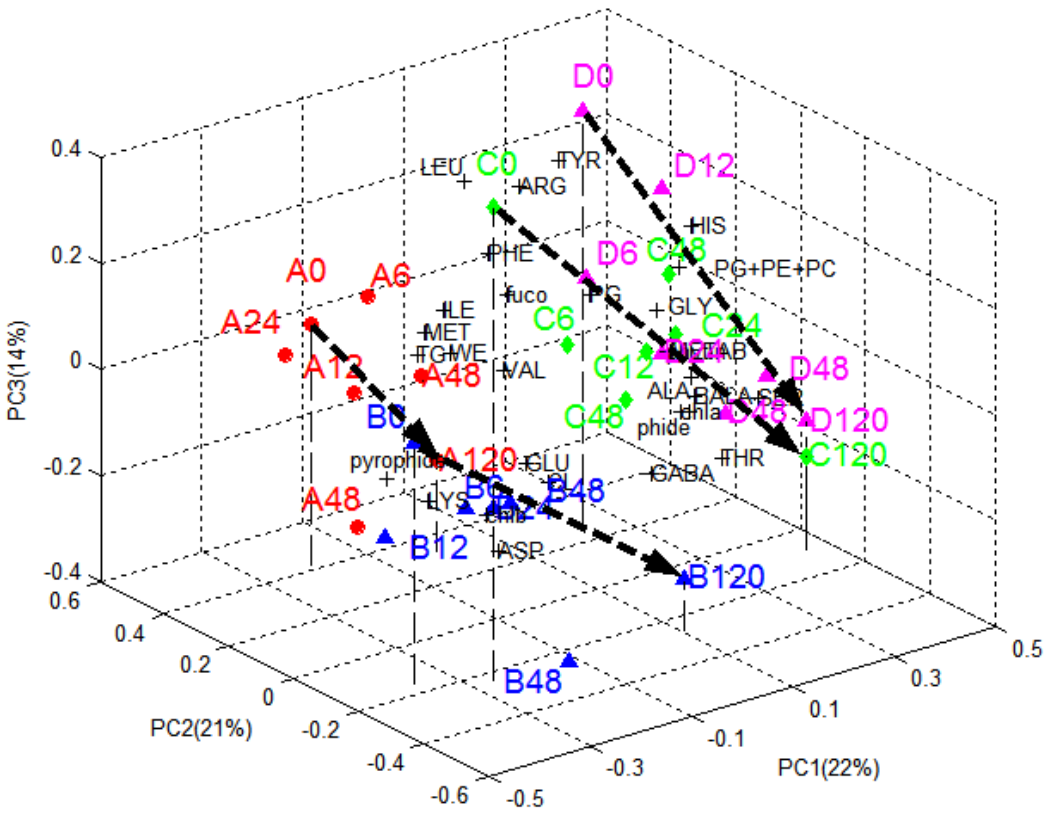


Figure 3.4 A test on the 3D plots (with two view angles,1 & 2) of the first 3 principal components for amino acids dataset of time series sediment trap samples in March to May 2003, DYFAMED site Mediterranean Sea with the variable amino acids MET deleted in original dataset.

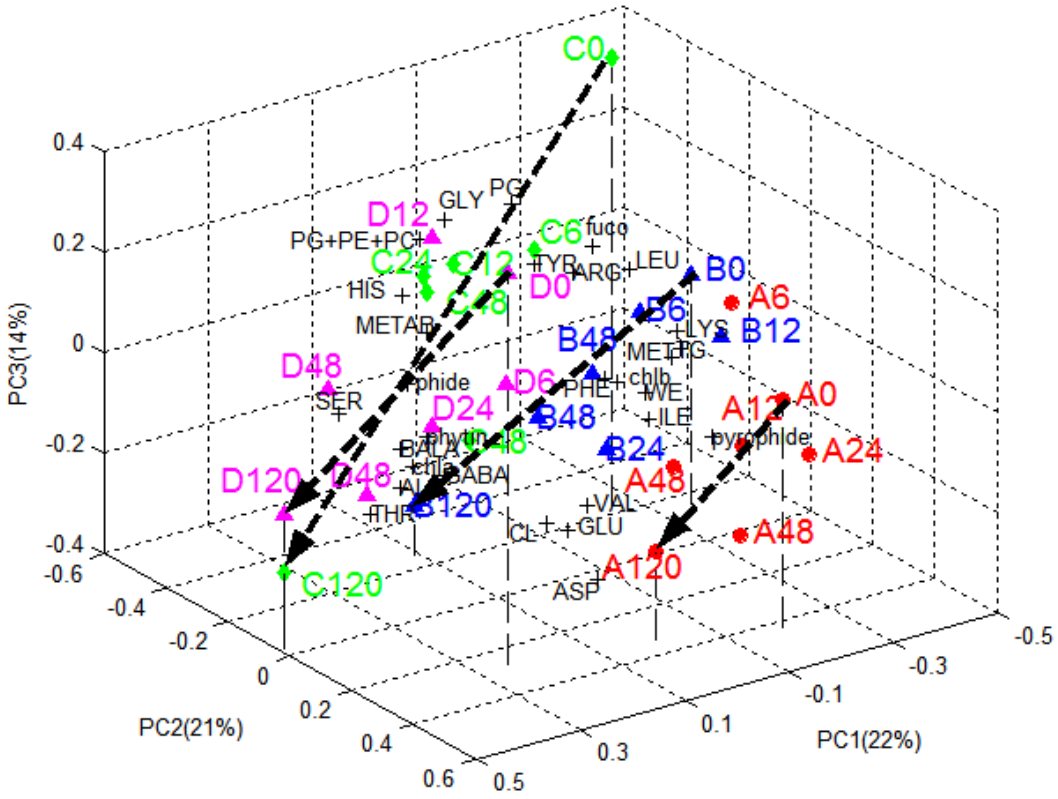
(a)


Eigenvalue
Component
0.25
0.2
0.15
0.1
0.05
0
0
5
10
15
20
25
30

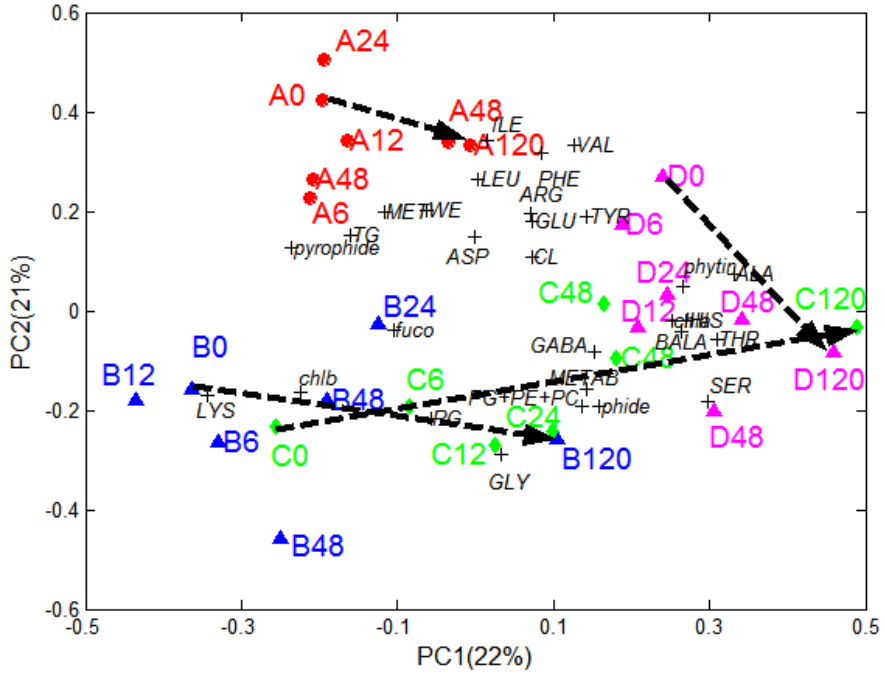
(b-1)


PC3(14%)
PC2(21%)
PC1(22%)
D0
C0
D12
D6
C48
C6
C24
C12
D48
C120
A0
A6
A24
A12
A48
B6
B12
B24
B48
B120
LEU
TYR
ARG
HIS
PHE
fuco
PG
PG+PE+PC
GLY
ILE
MET
TGWE
VAL
ALA
GABA
THR
GLU
ASP
LYS
phide

(b-2)


C0
PG
GLY
D12
PG+PE+PC
C6
fuco
C24
C12
TYR
ARG
LEU
B0
HIS
D0
A6
METAB
B6
LYS
B12
B48
MET
PG
D48
phide
chlb
D6
PHE
WE
SER
D24
B48
ILE
A12
A0
pyrophide
phytiq
BALA
C48
B24
A24
D120
D48
GABA
A48
ALA
B120
THR
VAL
CL
GLU
A48
C120
A120
ASP
PC1(22%)
PC2(21%)
PC3(14%)

(c)


A24
A0
A48
ILE
A12
A120
VAL
A48
LEU
PHE
ARG
D0
A6
MET
WE
GLU
TYR
D6
pyrophide
TG
ASP
CL
D24
phytiq
BALA
C48
B24
D12
D48
C120
fuco
chlb
GABA
THR
B0
B12
C6
MET
AB
SER
D120
chlb
B48
LYS
PG
PE+PC
phide
C24
B6
C0
C12
B120
D48
GLY
B48
PC1(22%)
PC2(21%)

(d)

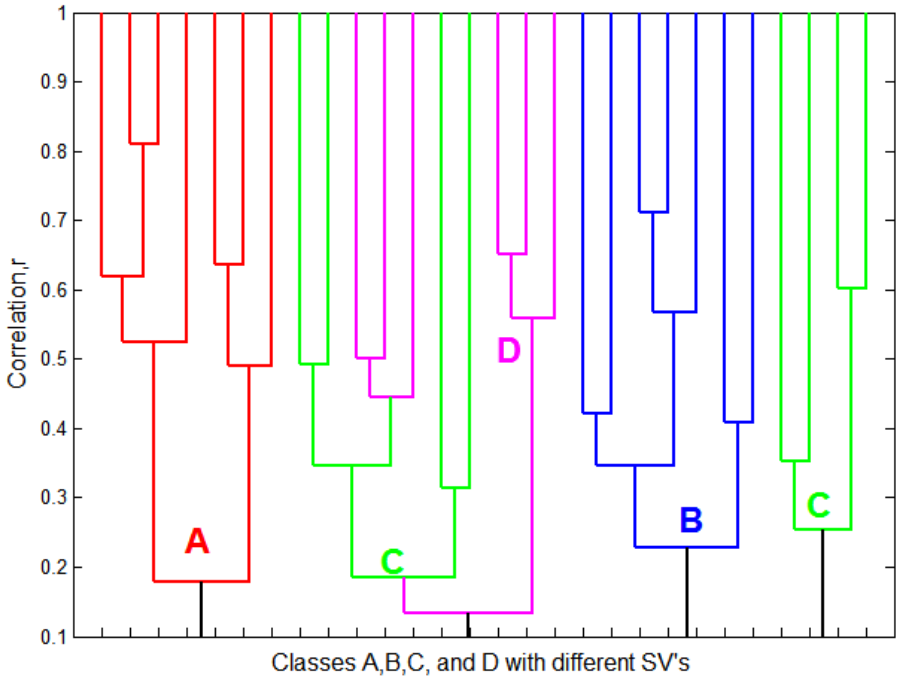
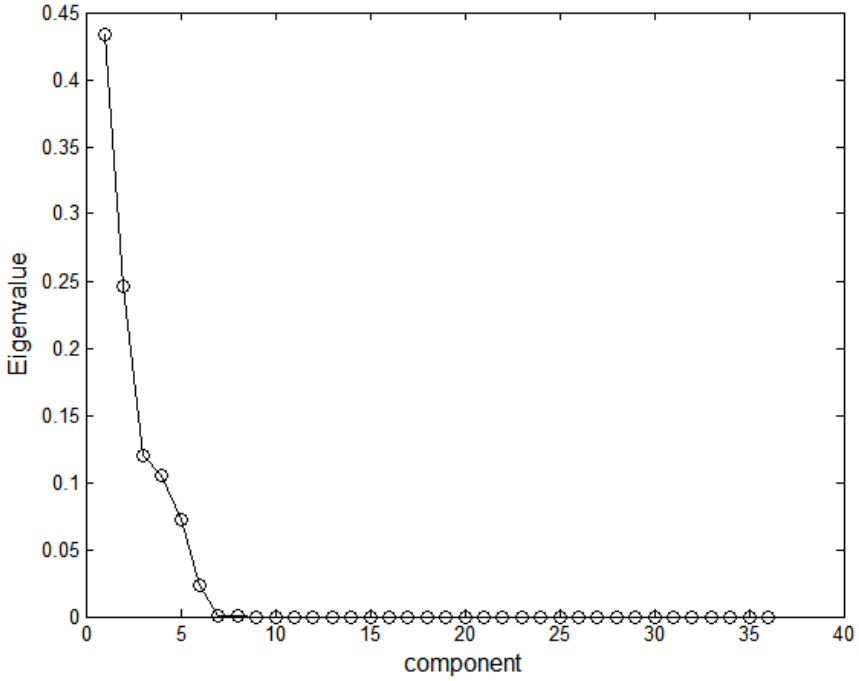
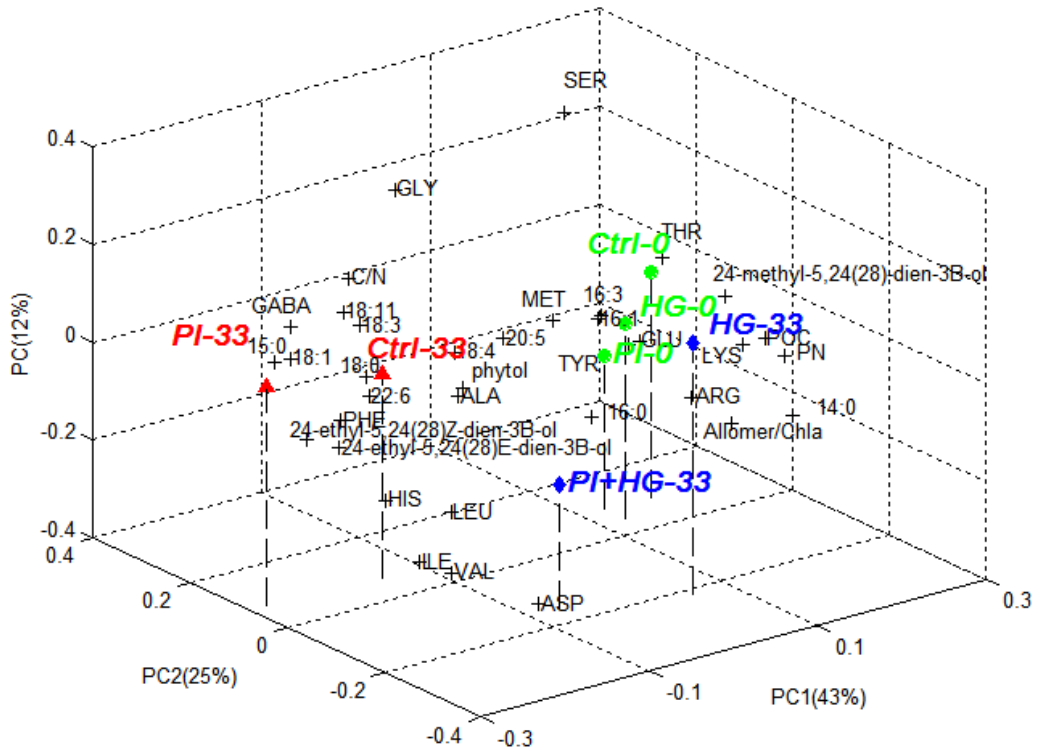


Figure 3.5 The (a) scree plot of eigenvalues; and plots of (b) the first 3 principal components (with two view angles, b-1 & b-2); (c) the first 2 principal components (see also Goutx et al., 2007), and (d) cluster analysis, to visualize the degradation trajectories of sinking particles with different settling velocities which were collected in MedFlux. Classes B and C were completely separated by their compositions on cluster analysis, which is consistent well with 3D PCA on Fig. 3.5(b), but they were not shown on 2D PCA (Fig. 3.5(c)).

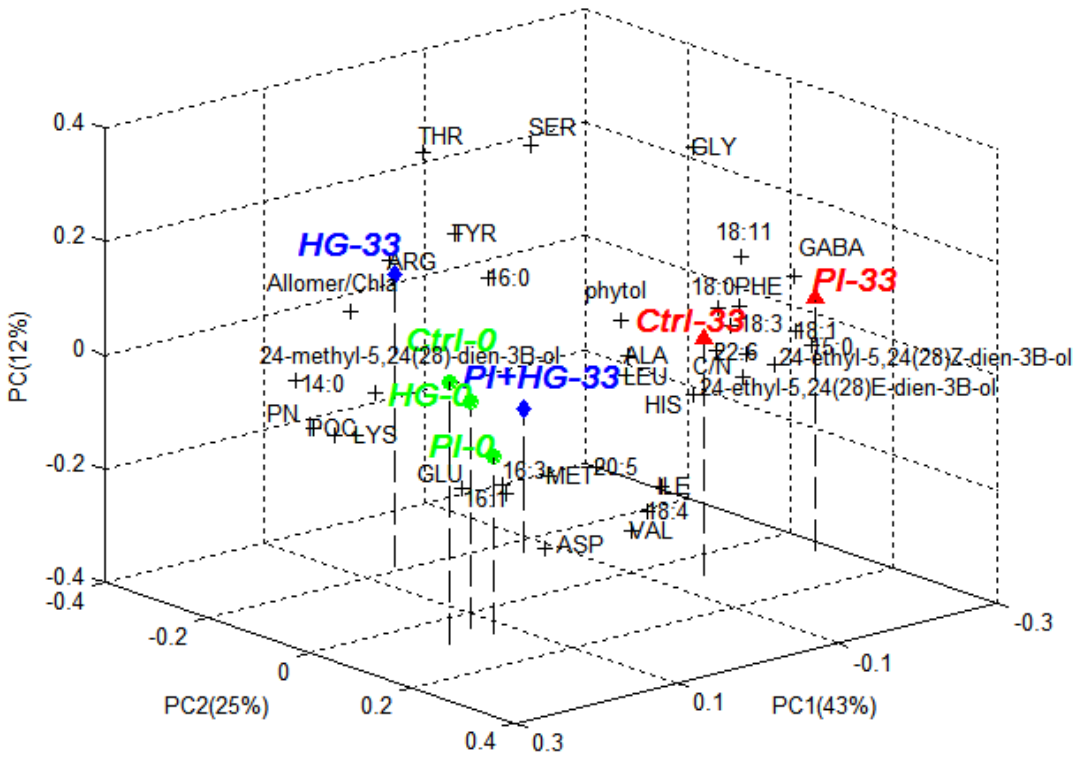
(a)


Eigenvalue
component
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
0
5
10
15
20
25
30
35
40

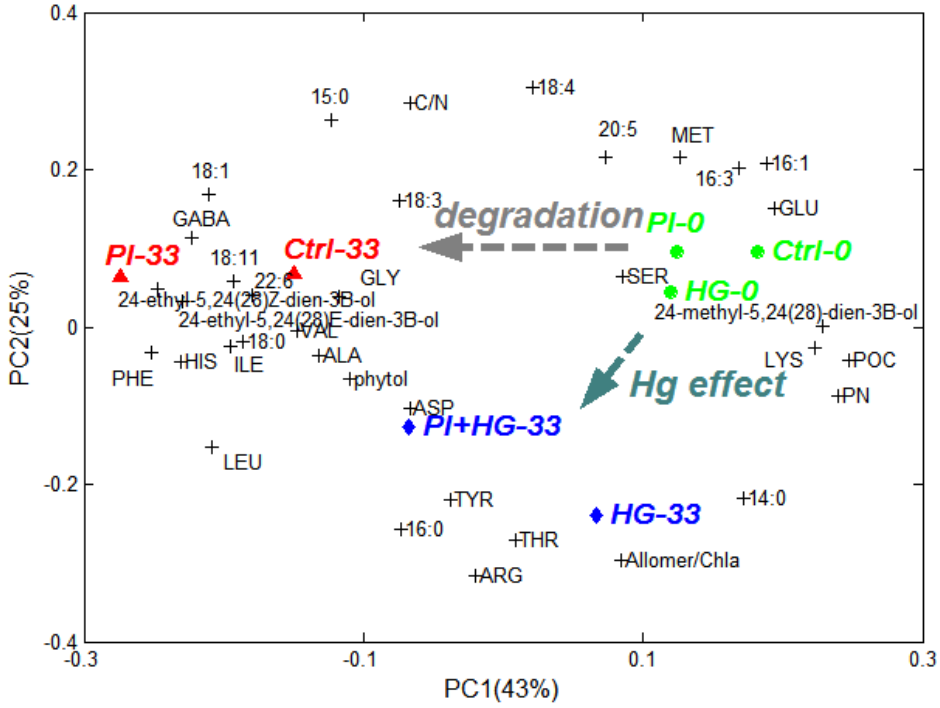
(b-1)


PC(12%)
PC2(25%)
PC1(43%)
SER
GLY
THR
C/N
GABA
MET
16:3
16:4
GLU
LYS
POC
PN
15:0
18:1
18:3
18:4
20:5
phytol
TYR
ALA
ARG
14:0
16:0
22:6
PHE
24-ethyl-5,24(28)Z-dien-3B-ol
24-ethyl-5,24(28)E-dien-3B-ol
24-methyl-5,24(28)-dien-3B-ol
Allomer/Chla
HIS
LEU
ILE
VAL
ASP
Ctrl-0
HG-0
HG-33
PI-0
PI-33
Ctrl-33
PI+HG-33

(b-2)


PC(12%)
PC2(25%)
PC1(43%)
THR
SER
GLY
HG-33
TYR
ARG
18:11
GABA
Allomer/Chla
16:0
phytol
18:0
PHE
PI-33
Ctrl-33
Ctrl-0
24-methyl-5,24(28)-dien-3B-ol
ALA
LEU
C/N
24-ethyl-5,24(28)Z-dien-3B-ol
14:0
PI+HG-33
HIS
24-ethyl-5,24(28)E-dien-3B-ol
HG-0
PN
POC
LYS
PI-0
GLU
16:3
MET
20:5
16:1
ILE
18:4
VAL
ASP

(c)


PC2(25%)
PC1(43%)
15:0
C/N
18:4
20:5
MET
16:3
16:1
18:1
18:3
GLU
GABA
degradation
PI-0
PI-33
18:11
Ctrl-33
Ctrl-0
GLY
SER
22:6
HG-0
24-ethyl-5,24(28)Z-dien-3B-ol
24-methyl-5,24(28)-dien-3B-ol
24-ethyl-5,24(28)E-dien-3B-ol
VAL
18:0
LYS
POC
HIS
ILE
ALA
PHE
phytol
Hg effect
PN
ASP
PI+HG-33
LEU
TYR
14:0
HG-33
16:0
THR
Allomer/Chla
ARG

(d)

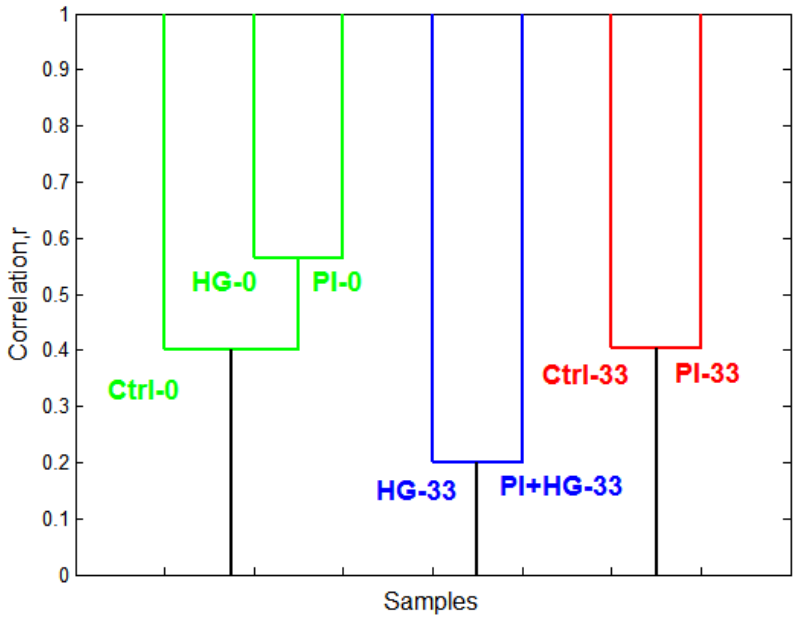
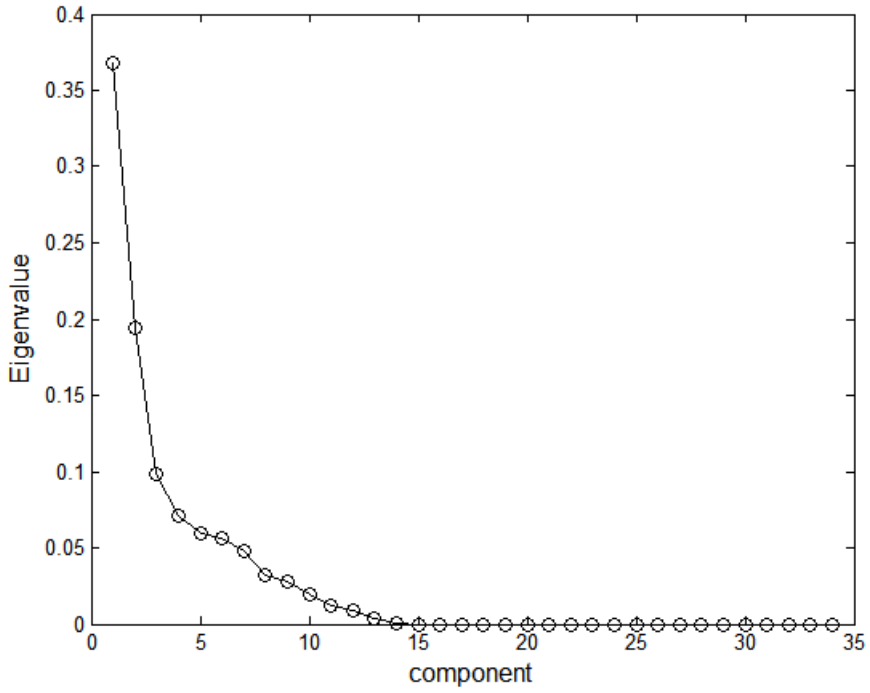
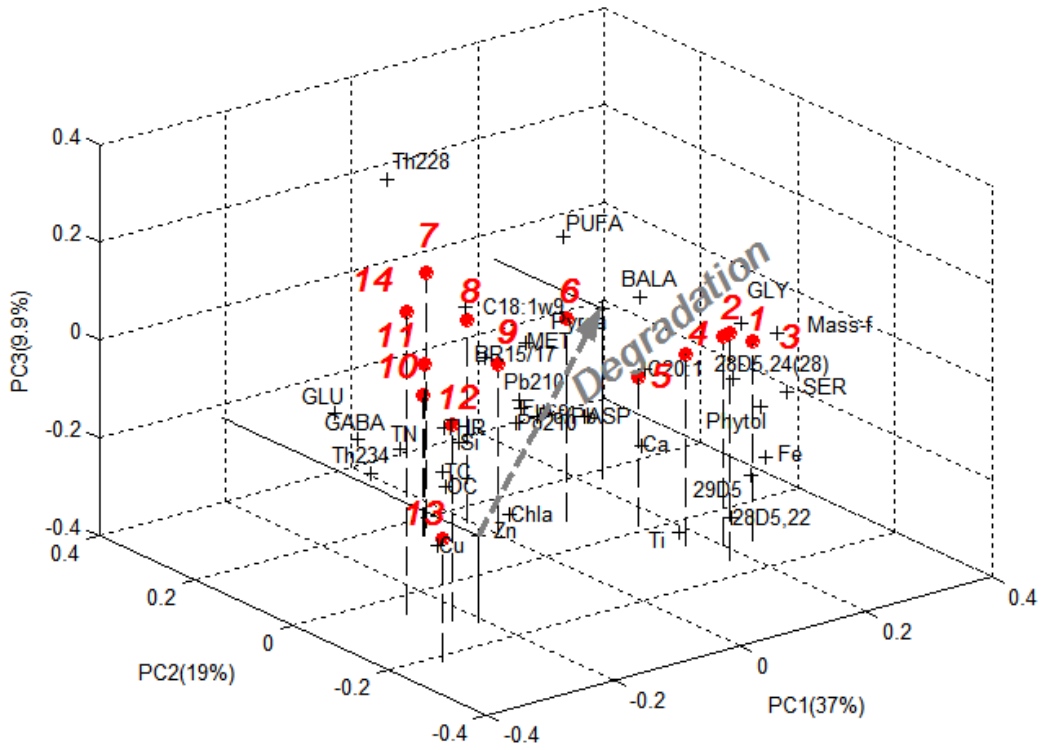


Figure 3.6 The (a) scree plot of eigenvalues; and plots of (b) the first 3 principal components (with two view angles, b-1 & b-2); (c) the first 2 principal components (see also Liu et al., 2006), (d) cluster analysis, to find out the effects of mercuric chloride and protease inhibitors on the degradation of particles in MedFlux. This cluster grouping information is consistent with both 2D and 3D PCA results.

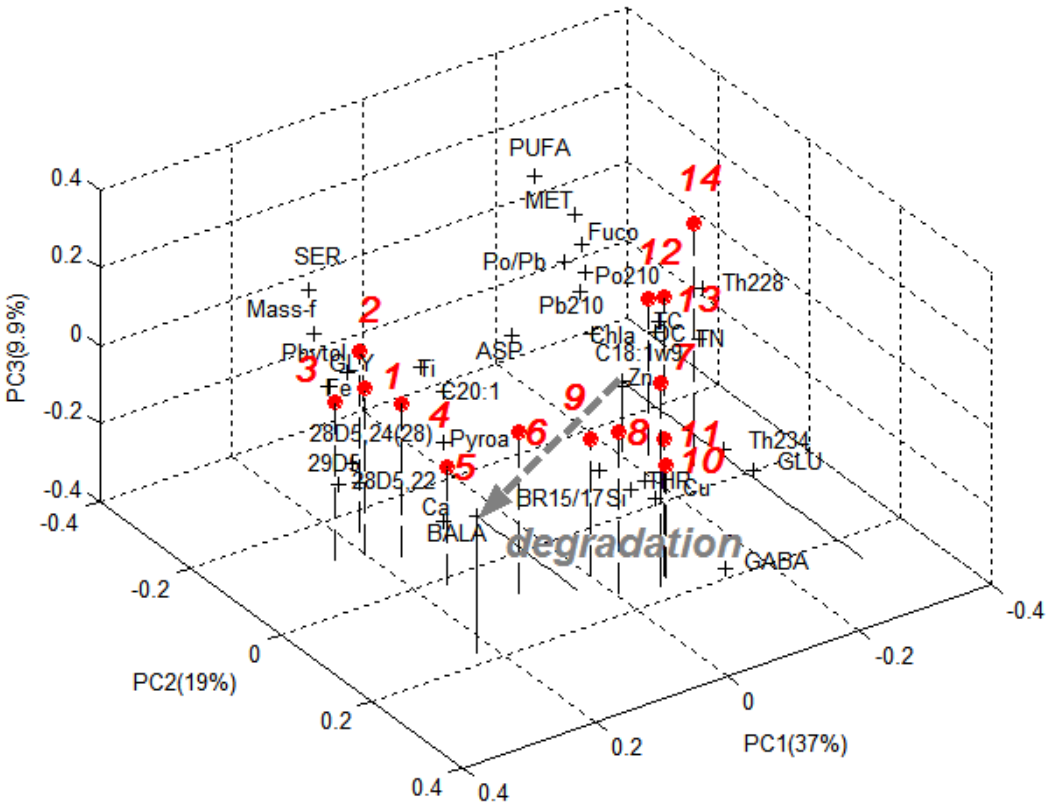
(a)


Eigenvalue
component
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
0
5
10
15
20
25
30
35

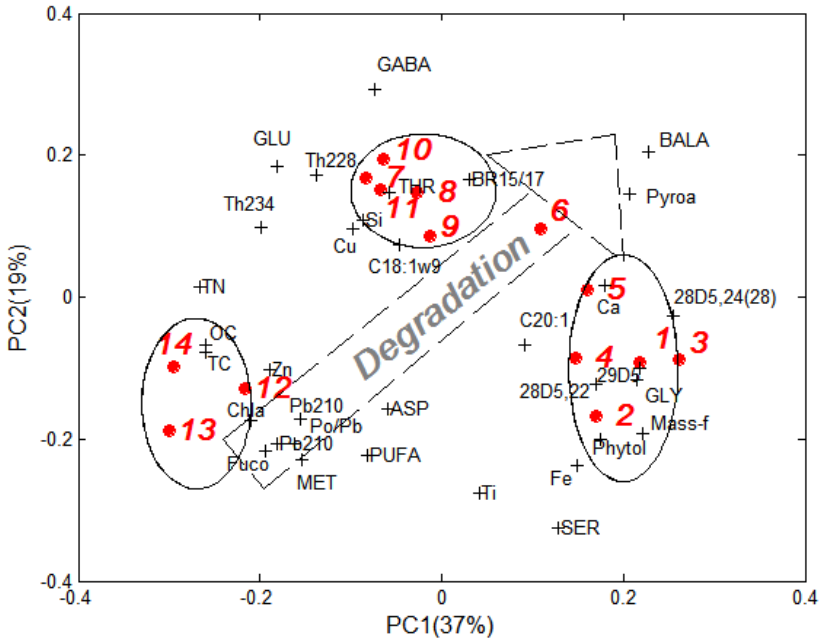
(b-1)


PC3(9.9%)
PC2(19%)
PC1(37%)
Degradation
Th228
PUFA
BALA
GLY
Mass-f
SER
Phytol
Ca
Fe
Ti
Chla
Zn
Cu
GLU
GABA
TN
Th234
Pb210
MET
C18:1w9
28D5,24(28)
29D5
28D5,22
1
2
3
4
5
6
7
8
9
10
11
12
13
14

(b-2)


PUFA
MET
Fuco
SER
Po/Pb
Po210
Th228
Mass-f
Pb210
Chla
OC
TN
Phytol
ASP
C18:1w9
Zn
Ti
C20:1
Fe
Pyroa
28D5,24(28)
29D5
28D5,22
Th234
GLU
THR
Cu
Ca
BR15/17
Si
BALA
degradation
GABA
PC3(9.9%)
PC2(19%)
PC1(37%)

(c)


GABA
GLU
Th228
BALA
THR
BR15/17
Pyroa
Th234
Si
Cu
C18:1w9
TN
OC
TC
Zn
Chla
Pb210
Po/Pb
ASP
Pb210
Fuco
PUFA
MET
Degradation
Ca
28D5,24(28)
C20:1
29D5
28D5,22
GLY
Mass-f
Phytol
Fe
Ti
SER
PC2(19%)
PC1(37%)

(d)

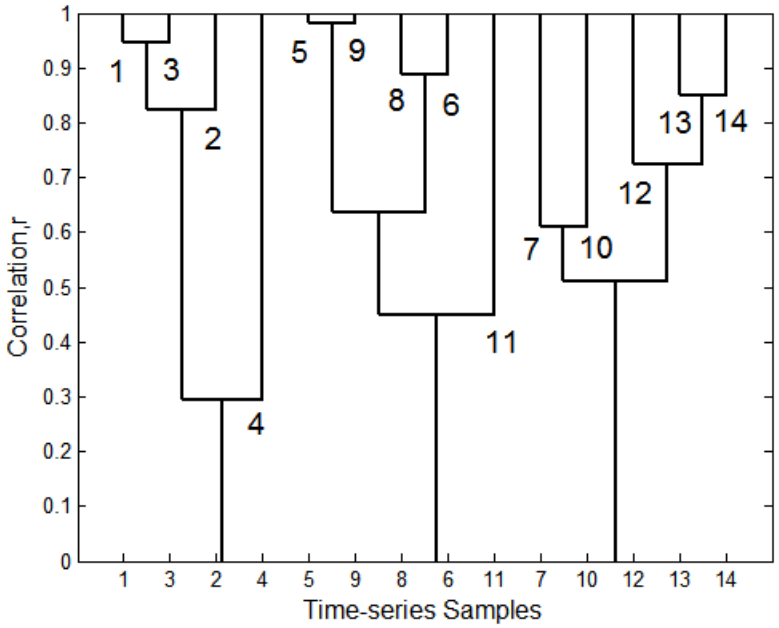
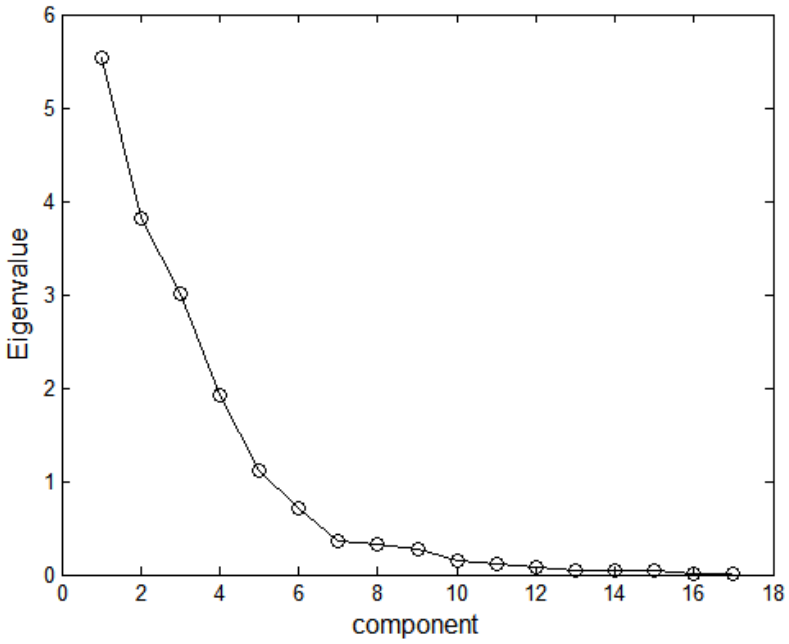
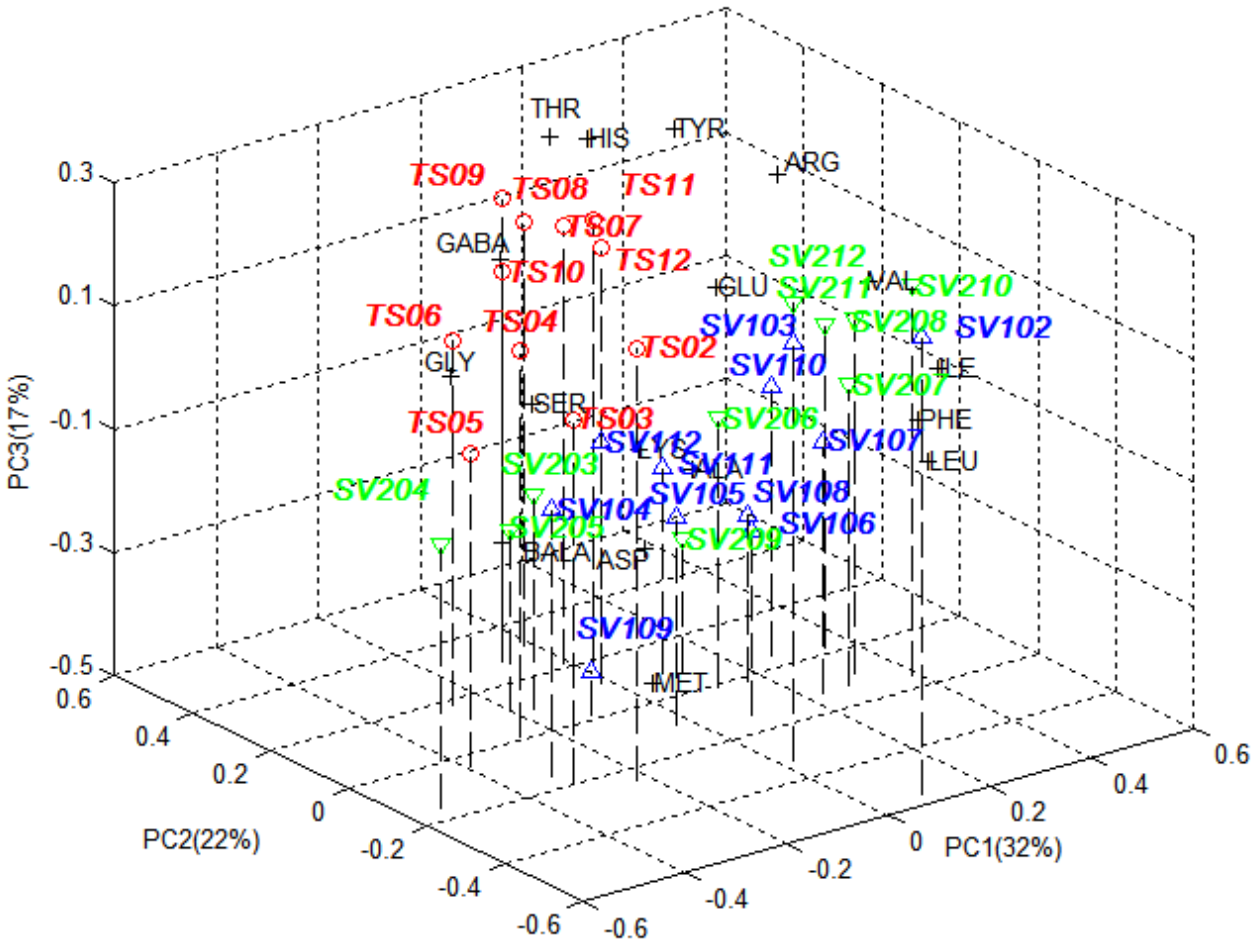


Figure 3.7 The (a) scree plot of eigenvalues; and plots of (b) the first 3 principal components (with two view angles, b-1 & b-2); (c) the first 2 principal components (See also Stewart et al., 2007); and (d) cluster analysis, to test the association of natural radioisotopes with organic compounds in sinking particles. This cluster grouping information is more consistent with the 3D PCA results on Fig. 3.7(b).

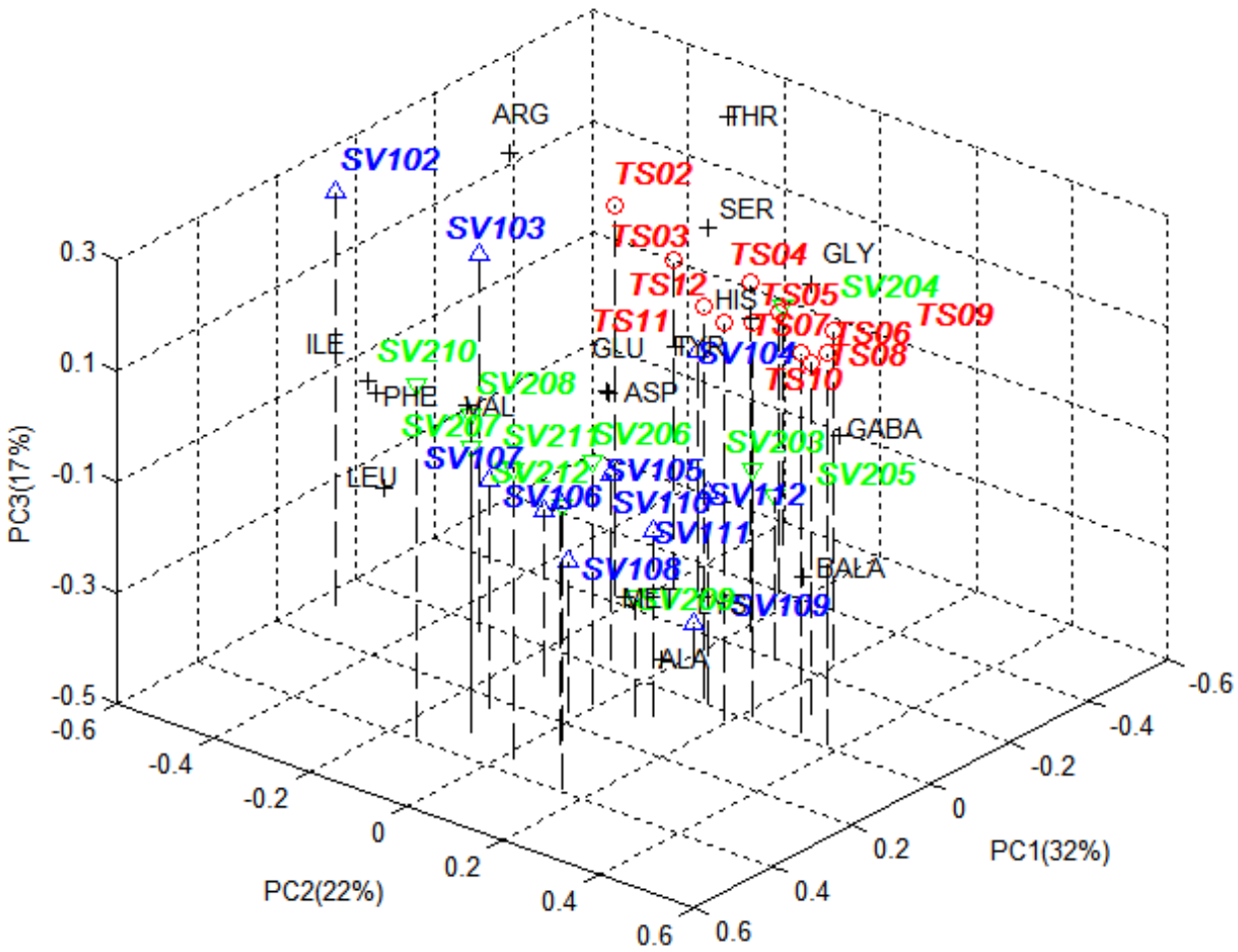
(a)



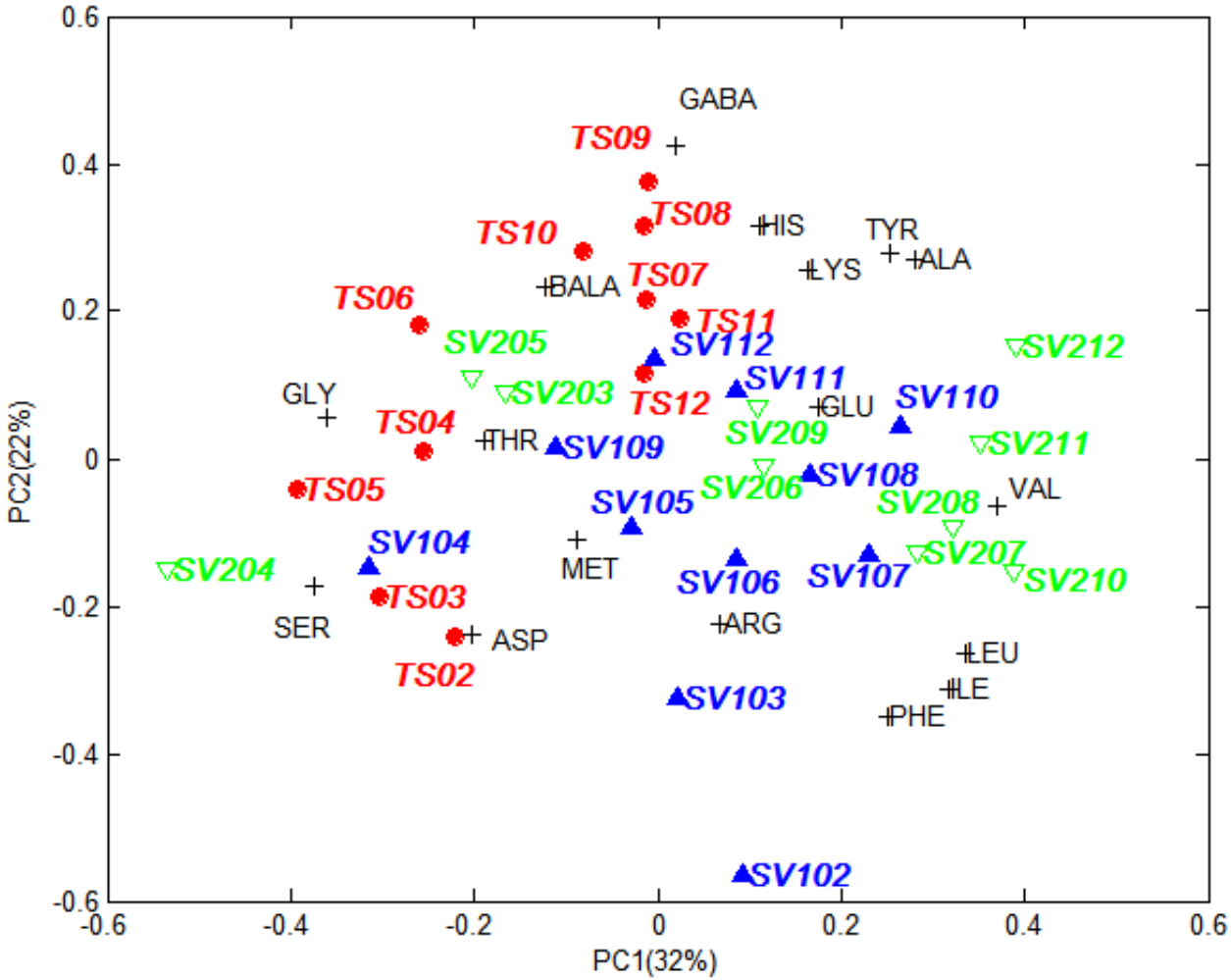
(b-1)


THR
HIS
TYR
ARG
TS09
TS08
TS11
TS07
GABA
TS10
TS12
SV212
GLU
SV211
VAL
SV210
TS06
TS04
SV103
SV208
SV102
TS02
GLY
SV110
ILE
SV207
SER
TS05
TS03
SV206
PHE
SV112
SV107
SV203
SV111
LEU
SV204
SV105
SV108
SV104
SV106
SV205
SV209
BALA
ASP
SV109
MET
PC3(17%)
0.3
0.1
-0.1
-0.3
-0.5
PC2(22%)
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
PC1(32%)
0.6
0.4
0.2
0
-0.2
-0.4
-0.6

(b-2)


ARG
THR
SV102
TS02
SER
SV103
TS03
TS04
GLY
TS12
HIS
TS05
SV204
TS11
TS07
TS06
TS09
ILE
SV210
GLU
SV104
TS08
TS10
SV208
PHE
VAL
ASP
SV207
SV211
SV206
GABA
SV107
SV203
SV212
SV105
SV205
LEU
SV106
SV110
SV112
SV111
SV108
BALA
SV209
SV109
ALA
PC3(17%)
PC2(22%)
PC1(32%)
0.3
0.1
-0.1
-0.3
-0.5
-0.6
-0.4
-0.2
0
0.2
0.4
0.6

(c)


GABA
TS09
TS08
TS10
HIS
TYR
ALA
LYS
BALA
TS07
TS06
TS11
SV205
SV112
SV212
SV111
SV203
TS12
GLU
SV110
GLY
TS04
THR
SV109
SV209
SV211
SV108
TS05
SV206
SV105
SV208
VAL
SV104
MET
SV207
SV204
SV106
SV107
SV210
TS03
SER
ARG
ASP
LEU
TS02
SV103
ILE
PHE
SV102
PC2(22%)
PC1(32%)
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.6
-0.4
-0.2
0
0.2
0.4
0.6

(d)

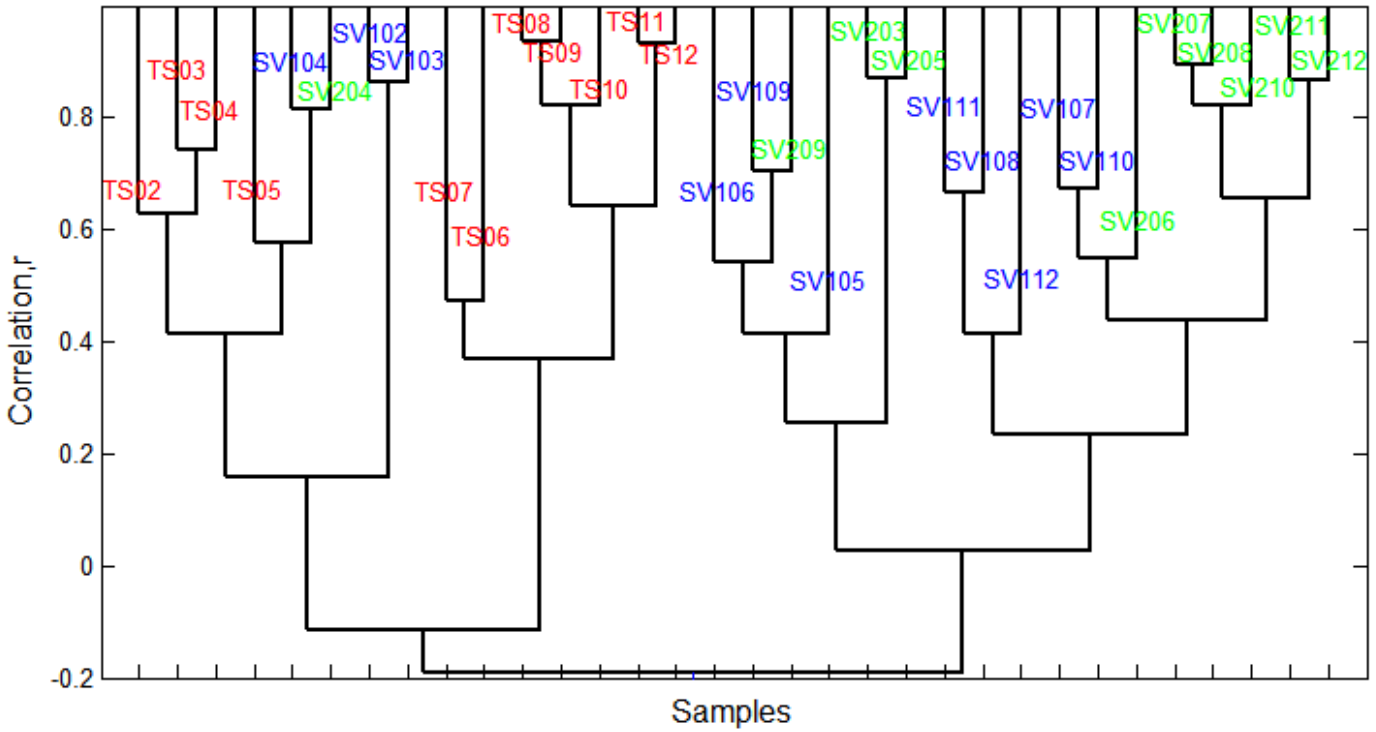


Figure 3.8 The (a) scree plot of eigenvalues; and plots of (b) the first 3 principal components (with two view angles, b-1 & b-2); (c) the first 2 principal components; and (d) cluster analysis, to compare the amino acids organic compounds in sinking particles collected from time-series sediment traps or SV mode sediment traps. This grouping information is more consistent with 3D PCA plot in Fig. 3.8(b).

CHAPTER FOUR: The difference (mass deficit) between measured and calculated masses of sinking particles in the ocean

Abstract

In 2003 and 2005, sinking particles were collected from depths of ~200m, ~800m, and ~1800m at the DYFAMED site in the Mediterranean Sea as part of the MedFlux project, using both time-series and settling-velocity sediment traps. We analyzed the major chemical components of these sinking particles, including organic matter (OM), opal, $CaCO_3$, and lithogenic minerals. At DYFAMED, the sum of the masses of these components is often less than the mass that is measured directly, leading to an observed mass deficit. For comparison, we also analyzed US JGOFS data from several sites, using the same algorithms as we had applied to MedFlux data. We found that the estimated mass deficits for sinking particles from the (relatively shallow) Ross Sea were large, but that particles collected from deep-ocean sites, including the Arabian Sea, equatorial Pacific, Southern Ocean, and North Atlantic, had mass deficits close to zero. Our analyses also showed that OM content is directly related to mass deficit, especially in shallow water where OM content is high. We hypothesize that water bound to organic molecules or to minerals in the sinking particles, especially those from shallow depths, may not be totally removed during oven-drying, and that tightly-bound water molecules may cause the observed mass deficit. Analytical error may also contribute variance to the

mass deficit, especially where sample size is small, although our study suggests that it is not the major factor contributing to mass deficit.

1. Introduction

Sinking particles are the main vehicles transferring carbon from surface to deep ocean waters, temporarily or permanently removing carbon from the active carbon cycle. In this respect, sinking particles play a very important role in reducing global warming (Sarmiento and Gruber, 2006). This carbon transport is often called the biological pump, and its efficiency is directly determined by the composition and settling velocity of sinking particles: how much carbon is carried by sinking particles and how fast it is transported determine the amount of remineralization that can occur with depth.

Sinking particles in the ocean typically consist of organic matter (OM) and ballast minerals. The ballast minerals, as defined here, include biogenic opal, calcium carbonate ($CaCO_3$), and lithogenic minerals. During the transit from surface waters to the ocean interior, most of the organic carbon in sinking particles is remineralized at depths of 100-1000 m, in the so-called twilight zone, releasing inorganic carbon to surrounding seawater (Wakeham et al., 1997). Sinking particles settle faster when they are by becoming large and when they contain ballast minerals that have greater density. By analyzing extensive sediment trap data, Klaas and Archer (2002) found that most of the organic matter (OM) (83%) in sinking particles is associated with $CaCO_3$, which is much denser than opal and more abundant than terrestrial minerals.

The flux and composition of sinking particles have been extensively studied in different ocean areas (e.g., Lee and Cronin, 1982; Wakeham et al., 1984; Honjo and Manganini, 1993; Honjo et al., 1995; Honjo et al., 1999; Collier et al., 2000; Honjo et al., 2000; Prahl et al., 2000). The total mass of sinking particles is an important parameter,

not only for representing the total amount of the bulk materials being transferred from surface to deep waters, but also for providing information about the relationship between organic and inorganic components. Total mass in sediment traps can be obtained directly by measuring the mass of sinking particles after the particles are oven dried at 60°C, or indirectly by summing the masses of the major chemical components. To calculate the mass of sinking particles indirectly, we summed the masses of four components: organic matter (OM), $CaCO_3$, opal, and lithogenic minerals. These masses were based on measured elemental concentrations: organic carbon (OC) for OM, inorganic carbon (IC) for $CaCO_3$, biogenic Si for opal, and lithogenic Al or Ti for lithogenic minerals (Honjo et al., 2000). By summing the four components, one can obtain a calculated mass. Since Al and Ti concentration data are not always available, or are relatively low in sinking particles and are difficult to accurately measure, lithogenic minerals are often calculated by subtracting the masses of OM, $CaCO_3$, and opal from the measured mass, under the assumption that the measured mass matches well with calculated mass (Haake et al., 1993; Takahashi et al., 2000). Using US JGOFS data, however, we will show that the sum of OM, opal, $CaCO_3$, and lithogenic minerals is systematically lower than the measured mass at several sites; that is, that there is a mass deficit. Here we define mass deficit *MD* as the difference between 1 and the ratio of the sum $\sum m_i$ of the 4 major components m_i and the measured total mass M_{meas}:

$$MD = 1 - \frac{\sum m_i}{M_{meas}} \tag{4.1}$$

We know of no previous studies that report a comparison between measured and calculated mass fluxes.

Aluminum or titanium is usually used to estimate lithogenic mineral contributions because the only source of Al and Ti is aeolian and fluvial transport of material from the continental crust, where Al/Ti is on average 15.4 (Taylor and Mclennan, 1995). However, recent studies have shown that a significant fraction of Al in sinking particles exists as "excess" Al, which is Al scavenged from the water column as the particle sinks; excess Al is defined as:

$$Al_{excess} = Al_{total} - (Al/Ti_{litho}) \times Ti_{total} \ . \qquad (4.2)$$

This excess Al can cause measured Al/Ti ratios to be much higher than 15.4; for example, they can be as much as 30 in the equatorial Pacific Ocean (Dymond et al., 1997; Kryc et al., 2003). Excess Al is mostly in the form of $Al(OH)_3$ and Al_2O_3 , and is a major component of particles collected at DYFAMED , where it can contribute as much as ten percent of total mass (Lee et al., submitted). However, even when excess Al hydroxide is included, total mass still does not reach 100% of the measured mass for many samples from the DYFAMED site and from some other open ocean sites.

In this chapter we calculate mass deficits using Eq. (4.1) for sinking particles collected in sediment traps at different open ocean sites (Mediterranean Sea, equatorial Pacific, Arabian Sea, Southern Ocean including Ross Sea, and North Atlantic), and investigate the relationship of mass deficit to measured mass, sampling depth, and chemical composition. Possible causes of mass deficits will then be discussed; we conjecture that water bound to compounds in sinking particles is the most likely cause.

2. Method

2.1. Sample collection

MedFlux time-series samples were collected using indented rotating sphere (IRS) sediment traps at the DYFAMED site (43°25'N, 7°52'E) in the northwestern Mediterranean Sea (http://alpha1.msrc.sunysb.edu/MedFlux/, see also Lee et al., submitted). Samples were collected from early March to early May 2003 (Period 1 or P1, 2003), mid-May to the end of June 2003 (Period 2 or P2, 2003), and early March to the end of April 2005 (Period 1 or P1, 2005) at two depths (see Lee et al., submitted, for details). Each time-series deployment collected material in 11 cups; each cup collected for 4 – 6 days. An aliquot (20% or 30%) from each cup sample was used to determine mass. Other sample splits were used for organic carbon, inorganic carbon, biogenic silica, Al, and Ti measurements. In 2005, only total silica was measured.

Settling velocity (SV) traps were used for the first time in 2003; they were deployed in 2003 at 238 and 117 m, and in 2005 at 313, 524, and 1918 m; time-series IRS sediment traps were deployed simultaneously (Peterson et al., 2005; Lee et al., submitted). A technical description of the settling velocity traps is given in Peterson et al. (2005). Sinking particles were collected and separated into fractions having minimum in situ sinking velocities from 0.68 m/d to 979 m/d. Samples were then analyzed for total mass, OC, IC, biogenic Si, Al and Ti.

In addition to the data from MedFlux, we also analyzed time-series sediment trap data from the US JGOFS Arabian Sea Process Study (ASPS), equatorial Pacific Ocean study (EqPac), North Atlantic Bloom Experiment (NABE), Antarctic Environment and Southern Ocean Process Study (AESOPS), and Ross Sea component of AESOPS. These

last data are available from the US JGOFS website:

http://usjgofs.whoi.edu/mzweb/data/Honjo/sed_traps.html.

2.2. Total mass and chemical composition measurements and calculations

Particles collected by each sediment trap cup were split using a McLane^{TM} WSD splitter and were filtered onto Nuclepore or glass fiber filters (Peterson et al., 2005; Wakeham et al., submitted). Mass was measured with a semi-micro balance (10^{-5} g capability) on a 30% split (in 2003) or 20% split (in 2005) after washing with 5 mL deionized (DI) water and drying overnight at 60°C (Lee et al., submitted). Total carbon was measured using a Carlo Erba model 1106 Elemental Analyzer (in 2003) or model 1602 CNS analyzer (in 2005) (Peterson et al., 2005; Lee et al., submitted). IC was determined by subtracting OC mass from the mass of the non-HCl-treated split. In 2003, samples were sequentially leached with base and acid to separate biogenic and lithogenic silica; Si was then measured by graphite furnace atomic absorption spectrometry (GF-AAS) using a Perkin-Elmer AAnalyst 800 spectrometer. Al and Ti were measured using Inductively Coupled Plasma Mass Spectrometry (ICP-MS) on a ThermoFinnigan Element II ICP-MS. In 2005, total Si, Al, and Ti were measured by GF-AAS using a Perkin-Elmer AAnalyst 800 spectrometer after samples were digested in 750 uL concentrated HCl, 250 uL concentrated HNO_3, and 50 uL concentrated HF (all trace metal clean) for one week and then diluted to 10mL with MilliQ water (Lee et al., submitted).

The major chemical components, OM, opal, $CaCO_3$, lithogenic minerals, and excess Al hydroxides, were calculated based on elemental percentages in both MedFlux and US JGOFS samples. To obtain OM, we multiplied OC by 2.199 (Klaas and Archer, 2002). $CaCO_3$ was calculated stoichiometrically from inorganic carbon (IC), with the assumption that all IC is present as calcium carbonate (Lee et al., submitted). Opal is present as $SiO_2.nH_2O$, where *n* can vary from 0.4 to 1.5 (Mortlock and Froelich, 1989; Collier et al., 2000; Honjo et al., 2000); we used $SiO_2.H_2O$ (Lee et al., submitted) in this study. In 2003 we calculated opal from biogenic Si; in 2005 we calculated opal from {biogenic Si} = {total Si} – {lithogenic Si}, assuming a ratio of 3 between lithogenic Si and lithogenic Al (Klaas and Archer, 2002). Lithogenic mineral is calculated from lithogenic Al ($15.4 \times$ Ti) times 11.9 (Lee et al., submitted). Hydrated Al was calculated from {excess Al} = {total Al} – {lithogenic Al}, multiplied by 3 (= 81/27), where 81 is the molecular weight of hydrated Al in seawater assuming it is 76.2% $Al(OH)_4^-$ and 23.7% $Al(OH)_3$ (Vandenberg et al., 1994; Lee et al., submitted) and 27 is the atomic weight of Al. However, all Al was assumed to be lithogenic in the NABE and AESOPS-noRS data, as Ti data were not available. In the Southern Ocean Al is present only in trace amounts (~0.1%), so the possible presence of excess Al in these samples was ignored.

3. Results

3.1. Mass deficits in different oceans

3.1.1. Mass deficits in the MedFlux study

Ternary plots are often used to depict the relative percentages of three components in a sample that sum to a constant. Ternary plots have been widely used in geosciences, for example, to depict rock classifications according to their compositions (e.g., percentages of quartz, feldspar, and mica) (Philip and Watson, 1988). Since mass deficit *MD* is defined as the difference between 1 and the sum of the OM / M_{meas} and {ballast minerals} / M_{meas} (Eq. 4.1), a ternary diagram can be used to display relationships among these three variables.

In a ternary plot, each side of the triangle represents 0% of one of the components, and the corner of the triangle opposite the base represents 100%. A typical MedFlux example is shown in Fig. 4.1a; the percentages of OM, ballast minerals, and mass deficit in this example are 30%, 23%, and 47%, respectively. Samples with composition percentages outside the range 0 – 100% are located outside the triangle. These samples have negative mass deficit; i.e., the sum of OM and ballast minerals is greater than the measured total mass.

Sinking particle compositions from MedFlux time-series (TS) traps show large ranges in their relative percentages (Fig.1a and Table 4.1): OM / M_{meas} (0.05 – 0.66), {ballast minerals} / M_{meas} (0.25 – 1.72), and mass deficit (-1.05 – 0.56). There is clearly a seasonal trend for the MedFlux samples: most spring (P1) 2003 and spring 2005 sinking particles have similar OM content (0.05 – 0.40), but 2005 samples (0.45 – 1.10) have much higher ballast minerals content than in 2003 (0.25 – 0.73). Much wider range of OM / M_{meas} (0.10 – 0.66) and {ballast minerals} / M_{meas} (0.26 – 1.50) were observed for

summer (P2) 2003 samples. The changing intervals in mass deficit were (0.10 – 0.56) for most P1 2003 samples, (0.20 – 0.37) for most P1 2005 samples, and (-1.05 – 0.39) for P2 2003 samples. During each collection period, significant differences in the major components of particles were found in shallow-water (117 – 313 m) samples, but not in deep-water (771 – 1918 m) samples. All P1 2003 samples from 238 m have positive mass deficits (average 0.34 with a standard deviation of 0.09). In contrast, the mass deficit is quite negative (average -0.42, with standard deviation 0.40) for P2 2003 samples from 313 m.

MedFlux samples from sinking velocity (SV) traps are very similar in OM, ballast mineral, and mass deficit composition to those from concurrently deployed time-series traps (Fig. 4.1b and Table 4.1). The only exception is for SV trap samples collected at 117 m in P2 2003, most of which have positive mass deficits, while time-series traps from the same period have negative mass deficits. Particles collected in SV traps in P1 2005 show compositional changes with depth: shallow-water samples (313 m) have significantly higher OM / M_{meas} (range (0.06 – 0.37) and mean 0.17) and lower {ballast minerals} / M_{meas} (range (0.22 – 0.78) and mean 0.56) than those from deeper (1918 m) water (OM / M_{meas} range (0.04 – 0.18) and mean 0.07; {ballast minerals} / M_{meas} range (0.52 – 1.51) and mean of 0.80) ; deep-water (1918 m) samples show only a slightly lower mass deficit than shallow-water (313 m) samples.

OM / M_{meas}, {ballast minerals} / M_{meas}, and mass deficit were also calculated, using with the same assumptions that were applied to the MedFlux data, for sinking particles collected at US JGOFS sites. Particles from the Ross Sea were all from shallow water (200 – 481 m). Similar to MedFlux samples, they show wide ranges in OM / M_{meas}

(0.10 – 0.90), {ballast minerals} / M_{meas} (0.13 – 0.95), and mass deficit (-0.10 – 0.45) (Fig. 4.1c and Table 4.1). Particles from other sites (ASPS, EqPac, NABE, and AESOPS-noRS (AESOPS with Ross Sea excluded)) are all from deeper waters (> 800 m) and are more uniform in their compositions: OM / M_{meas} (0.05 – 0.30), {ballast minerals} / M_{meas} (0.60 – 1.10), and mass deficit (-0.20 – 0.20) (Fig. 4.1c and Table 4.1).

3.1.2. Mass deficits in both MedFlux and US JGOFS studies

Mass deficits for all samples are shown in Fig. 4.2, along with their sampling depths. Most of the samples in MedFlux and the Ross Sea have positive mass deficits; they were mostly collected in shallow waters (< 1000 m). All other samples from US JGOFS studies were collected in deeper waters (> 800 m), and have mass deficits that are mostly within the range (-0.20 – 0.20).

A box plot is a simple graphic way of depicting a data distribution by plotting 5 summary statistics: the median, the 25% and 75% quartiles, and the maximum and minimum values. Here we define the range between the 25% and 75% quartiles, the middle half of the data, as the *box width*, and the range of this box as the *box range*. Box plots for each study are shown in Fig. 4.3. Samples from the MedFlux study have the highest median mass deficit (0.24), and its box range is positive (0.12 – 0.33). Samples from Ross Sea, which are all at shallow depths (200 – 481 m), also have a higher median for mass deficit (0.08) and a positive box range (0.02 – 0.24). Mass deficits in the other sites (ASPS, EqPac, NABE, and AESOPS-noRS), however, all have medians that are close to 0. Their box widths all include zero or are very close to zero (Fig. 4.3 and Table

4.2). Therefore, the mass deficits in MedFlux and Ross Sea are substantially positive, whereas the mass deficits in other open oceans are not.

3.1.3. Sensitivity study of MD by chosen different parameters for the calculation of major components in both MedFlux and JGOFS

The five major components of total mass were calculated by their corresponding elements using assumed parameter values to convert elements to compounds. However, most of these parameters values are still the subject of discussion. For example, Dymond et al. (1997) used 15.4 as the average of the Al/Ti ratio from a typical crustal source, while Formental et al. (2003) give a higher value (18.5) for average Saharan dust. The Al/Ti ratio was used in the present study to differentiate lithogenic Al from the excess Al, which exists mainly as hydrated Al. Lithogenic minerals have been calculated in previous studies as 11. 9 or 12.15 times lithogenic Al (e.g., Honjo et al., 2000; Lee et al., submitted). For opal, the number of water molecules associated with each biogenic Si ranges from 0.4 to 1.5 (Mortlock and Froelich, 1989; Collier et al., 2000; Honjo et al., 2000). The choice of parameter values might therefore make a substantial difference in the calculation of MD.

For the MedFlux study, box plots were made for MD calculated by choosing six sets of parameters for opal, lithogenics, and excess Al (see legends of Fig. 4.4 for details). Crustal Al/Ti was set at either 15.4 or 18.5; two values (11.9 or 12.15) were used for the estimation of lithogenic minerals; the number of water molecules associated with each Si atom was varied between 0.4 and 1.5 for each mole of biogenic silicon. Although the

medians of MD for these six cases vary from 0.21 to 0.25, all of these medians are still significantly positive and the box position for MD's are all greater than 0. No matter which set of parameters are used for calculations, MD is substantially positive in MedFlux.

The content of Al is much lower in all the JGOFS studies (~1% in ASPS and NABE, ~0.06% in EqPac, and ~0.5% in AESOPS) than at the DYFAMED site (~5%). MD varies within 0.25% of measured total mass for the choice of parameter value (11.9 – 12.15) for lithogenics in these JGOFS studies. The chosen of parameter value for the calculation of lithogenic minerals does not affect MD substantially. Therefore Fig. 4.5 shows only a comparison of MD by changing numbers for the calculation of opal for both MedFlux and JGOFS. When more water is assumed to be associated with opal ($SiO_2.1.5H_2O$), MD is generally lower (by 0.02 – 0.10) in all studies (Fig. 4.5a) than shown in Fig. 4.3 ($SiO_2.H_2O$). If a lower level of hydration is assumed ($SiO_2.0.4H_2O$), MD was higher by 0.01 – 0.07 (Fig. 4.5b) than in Fig. 4.3. The water assigned to opal has its largest effect on MD in AESOPS because opal is the major composition of total mass in most samples collected in AESOPS. The medians for AESOPS-noRS cross zero, from -0.08 in Fig. 4.5a to 0.09 in Fig. 4.5b; the median in the Ross Sea varies between 0.02 and 0.14, but stays positive. In general, MD is substantially positive and higher in both MedFlux and Ross Sea, and it is closer to zero in other JGOFS studies, no matter how much water chosen for the calculation of opal.

3.2. Mass deficits and measured masses

The IRS traps used in MedFlux have 0.0184 m^2 collection areas (Peterson et al., 2005), which is much smaller than the traps used at the other sites (0.5 m^2) (Honjo and Manganini, 1993; Honjo et al., 1995; Honjo et al., 1999; Collier et al., 2000; Honjo et al., 2000; Prahl et al., 2000; Peterson et al., 2005). The quantity of particles collected by each cup in MedFlux (0.4 – 93 mg) for the measurement of total mass is therefore one or two orders of magnitude lower than those from ASPS, EqPac, NABE, and AESOPS, which are usually hundreds to thousands of milligrams. Low sample sizes might have led to greater analytical errors in measurements of both total mass and of the mass of each element in MedFlux. To investigate this possibility, total measured masses for MedFlux were compared to those from JGOFS studies.

Total flux or dry mass was determined by measuring only a fraction of each sample; the other splits were used for the analysis of OM, IC, biogenic Si, Al, and other elements. In MedFlux, 20% or 30% of the sample was used for the mass measurement. For the JGOFS samples, mass splits were 30% (ASPS, EqPac, NABE, and AESOPS-noRS) and 20% (Ross Sea) (Honjo and Manganini, 1993; Honjo et al., 1995; Honjo et al., 1999; Collier et al., 2000; Prahl et al., 2000). As shown in Fig. 4.6, measured masses could be high as thousand of milligrams in waters shallower than 2000 m, or could be hundreds of milligrams in waters deeper than 2000 m. In contrast, measured masses in MedFlux study were all less than 50 mg. Mass deficits were usually seen in smaller size samples (Fig. 4.7a); as measured total masses increase, the range of mass deficits also become smaller. This relationship also holds for each study separately (Fig. 4.7b-4.7e): as measured masses increase, mass deficits approach zero, except in MedFlux and Ross Sea samples, where mass deficits are slightly positive (details are given below).

Analytical errors can be estimated from measurement errors for each element. In MedFlux for example, the analytical error for C is ±2%, for biogenic Si is high as 50%, and for Al is about 20%. These analytical errors were *propagated* to the calculation of MD and are shown in Fig. 4.8 by each MedFlux samples, for which the average error is ±0.08. With the consideration of error bars for each sample, the MD in MedFlux is still substantially positive.

Overall, analytical error in the measurement of mass or of other chemical components may be one of the important factors leading to the relatively larger variance in mass deficits in MedFlux and Ross Sea samples, especially for the smallest samples. However, the mass deficits in MedFlux and Ross Sea samples are systematic, suggesting that an explanation other than measurement error is needed.

3.3. Mass deficit and the sample collection depths

To investigate the possible effect of depth on mass deficit, data from all studies were grouped into 6 based on their collection depth: 0-750 m, 750-1500 m, 1500-2500 m, 2500-3400 m, 3400-4000 m, and >4000 m. These figures were chosen because their endpoints correspond to gaps in the data; see Fig. 4.2. As mentioned earlier, most of MedFlux and all Ross Sea samples were collected in shallow sediment traps and fell into the 0-750 m group, all the other US JGOFS samples were collected in deeper waters (≥ 800 m). Box plots of both mass deficits and measured masses are shown for each of the 6 depth groups (Figs. 4.9a & 4.9b). Samples collected from 0-750 m, which are all MedFlux and Ross Sea particles, have the largest median (0.17) and the most positive

box width (0.04 – 0.30) of mass deficit; samples collected in the other 5 groups (> 750 m) have much lower medians (0.02, -0.01, 0, -0.01 and -0.02) and box widths (all within (-0.05 – 0.05)) for their mass deficit. A two-sample t test shows that the mass deficits in the first group are significantly (0.05% level) higher than any other 5 groups. No significant relationship between mass deficit and sample collection depth was found in deeper waters (> 750 m). Meanwhile, the 0-750 m group has the smallest measured masses (10.2 mg for the median and (3.9 – 31.6) mg for the box width); the other 5 groups (> 750 m) have much higher measured masses (e.g., mean 136 mg and (54 – 290) mg box width for group 750-1500 m)

Note that most of the shallow-water samples (e.g., MedFlux samples) also have lower measured masses; it is therefore necessary to determine whether the high mass deficits in 0-750 m group were caused by low measured masses or by a depth effect. As shown in Fig. 4.7c, most of the mass deficits in the Ross Sea (all collected between 200 and 500 m) were positive, even when the measured masses were as high as thousands of milligrams, where the effect of low mass should be small. The AESOPS-noRS study, in which all samples were collected deeper than 980 m, had very similar ranges for measured masses to those samples in Ross Sea. There should also be no "study effect" caused by differences in measurement protocols for total mass and for the other chemical components, since the Ross Sea study was part of AESOPS. The mass deficits in AESOPS-noRS were, however, mostly within (-0.20 – 0.20), and had both median and mean close to 0. These observations provide evidence that depth does have an effect on samples from shallow water (less than 1000 m).

3.4. Sensitivity study of the parameters for calculating five major components on mass deficit

As mentioned earlier, calculated mass is obtained by summing the masses of the four major chemical components, or five if excess Al is included, based on their elemental concentrations, using certain constant parameters for the ratio of chemical component to its corresponding measurement. For example, the 2.199 for OM and OC (Table 4.3) is based the ratio used by Klaas and Archer (2002). Inaccuracy in these parameters may therefore cause calculated mass values to depart from measured mass. Assuming no mass deficit for all studies, we applied multiple regression analysis to test the reasonableness of these parameter values. Samples that have mass deficits in their box width (see Table 4.3 and Fig. 4.2) were chosen as the most representative samples for each site, and were used in the multiple regression tests, so that the random errors caused by low sample sizes were mostly avoided. Results are shown in Table 4.3. The numbers of samples that have mass deficits in box width, which were used in the multiple regressions, are also listed.

Literature parameters or their ranges are listed in Table 4.3. Parameters with their 95% confidence intervals that were estimated using multiple linear regressions are also listed in Table 4.3. Estimated parameter values that are beyond ±20% of literature parameters or their averages are shown in bold in this table. They include parameters for OC in MedFlux, LithoAl (lithogenic Al) and XsAl (excess Al) in MedFlux, Ross Sea, and EqPac, and XsAl in ASPS. No parameter is assigned to excess Al in NABE and AESOPS-noRS because all Al was assumed to be lithogenic in these two studies. Since

Al is very low in samples from the Ross Sea (~ 0.03%) and EqPac (~ 0.06%), and there is almost no excess Al in ASPS, the 5 black boldface parameters without highlights would not be expected to have a large effect on the calculation of total mass. The 3 boldface parameters (all from MedFlux) highlighted with gray in Table 4.3 are all significantly larger than literature values. Since OM is a very important contributor to total mass especially in shallow waters (~ 20% in MedFlux and ~ 30% in the Ross Sea; see Table 4.1), and the Al concentration in MedFlux is 5 – 135 times higher than at the other JGOFS study sites (Chou and Wollast, 1997), the choice of literature values for these three parameters might have significantly underestimated calculated mass, which in turn might have caused the large mass deficits at MedFlux and the Ross Sea. The higher the OM and Al contents in MedFlux and Ross Sea, the lower the calculated total masses and the greater the mass deficits are for these two sites. Reasons for the departure of the multiple regression results from literature values will be discussed below.

To investigate further the effects of each elemental composition on mass deficit, the dependencies of mass deficit on OC, IC, biogenic Si, lithogenic Al, and excess Al, are shown in Fig. 4.10. MedFlux and Ross Sea data are separated from the other JGOFS data because they are different not only in the total measured mass, but also in the depths at which the sediment traps were deployed. Most samples of TS 117m in MedFlux 2003 P2 were outliers from the rest of MedFlux study, and were excluded in these regressions. Four relationships, OC and MD, BioSi and MD in the MedFlux and Ross Sea data, IC and MD in all studies, are significant (see their r's in Fig. 4.10).

4. Discussion

Mass deficits of sinking particles collected from different open oceans are mostly in the range (-0.20 – 0.20), but are especially large for samples collected at the DYFAMED site in the Mediterranean Sea, in the range of (-1.04 – 0.57) (Fig. 4.2 and Table 4.1). The sample treatment procedure might be one factor leading to the large observed mass deficits. To determine the mass of sinking particles, samples collected by sediment trap are pretreated by removing the salt and drying the samples in an oven (Honjo and Manganini, 1993; Lee et al., submitted). Rinsing the samples with DI water removes salts absorbed onto the filter and particle surfaces. Any residual salt would cause an overestimate of the mass of the samples. We used 5 ml of DI water, which should be enough to thoroughly rinse the salt from the 0.4 um Nuclepore filters used in MedFlux. In general, with enough care, the residual salts on the particles and the filter surface are negligible. The other factor is that the analytical error for total and elemental masses might be large for low-mass samples, as suggested by the relationship between measured masses and mass deficits (Fig. 4.7), especially considering that the samples must be split into subsamples. For example, large variations in mass deficit were found in MedFlux, Ross Sea, and AESOPS-noRS when total measured mass was less than 50 mg. When the measured mass increases to 200 mg, all the mass deficits decreased to within ±0.20. Mass deficit is therefore at least partially random because of relatively large analytical errors for samples with low mass.

For samples with high mass in which analytical error cannot explain the mass deficit, there must be some other factors. Our multiple regression results (Fig. 4.10) indicate that organic matter is directly related to the mass deficit: the higher the content

of organic matter, the higher the mass deficit. One possible factor that could lead to an overestimate of mass is water trapped or bound within the particles. Tightly-bound water molecules might not evaporate at the oven temperatures we used, leading to higher measured mass values than calculated mass. Thus the water trapped in the dried sample might explain the mass deficit. Bound water may be tightly bonded to organic matter through hydrogen bonds, which may be strong enough to withstand the drying conditions (Sutton and Sposito, 2005; Liu and Lee, 2006). Another factor to be considered is that organic matter in the surface ocean may be more hydrated than matter from depth. For example, transparent exopolymer particles (TEP), a type of highly hydrated gel, are abundant in sinking particles in the surface ocean, and it can be very difficult to totally dehydrate TEP during the drying procedure (Passow, 2002). In addition to organic matter, hydrated inorganic minerals such as $Al(OH)_3$, SiO_2, $MgCO_3$ and other carbonate-bearing minerals may also trap or contain extra water (Adams et al., 1991; Vandenberg et al., 1994; Gaffey, 1995; Ruiz et al., 1997; Railsback, 1999).. For example, each $Al(OH)_3$ may be associated with two crystallized water molecules in complexes.

Based on Redfield's CNP ratios, the representative molecular formula of sinking organic material is $(CH_2O)_{106}(NH_3)_{16}(H_3PO_4)$. The water and oxygen of this formula was once thought to be seriously overestimated due to the combining of nominal structural units (CH_2O and NH_3) into polysaccharides and proteins during dehydration (Hedges et al., 2002). While the water molecules in this structure may be overestimated, a small amount of bound water molecules may still exist within the dried particles. Hedges et al. (2002) also pointed out that it is difficult to accurately measure the elemental percentage for organic hydrogen, oxygen, and sulfur. Therefore, the drying process during sample

pretreatment might not expel all the bound water molecules, leading to higher measured mass than calculated mass. Alternatively, some smectite water might be lost from $CaCO_3$ or lithogenic minerals by over-drying the sample, which would lead to lower measured mass than calculated mass. In our calculations, no H_2O is assigned to organic matter, $CaCO_3$, and lithogenic minerals, and only one H_2O is assigned to each biogenic Si; whether there is additional H_2O in these components is not known.

In addition, the multiple linear regression results (Table 4.3) show that the parameters used for OC and Al (both lithogenic Al and excess Al) used in calculating total mass may have been underestimated in the MedFlux study. However, these same results may be interpreted not as showing that the parameter values were wrong, but instead that some water was trapped in either OM or lithogenic materials. Since both OC and Al are major contributors to total mass in MedFlux samples, water bound with either OC or Al might significantly affect the mass. As shown by Fig. 4.1c, OC is also a dominant component in particles from the shallower waters of the Ross Sea (Fig. 4.1c). Therefore, the high concentration of OM, with trapped or bound water, is a possible explanation for positive mass deficit in shallow waters.

The statistical results show that the mass deficit is significantly negatively correlated with IC for MedFlux and Ross Sea samples ($r=0.69$; $n=222$) (the 2003 P2 TS samples from 117m have not been included in the regression as these samples were outliers from the rest of MedFlux samples) and for the rest of the JGOFS data ($r=0.34$, $n=898$). However, no significant correlation between mass deficit and other major components was found for these data (Fig. 4.10). So it appears that mass deficit is large with high amounts of OM, opal, and lithogenic minerals, or with low amount of $CaCO_3$.

This suggests that OM, opal or lithogenic minerals are the main water-carrying reagent rather than $CaCO_3$ according to our hypothesis about bound water in the sinking particles.

The averaged mass deficits are positive for particles of MedFlux and Ross Sea, while they are near zero for particles from the other sites (Fig. 4.3). Both MedFlux and Ross Sea particles are from shallow waters, and their OM percentages vary substantially. The calculated total mass is therefore, more consistent with its measured mass for sinking particles collected in deeper oceans using the current calculation method. This consistency for deep ocean samples suggests that our protocol for mass calculation is appropriate, especially the coefficients for calculating different components. It is not clear though how much water is trapped into either OM or lithogenic materials at shallow depths. A rough estimate of the water content in OM, lithogenic mineral, and excess Al in MedFlux study can be made by assuming that the departure from literature values of all the multiple regression results were caused by H_2O bound to each component. For example, the H_2O bounded into OM in MedFlux is estimated by:

$$\frac{3.1 - 2.199}{3.1} \times 100\% = 29\%$$

where 3.1 is the multiple regression result for the ratio between OM and OC in MedFlux in Table 4.3, and 2.199 is the literature parameter for it. In this way, we estimate that the H_2O bounded into OM, lithogenic minerals and excess Al during the measurement of total mass in MedFlux study are approximately 29%, 27%, and 23% by weight, respectively.

We know that biogeochemical properties of sinking particles change dramatically in the twilight zone (Wakeham et al., 1997). For example, sinking particles from the surface ocean are fresh and are made of large biopolymers, in which water can be

encapsulated inside the polymer networks. However, these particles maybe degraded extensively, becoming more amorphous (Liu et al., unpublished results), so that water can be more easily dissipated during the drying. Inorganic minerals might also change with depth. For example, biominerals like $CaCO_3$ and opal tend to dissolved with depth. How the change of minerals in the sinking particles affect particle affinity for water is not yet clear.

5. Conclusions and implications

The mass of sinking particles calculated by summing up the major chemical components (OM, opal, $CaCO_3$, and lithogenic minerals) is generally different from the measured mass, by about -0.20 – 0.20 in open ocean samples. This mass deficit is systematically positive for particles mostly collected from shallow water (<800 m) in the Mediterranean Sea and Ross Sea. However, mass deficits at the other sites, at depths deeper than 800 m, are close to zero. We hypothesize that water bound in OM, which is an important content in particles of shallow water, may partly explain mass deficits. Ballast minerals may also be important in trapping water in surface particles. Analytical errors on samples with lower mass can only contribute to mass deficit in a random way.

Our results here also address a concern about the definition of mass. Operationally, mass is obtained when the sample weight reaches a constant value in an oven at 60 °C. How much associated water remains on the particle are unknown, and may depend on the particle types or composition. Unfortunately, it is very difficult to quantify the water left in the particle samples after being oven-dried. How to define the mass,

whether the bound water should be included, and the mechanisms leading to the mass deficit, are questions for future investigations.

References

Adams, S. J., Hawkes, G. E., and Curzon, E. H., 1991. A solid-state Si-29 nuclear-magnetic-resonance study of opal and other hydrous silicas. *American Mineralogist* **76**, 1863-1871.

Chou, L. and Wollast, R., 1997. Biogeochemical behavior and mass balance of dissolved aluminum in the western Mediterranean Sea. *Deep-Sea Research II* **44**, 741-768.

Collier, R., Dymond, J., Honjo, S., Manganini, S., Francois, R., and Dunbar, R., 2000. The vertical flux of biogenic and lithogenic material in the Ross Sea: Moored sediment trap observations 1996-1998. *Deep-Sea Research II* **47**, 3491-3520.

Dymond, J., Collier, R., McManus, J., Honjo, S., and Manganini, S., 1997. Can the aluminum and titanium contents of ocean sediments be used to determine the paleoproductivity of the oceans? *Paleoceanography* **12**, 586-593.

Formenti, P., Elbert, W., Maenhaut, W., Haywood, J., and Andreae, M. O., 2003. Chemical composition of mineral dust aerosol during the Saharan Dust Experiment (SHADE) airborne campaign in the Cape Verde region. *Journal of Geophysical Research* **108**, No. D18, 8576, doi:10.1029//2002JD002648.

Gaffey, S. J., 1995. H_2O and OH in echinoid calcite - a spectroscopic study. *American Mineralogist* **80**, 947-959.

Haake, B., Ittekkot, V., Rixen, T., Ramaswamy, V., Nair, R. R., and Curry, W., 1993. Seasonality and interannual variability of particle fluxes to the deep Arabian Sea. *Deep-Sea Research I* **40**, 1323-1344.

Hedges, J. I., Baldock, J. A., Gelinas, Y., Lee, C., Peterson, M. L., and Wakeham, S. G., 2002. The biochemical and elemental compositions of marine plankton: A NMR perspective. *Marine Chemistry* **78**, 47-63.

Honjo, S., Dymond, J., Collier, R., and Manganini, S. J., 1995. Export production of particles to the interior of the equatorial Pacific-ocean during the 1992 EqPac experiment. *Deep-Sea Research II* **42**, 831-870.

Honjo, S., Dymond, J., Prell, W., and Ittekkot, V., 1999. Monsoon-controlled export fluxes to the interior of the Arabian Sea. *Deep-Sea Research II* **46**, 1859-1902.

Honjo, S., Francois, R., Manganini, S., Dymond, J., and Collier, R., 2000. Particle fluxes to the interior of the Southern Ocean in the Western Pacific sector along 170 degrees W. *Deep-Sea Research II* **47**, 3521-3548.

Honjo, S. and Manganini, S. J., 1993. Annual biogenic particle fluxes to the interior of the north-Atlantic ocean - studied at 34-degrees-N 21-degrees-W and 48-degrees-N 21-degrees-w. *Deep-Sea Research II* **40**, 587-607.

Klaas, C. and Archer, D. E., 2002. Association of sinking organic matter with various types of mineral ballast in the deep sea: Implications for the rain ratio. *Global Biogeochemical Cycle* **16**, 1-16.

Kryc, K. A., Murray, R. W., and Murray, D. W., 2003. Al-to-oxide and Ti-to-organic linkages in biogenic sediment: relationships to paleo-export production and bulk Al/Ti. *Earth & Planetary Science Letters* **211**, 125-141.

Lee, C. and Cronin, C., 1982. The vertical flux of particulate organic nitrogen in the sea - decomposition of amino-acids in the Peru upwelling area and the equatorial Atlantic. *Journal of Marine Research* **40**, 227-251.

Lee, C., Peterson, M. L., Wakeham, S. G., Armstrong, R. A., Cochran, J. K., Miquel, J. C., Fowler, S., Hirschberg, D., Beck, A., and Xue, J., submitted. Particulate organic matter and ballast fluxes measured using Time-Seriees and Settling Velocity sediment traps in the northwestern Mediterranean Sea. *Deep-Sea Research II*

Liu, Z. F. and Lee, C., 2006. Drying effects on sorption capacity of coastal sediment: The importance of architecture and polarity of organic matter. *Geochimica et Cosmochimica Acta* **70**, 3313-3324.

Mortlock, R. A. and Froelich, P. N., 1989. A simple method for the rapid-determination of biogenic opal in pelagic marine-sediments. *Deep-Sea Research* **36**, 1415-1426.

Passow, U., 2002. Transparent exopolymer particles (TEP) in aquatic environments. *Progress in Oceanography* **55**, 287-333.

Peterson, M. L., Wakeham, S. G., Lee, C., Askea, M. A., and Miquel, J. C., 2005. Novel techniques for collection of sinking particles in the ocean and determining their settling rates. *Limnology & Oceanography-Methods* **3**, 520-532.

Philip, G. and Watson, D., 1988. Determining the representative composition of a set of sandstone samples. *Geological Magazine* **125**, 267-272.

Prahl, F. G., Dymond, J., and Sparrow, M. A., 2000. Annual biomarker record for export production in the central Arabian Sea. *Deep-Sea Research II* **47**, 1581-1604.

Railsback, L. B., 1999. Patterns in the compositions, properties, and geochemistry of carbonate minerals. *Carbonate Evaporite* **14**, 1-20.

Ruiz, J. M., McAdon, M. H., and Garces, J. M., 1997. Aluminum complexes as models for Broensted acid sites in zeolites: Structure and energetics of [Al(OH)(4)](-), [Al(H2O)(6)](3+), and intermediate monomeric species [Al(OH)(x)(H2O)(n-x)center dot mH(2)O](3-x) obtained by hydrolysis. *Journal of Physical Chemistry B* **101**, 1733-1744.

Sarmiento, J. L. and Gruber, N., 2006. *Ocean biogeochemical dynamics*. Princeton University Press, Princeton, N.J.

Sutton, R. and Sposito, G., 2005. Molecular structure in soil humic substances: The new view. *Environmental Science & Technology* **39**, 9009-9015.

Takahashi, K., Fujitani, N., Yanada, M., and Maita, Y., 2000. Long-term biogenic particle fluxes in the Bering Sea and the central subarctic Pacific Ocean, 1990-1995. *Deep-Sea Research I* **47**, 1723-1759.

Taylor, S. R. and Mclennan, S. M., 1995. The geochemical evolution of the continental-crust. *Reviews of Geophysics* **33**, 241-265.

Vandenberg, C. M. G., Boussemart, M., Yokoi, K., Prartono, T., and Campos, M. L. A. M., 1994. Speciation of aluminum, chromium and titanium in the NW Mediterranean. *Marine Chemistry* **45**, 267-282.

Wakeham, S. G., Lee, C., Farrington, J. W., and Gagosian, R. B., 1984. Biogeochemistry of particulate organic-matter in the oceans - results from sediment trap experiments. *Deep-Sea Research* **31**, 509-528.

Wakeham, S. G., Lee, C., Hedges, J. I., Hernes, P. J., and Peterson, M. L., 1997. Molecular indicators of diagenetic status in marine organic matter. *Geochimica et Cosmochimica Acta* **61**, 5363-5369.

Wakeham, S. G., Lee, C., Peterson, M. L., Liu, Z., Szlosek, J., Putnam, I., and Xue, J. Organic compound composition and fluxes in the Twilight Zone - time series and settling velocity sediment traps during MEDFLUX. *Deep-Sea Research II* (submitted)

	time or site	depth (m)	n	OM mean	CaCO3 mean	opal mean	lithogenic minerals mean	Al hydroxide mean	mass deficit mean	mass deficit median	mass deficit range
MedFlux	2003-P1	TS 238	11	0.22	0.18	0.14	0.06	0.05	0.34	0.36	0.16-0.46
		TS 771	11	0.2	0.27	0.1	0.05	0.04	0.34	0.29	0.11-0.57
		SV1 238	8	0.21	0.3	0.09	0.08	0.07	0.26	0.27	0.11-0.38
		SV2 238	8	0.18	0.22	0.17	0.03	0.03	0.37	0.33	0.28-0.54
	2003-P2	TS 117	9	0.5	0.73	0.11	0.06	0.02	-0.42	-0.47	-1.04-0.06
		TS 1918	10	0.2	0.27	0.06	0.1	0.07	0.3	0.34	0.16-0.40
		SV1/2 117	11	0.55	0.34	0.02	0.04	0.01	0.04	0.27	-1.49-0.59
	2005	TS 313	11	0.12	0.31	0.21	0.24	0.14	0.02	0.11	-0.95-0.19
		TS 924	11	0.1	0.24	0.09	0.26	0.13	0.2	0.22	-0.18-0.39
		SV1 313	10	0.22	0.08	0.05	0.22	0.1	0.33	0.34	0.08-0.44
		SV2 313	11	0.12	0.15	0.2	0.17	0.13	0.22	0.2	0.09-0.44
		SV1 524	11	0.09	0.14	0.2	0.24	0.17	0.17	0.16	0.08-0.27
		SV2 524	11	0.09	0.17	0.07	0.28	0.13	0.26	0.28	0.09-0.36
		SV1 1918	11	0.07	0.23	0.12	0.24	0.13	0.21	0.19	0.03-0.40
		SV2 1918	11	0.07	0.36	0.11	0.29	0.12	0.05	0.09	-0.69-0.36
Ross Sea	M6	200	13	0.29	0.38	0.27	0.0001	0	0.06	0.01	-0.02-0.45
		460	11	0.14	0.33	0.46	0.01	0	0.06	0.04	-0.08-0.28
	M7a	465	21	0.31	0.16	0.33	0.004	0	0.2	0.2	-0.01-0.41
	M7b	206	14	0.44	0.22	0.31	0.005	0	0.03	0.03	-0.10-0.24
		481	17	0.25	0.18	0.35	0.01	0	0.21	0.15	0.01-0.44
ASPS	2NE	903	21	0.22	0.5	0.19	0.11	0	-0.01	0	-0.15-0.06
		1974	21	0.18	0.46	0.24	0.12	0	0.01	0.01	-0.15-0.15
		3141	21	0.14	0.52	0.26	0.13	0	-0.05	-0.05	-0.12-0.01
	3NE	858	12	0.21	0.51	0.15	0.08	0.0006	0.07	0.07	0.01-0.11
		1857	20	0.17	0.48	0.21	0.09	0.0005	0.05	0.05	0.02-0.08
		2871	18	0.14	0.49	0.22	0.11	0.0001	0.05	0.05	0.01-0.08
	4NE	821	21	0.2	0.52	0.15	0.05	0	0.09	0.07	0.02-0.24
		2229	21	0.14	0.5	0.25	0.07	0	0.04	0.03	0.01-0.08
		3478	21	0.12	0.49	0.25	0.09	0	0.04	0.03	0.01-0.11
	2SW	924	12	0.15	0.52	0.19	0.15	0	-0.01	0	-0.03-0.02
		1996	12	0.13	0.53	0.18	0.16	0	0	0	-0.04-0.03

		3159	12	0.11	0.49	0.2	0.17	0	0.04	-0.02	-0.27-0.77
	3SW	888	13	0.16	0.54	0.21	0.11	0	-0.01	-0.01	-0.06-0.03
		1882	18	0.15	0.5	0.26	0.12	0	-0.03	-0.03	-0.10-0.01
		2991	19	0.13	0.5	0.22	0.13	0	0.02	0.01	-0.02-0.09
	4SW	807	9	0.18	0.54	0.15	0.1	0	0.04	0.04	0.00-0.08
		2215	20	0.14	0.55	0.2	0.09	0	0.02	0.01	-0.02-0.09
		3489	21	0.12	0.54	0.2	0.1	0	0.03	0.02	-0.01-0.13
	5	800	9	0.18	0.63	0.13	0.07	0.0001	0	0	-0.02-0.01
		2363	21	0.13	0.62	0.14	0.11	0	0	-0.01	-0.03-0.14
		3915	21	0.13	0.59	0.15	0.14	0	-0.01	-0.01	-0.04-0.03
EqPac	12S	1292	21	0.13	0.73	0.11	0.003	0	0.03	0.04	-0.14-0.10
		3594	21	0.08	0.78	0.17	0.008	0	-0.03	-0.02	-0.17-0.06
	5S	2099	21	0.1	0.67	0.24	0.003	0	-0.01	-0.01	-0.06-0.05
		2209	21	0.1	0.69	0.3	0.003	0.0002	-0.09	-0.09	-0.19-0.09
		2316	21	0.1	0.66	0.25	0.003	0.0002	0	0.01	-0.39-0.42
	Equator	880	21	0.12	0.71	0.2	0.002	0.001	-0.04	-0.06	-0.19-0.18
		2284	21	0.1	0.74	0.28	0.004	0.001	-0.12	-0.04	-0.90-0.05
		3618	21	0.1	0.67	0.32	0.003	0.001	-0.09	-0.08	-0.17-0.01
	5N	1200	21	0.19	0.59	0.24	0.006	0	-0.03	-0.03	-0.07-0.04
		2100	21	0.13	0.66	0.24	0.005	0.0003	-0.04	-0.04	-0.06-0.02
		3800	14	0.13	0.65	0.26	0.008	0	-0.05	-0.05	-0.06-0.04
	9N	2150	18	0.15	0.62	0.27	0.02	0	-0.06	-0.05	-0.13-0.01
		2250	20	0.15	0.59	0.29	0.01	0.0003	-0.04	-0.05	-0.06-0.01
NABE	34N	1070	3	0.27	0.48	0.22	0.2	-	-0.16	-0.14	-0.28- -0.05
		1248	12	0.1	0.7	0.1	0.08	-	0.01	0.01	-0.05-0.09
		2067	13	0.1	0.64	0.15	0.15	-	-0.04	-0.04	-0.16-0.03
		1894	13	0.1	0.62	0.11	0.1	-	0.08	0.05	-0.02-0.24
		4565	10	0.08	0.63	0.14	0.19	-	-0.03	-0.03	-0.10-0.10
		4391	13	0.09	0.6	0.11	0.15	-	0.05	0.01	-0.05-0.27
	48N	1018	12	0.21	0.54	0.22	0.04	-	-0.02	-0.02	-0.12-0.04
		1202	13	0.14	0.59	0.19	0.05	-	0.03	0.02	-0.01-0.08
		2018	13	0.11	0.57	0.28	0.05	-	-0.01	0	-0.11-0.05
		2200	3	0.13	0.59	0.2	0.06	-	0.01	0.02	0.01-0.02

		3718	11	0.09	0.6	0.28	0.07	-	-0.03	-0.03	-0.09-0.02
		3749	13	0.08	0.61	0.21	0.1	-	0	-0.01	-0.03-0.09
AESOPS	M1	986	18	0.13	0.71	0.1	0.03	-	0.03	0.02	-0.03-0.15
-noRS		1981	3	0.30	0.47	0.14	0.03	-	0.06	-0.03	-0.07-0.28
	M2	982	16	0.14	0.42	0.44	0.01	-	-0.01	0	-0.10-0.05
		4224	19	0.07	0.46	0.45	0.04	-	-0.02	-0.02	-0.06-0.01
	M3	1003	13	0.11	0.24	0.01	0.003	-	0.03	0.03	-0.08-0.21
		1103	8	0.06	0.26	0.68	0.004	-	0	-0.01	-0.12-0.16
	M4	1031	13	0.08	0.10	0.82	0.001	-	0	0.02	-0.20-0.05
		2026	9	0.05	0.11	0.81	0.002	-	0.02	0.03	-0.03-0.06
	M5	937	13	0.18	0.04	0.76	0.001	-	0.02	0.01	-0.16-0.18
		1033	11	0.18	0.08	0.71	0.002	-	0.02	0.02	-0.12-0.13
		1842	11	0.15	0.09	0.79	0.01	-	-0.04	0.01	-0.35-0.08
		2311	10	0.16	0.16	0.67	0.01	-	0.01	0.02	-0.05-0.08

Table 4.1 Averages of OM, $CaCO_3$, opal, lithogenic minerals, and excess Al hydroxide at different open ocean sites, with the medians, means, and ranges for mass deficit. Mass deficit is calculated using Eq. 4.2 for both MedFlux and US JGOFS samples.

		mass deficit value				
	number of samples	median	25%	75%	min	max
MedFlux	155	0.241	0.115	0.328	-1.488	0.590
Ross Sea	76	0.075	0.019	0.235	-0.102	0.446
ASPS	363	0.011	-0.013	0.043	-0.267	0.773
EqPac	262	-0.036	-0.064	-0.011	-0.901	0.416
NABE	129	-0.006	-0.031	0.024	-0.281	0.274
AESOPS-noRS	144	0.007	-0.025	0.037	-0.346	0.245

Table 4.2 Box plot, including the medians, 25% and 75% quantiles, minima and maxima,of mass deficits for each of the 6 studies (MedFlux, Ross Sea, ASPS, EqPac, NABE, and AESOPS-noRS).

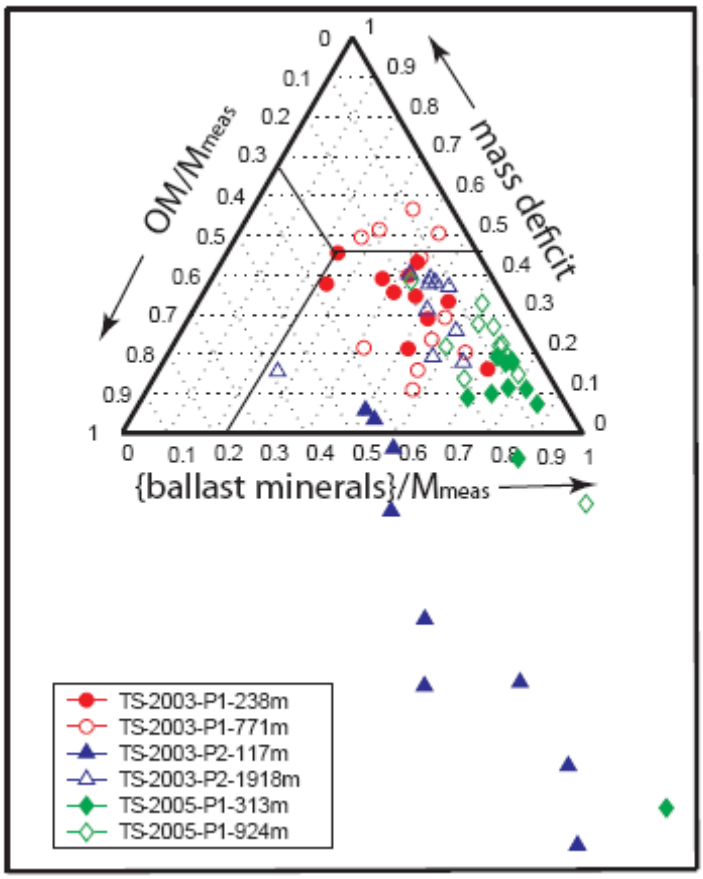
literature parameters	n*	r-square*	OC 2.199	IC 8.33	BioSi 2.4-3.22	LithoAl 11.9-12.15	XsAl 3
MedFlux	78*	0.82	**3.1□0.3**	9.9□1.4	3.3□0.6	**16.4□2.4**	**3.9□1.4**
Ross Sea	39*	0.85	2.5□□	8.6□0.7	3.0□0.2	**0.2□130**	**214□645**
ASPS	182*	0.89	2.4□0.1	8.3□0.1	2.9□0.1	11.3□0.6	**88□55**
EqPac	132*	0.93	2.3□0.1	8.2□0.1	2.5□0.1	**4.2□5.8**	**1.1□21**
NABE	66*	0.89	2.3□0.2	8.5□0.2	2.7□0.1	10.2□0.9	-
AESOPS-noRS	73*	0.92	2.2□0.1	8.4□0.1	2.8□0.03	10.1□3.1	-

* n is the number of samples between 25% and 75% quartiles that were used for multiple regression

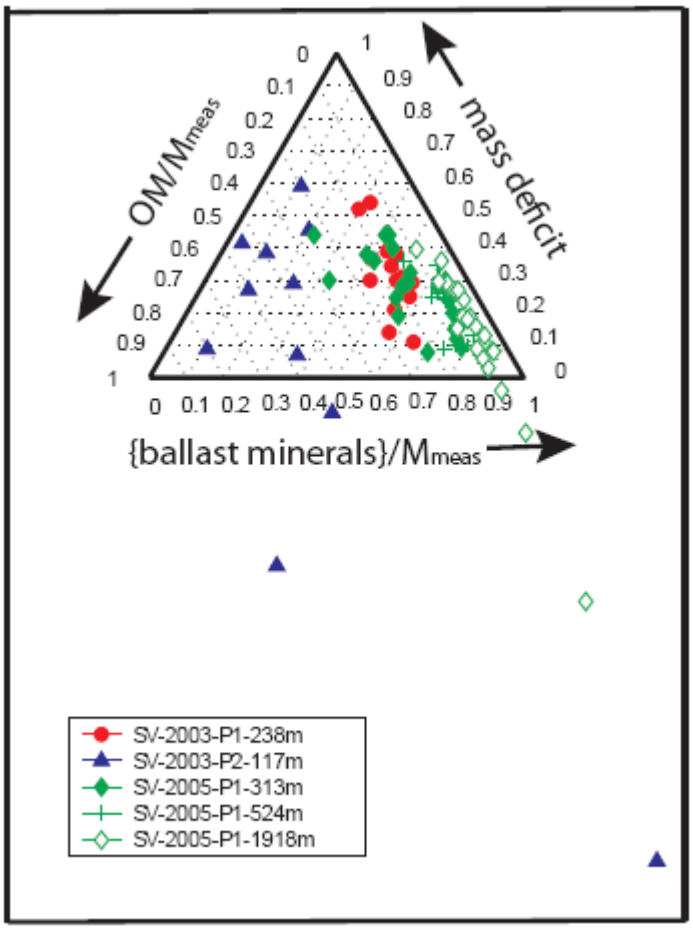
* r-square is based on the estimated OC with the observed OC, assuming IC, BioSi, LithoAl, and XsAl are the independent variables

Table 4.3 Samples that have mass deficit in their box width (the 25% to 75% intervals) were used for multiple regressions to obtain the parameters of major compositions. The parameters or ranges listed on the first row are from the literature. Parameter estimates (with their 95% confidence intervals) from the multiple linear regressions are also listed in Table 4.3. The r-square is given based on the estimated OC with the observed OC assuming IC, biogenic Si, lithogenic Al, and excess Al are the independent variables in the linear regression. Those estimated parameters that are beyond ±20% of literature parameters or their averages are shown in bold in this table. No parameter value is assigned to excess Al in NABE and AESOPS-noRS because all Al was assumed to be lithogenic. Since Al is very low in samples from the Ross Sea (~ 0.03%) and EqPac (~ 0.06%), and there is almost no excess Al in ASPS, the 5 black boldface parameters without highlight would not be expected to have a large effect on the calculation of total mass. The 3 parameters highlighted in gray, all from MedFlux, are all substantially larger than values from the literature.

(a)


OM/Mmeas
mass deficit
{ballast minerals}/Mmeas
TS-2003-P1-238m
TS-2003-P1-771m
TS-2003-P2-117m
TS-2003-P2-1918m
TS-2005-P1-313m
TS-2005-P1-924m

(b)


OM/Mmeas
mass deficit
{ballast minerals}/Mmeas
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
SV-2003-P1-238m
SV-2003-P2-117m
SV-2005-P1-313m
SV-2005-P1-524m
SV-2005-P1-1918m

(c)

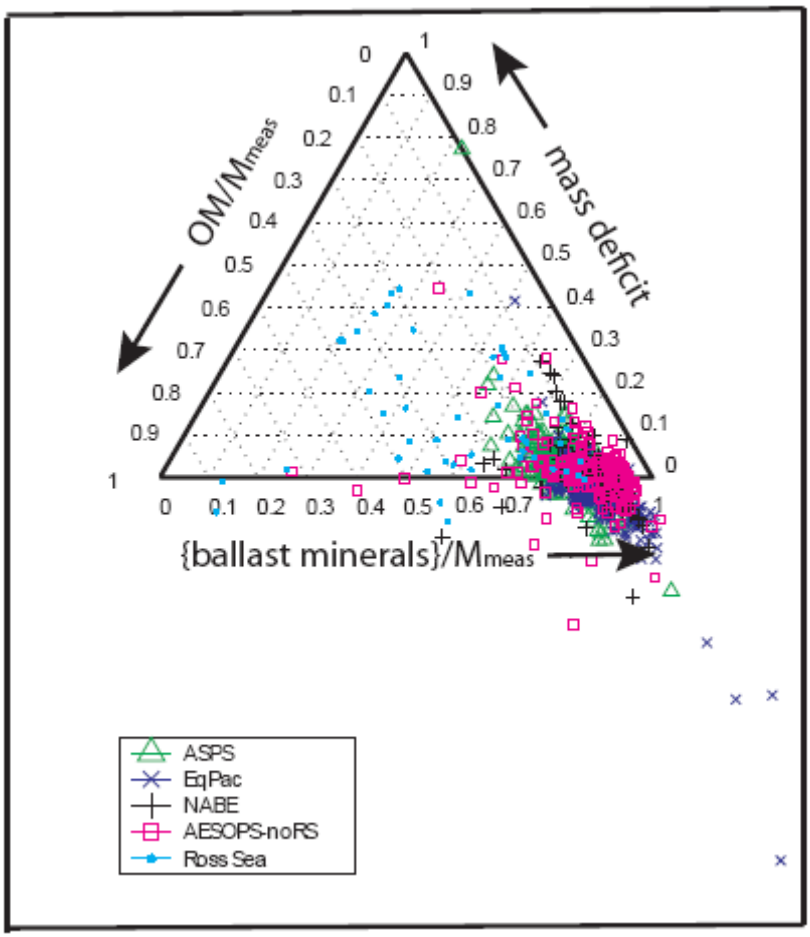


Figure 4.1 Ternary plot of particles composition based on organic matter (OM/M_{meas}; see Eq. 4.1 in text), ballast minerals (sum of opal, $CaCO_3$, lithogenic minerals, and excess Al hydrate, all divided by M_{meas}) and mass deficit for (a) MedFlux time-series sediment trap samples; (b) MedFlux settling velocity mode sediment trap samples; and (c) samples collected by time-series sediment traps from different JGOFS studies (ASPS, EqPac, NABE, AESOPS-noRS, and the Ross Sea). Those samples with compositions relative percentage out of the range 0~100% are located out of the triangle. Negative mass deficit means the sum of OM and ballast minerals is greater than the measured total mass.

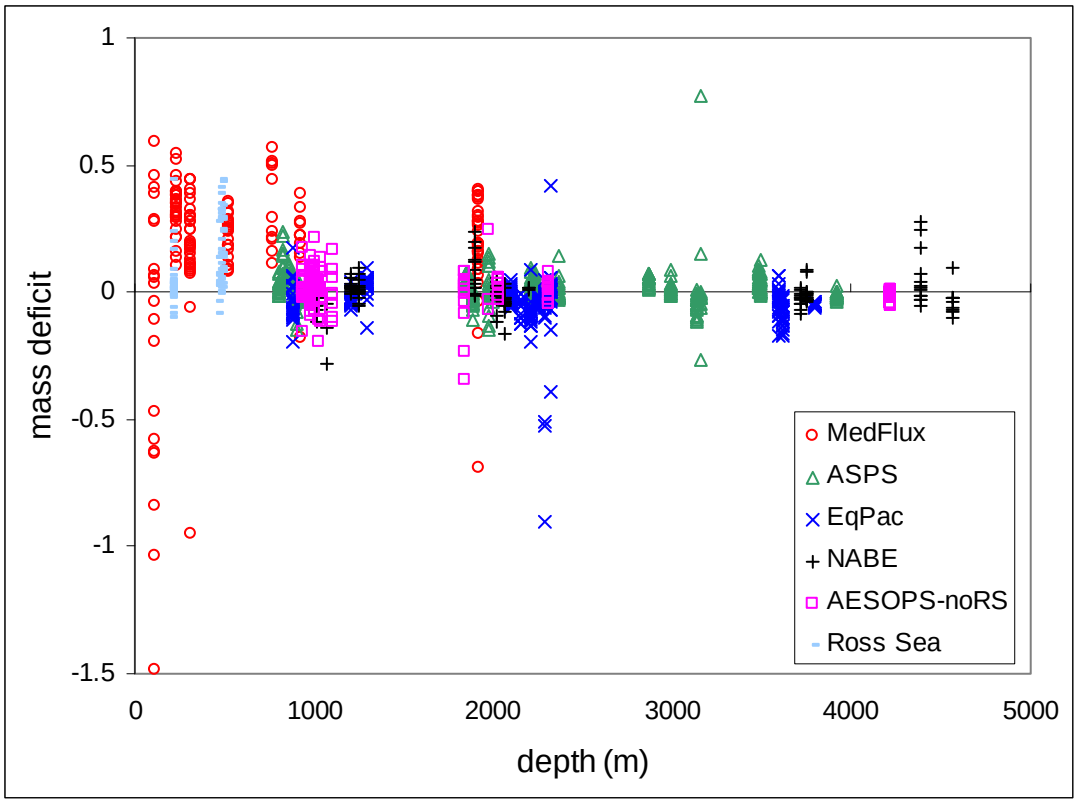


Figure 4.2 Distributions of mass deficits with the information of sampling depths for all samples.

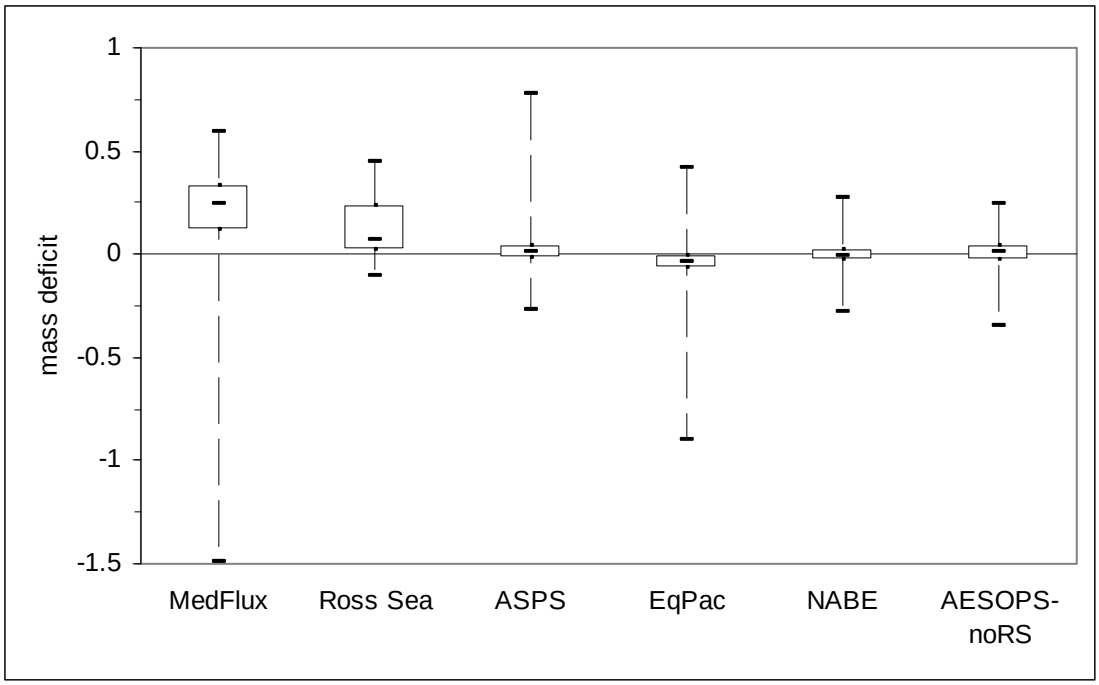


Figure 4.3 Box plots of mass deficit (including median, 25% and 75% quartiles, minimum, and maximum) for sinking particles collected by sediment traps from different open oceans (MedFlux, ASPS, EqPac, NABE, AESOPS-noRS, and the Ross Sea). Mass deficits in MedFlux and Ross Sea are significantly positive, whereas the mass deficits in other open oceans are not.

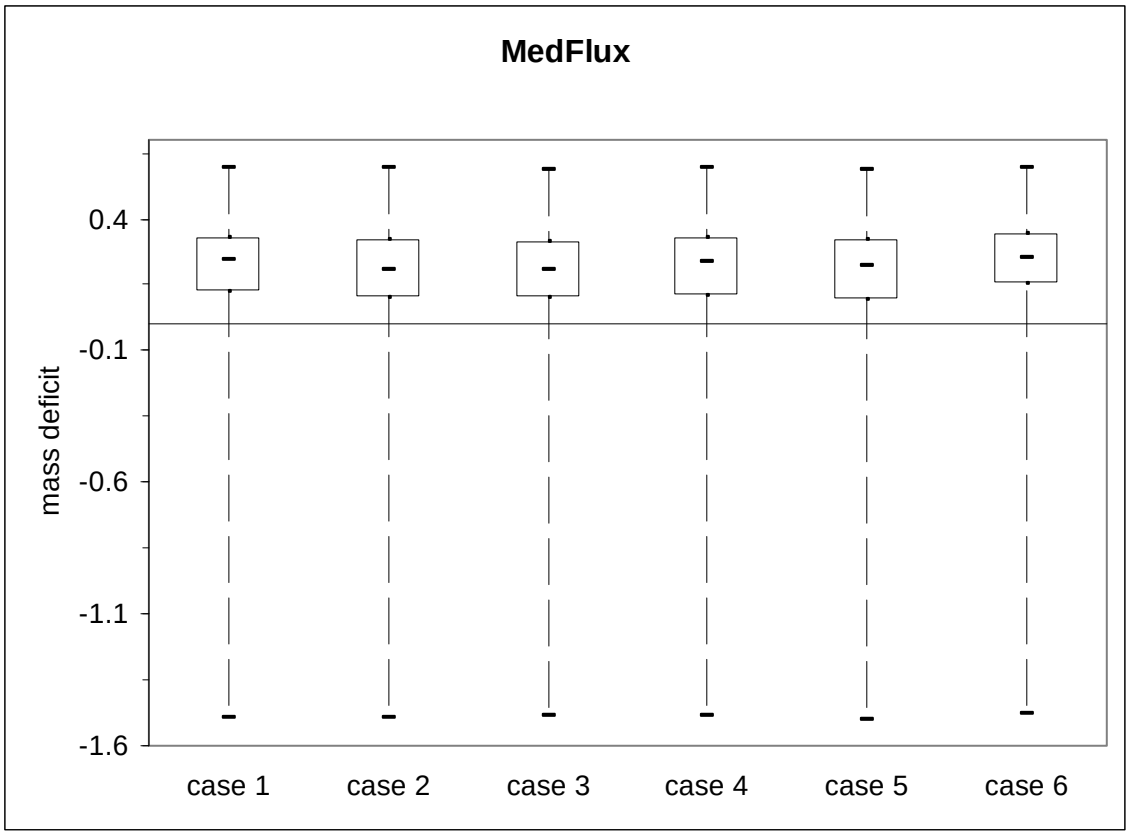
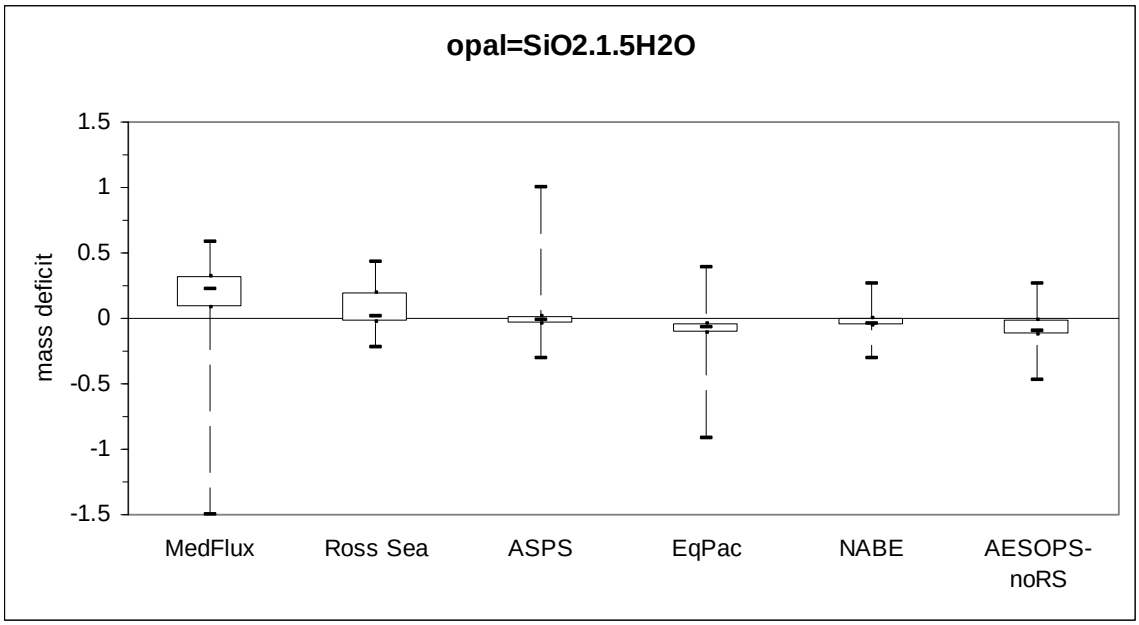


Figure 4.4 Sensitivity test of MD in six cases in MedFlux by changing parameters for the calculation of major components. Case 1: $(Al/Ti)_{litho}$=15.4; lithogenics=11.9*Al_{litho}; opal=$SiO_2.H_2O$; Case 2: $(Al/Ti)_{litho}$=18.5; lithogenics=11.9*Al_{litho}; opal=$SiO_2.H_2O$; Case 3: $(Al/Ti)_{litho}$=18.5; lithogenics=12.15*Al_{litho}; opal=$SiO_2.H_2O$; Case 4: $(Al/Ti)_{litho}$=15.4; lithogenics=12.15*Al_{litho}; opal=$SiO_2.H_2O$; Case 5: $(Al/Ti)_{litho}$=15.4; lithogenics=11.9*Al_{litho}; opal=$SiO_2.1.5H_2O$; Case 6: $(Al/Ti)_{litho}$=15.4; lithogenics=11.9*Al_{litho}; opal=$SiO_2.0.4H_2O$.

(a)



(b)

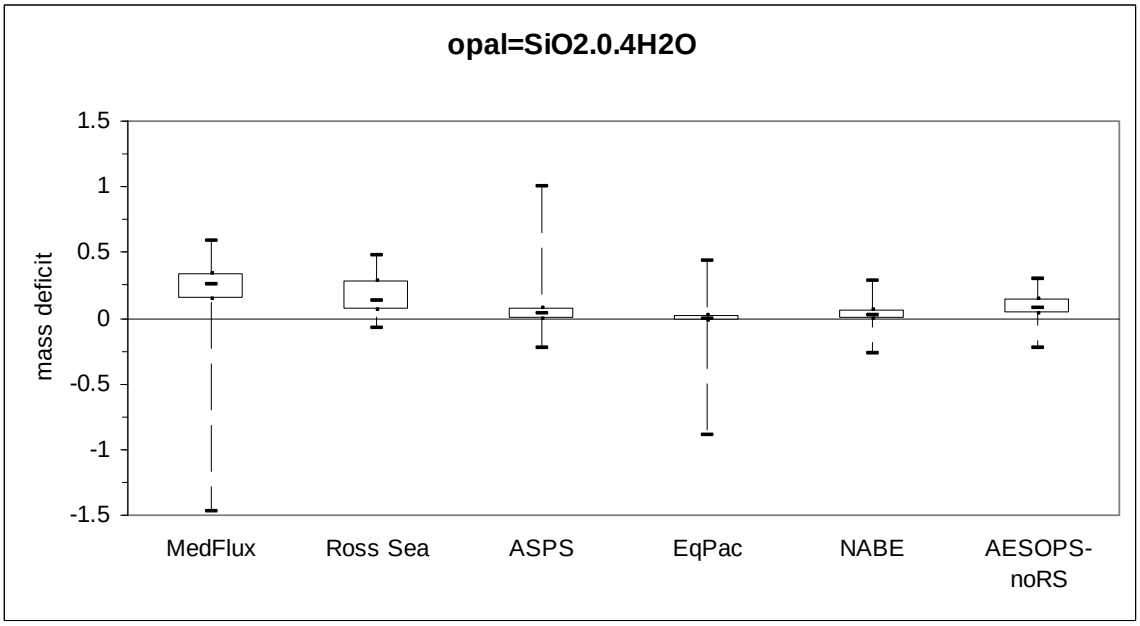


Figure 4.5 A comparison of MD by changing parameters for the calculation of opal for both MedFlux and JGOFS. Opal changes from (a) $SiO_2.1.5H_2O$ to (b) $SiO_2.0.4H_2O$.

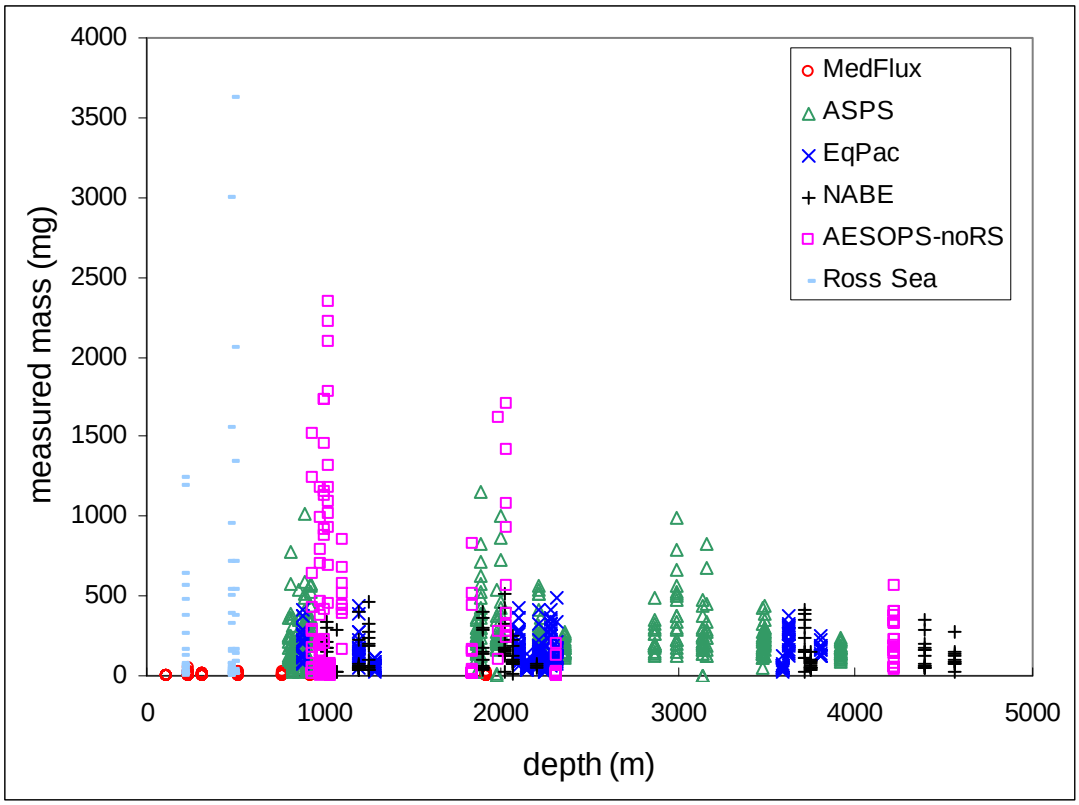
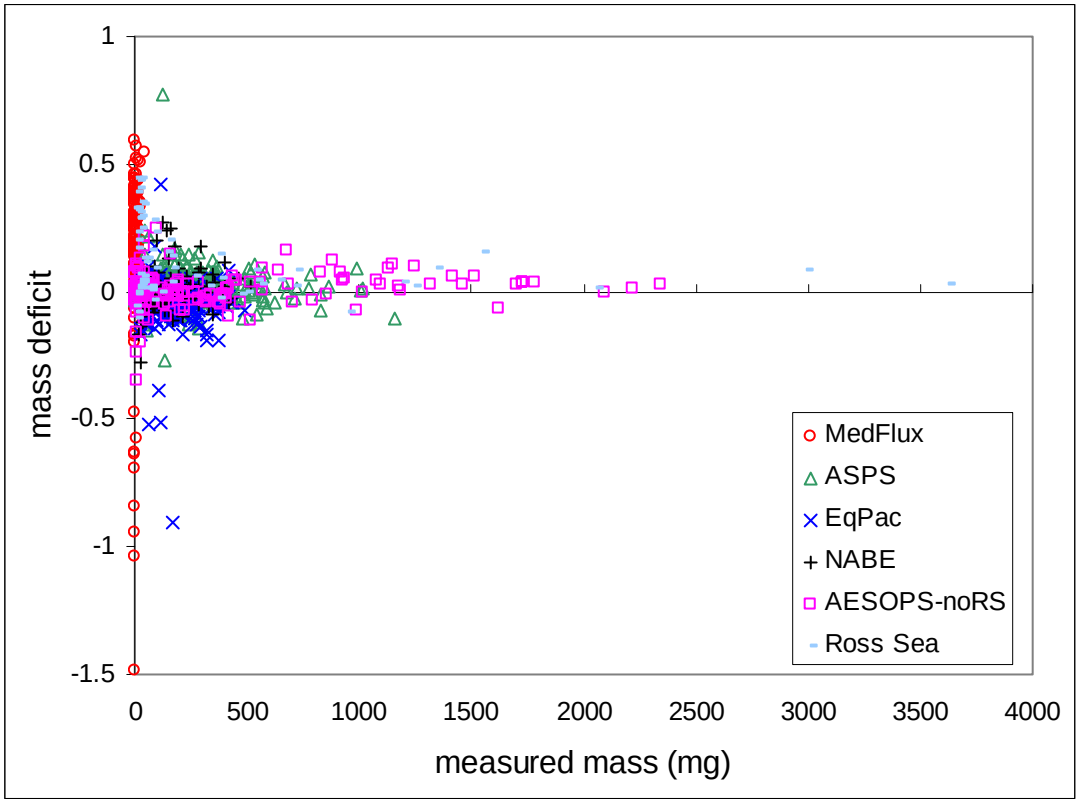
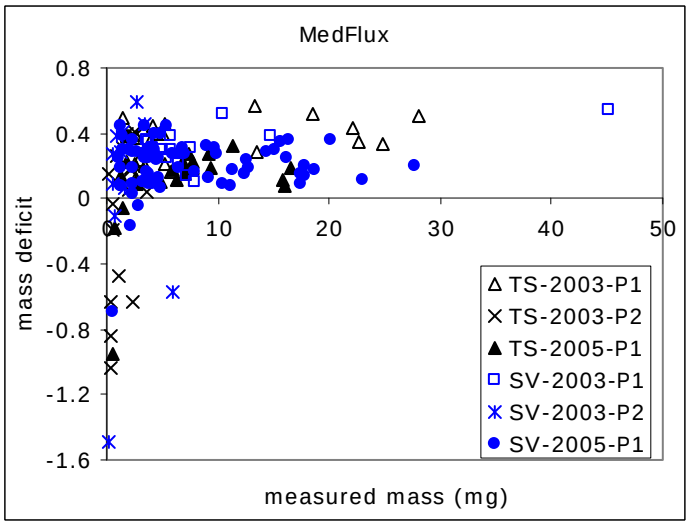


Figure 4.6 Measured masses plotted against depth sampling depth for all samples.

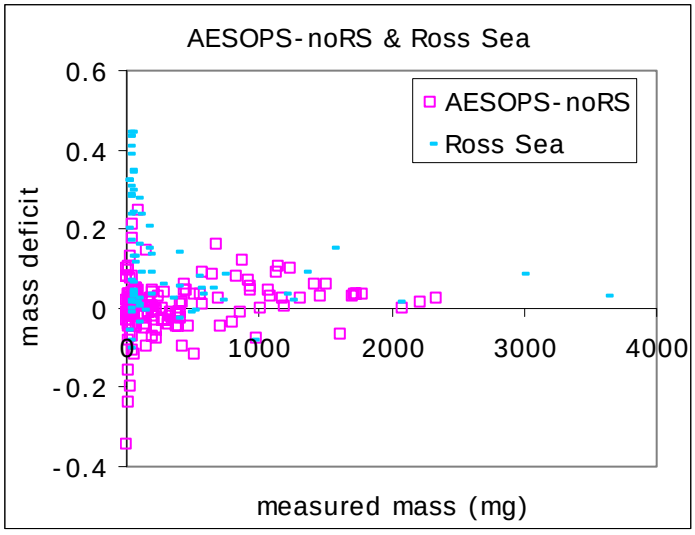
(a)


1
0.5
0
-0.5
-1
-1.5
mass deficit
0
500
1000
1500
2000
2500
3000
3500
4000
measured mass (mg)
MedFlux
ASPS
EqPac
NABE
AESOPS-noRS
Ross Sea

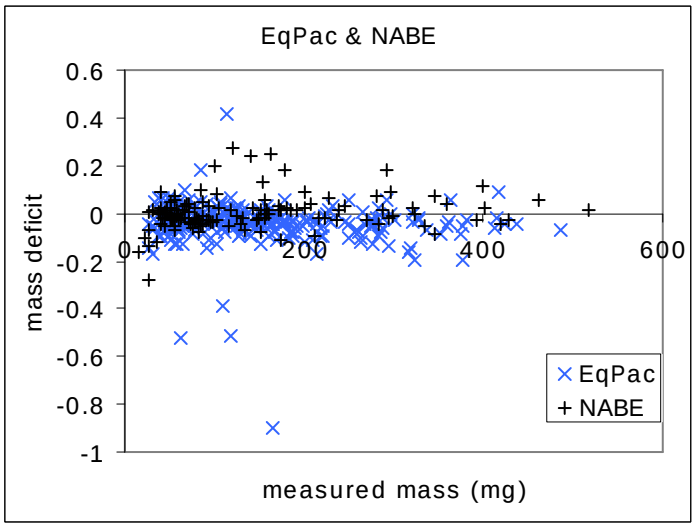
(b)


MedFlux
0.8
0.4
0
-0.4
-0.8
-1.2
-1.6
mass deficit
10
20
30
40
50
measured mass (mg)
TS-2003-P1
TS-2003-P2
TS-2005-P1
SV-2003-P1
SV-2003-P2
SV-2005-P1

(c)


AESOPS-noRS & Ross Sea
AESOPS-noRS
Ross Sea
mass deficit
measured mass (mg)
0.6
0.4
0.2
0
-0.2
-0.4
0
1000
2000
3000
4000

(d)


EqPac & NABE
EqPac
NABE
mass deficit
measured mass (mg)
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
0
200
400
600

(e)

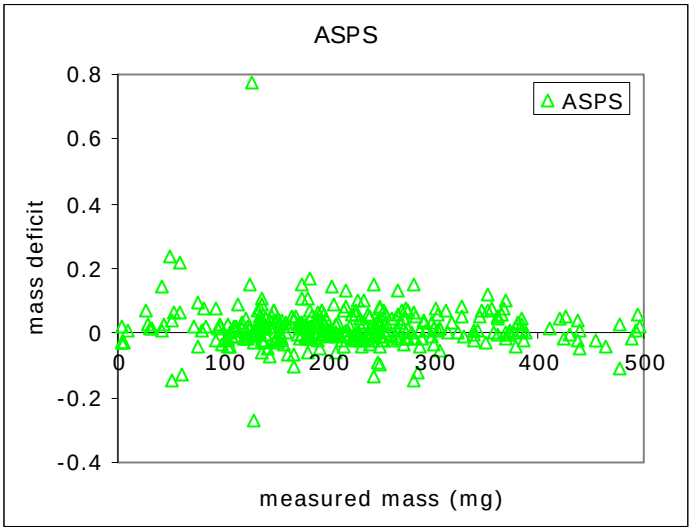


Figure 4.7 The distribution of mass deficits and the measured masses for sinking particles collected for (a) all of the studies; and for (b) MedFlux; (c) AESOPS-noRS and Ross Sea; (d) EqPac and NABE; and (e) ASPS. In MedFlux, the measured total mass is based on a 20% or 30% subsample. For the JGOFS samples, those splits were 30% (ASPS, EqPac, NABE, and AESOPS-noRS) and 20% (Ross Sea), respectively. Mass deficits were usually seen in smaller size samples. As measured total mass increases, the range of mass deficits become smaller; mass deficits approach zero, except MedFlux and Ross Sea samples, where mass deficits are slightly positive.

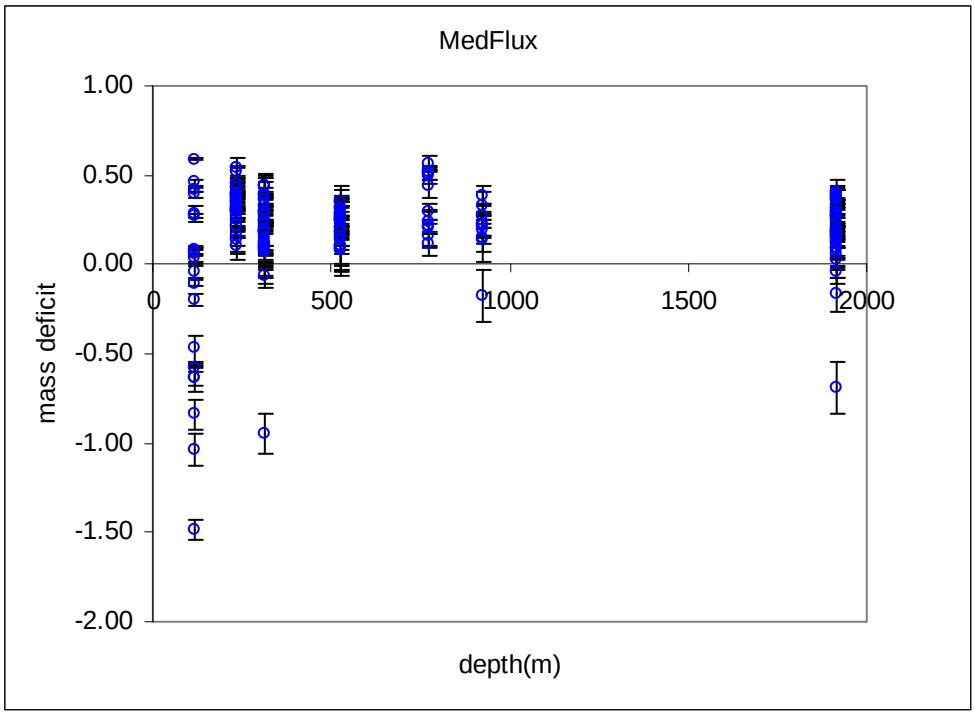
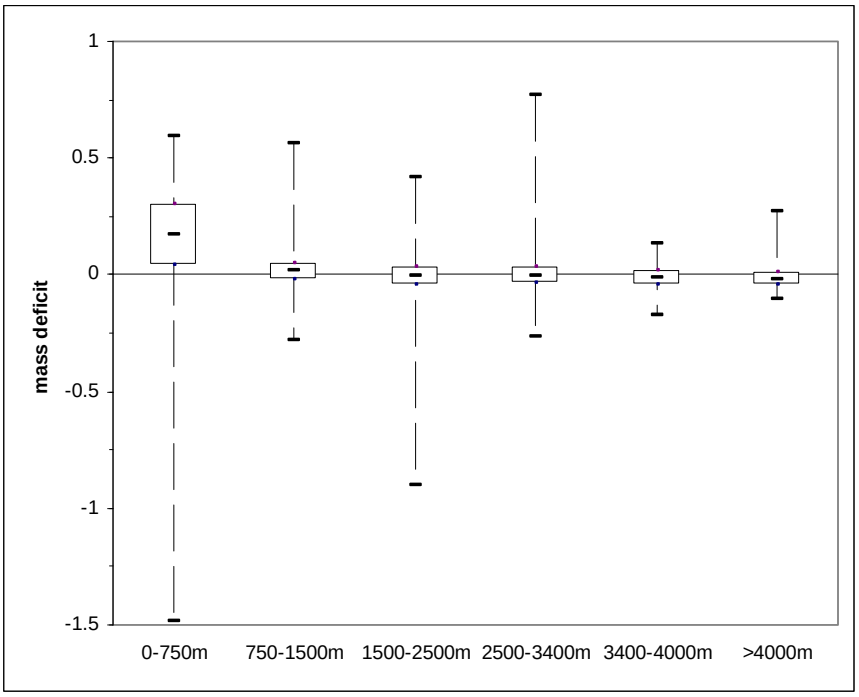


Figure 4.8 The MD and their analytical errors in MedFlux, where the analytical error for C is ±2%, for biogenic Si is high as 50%, and for Al is about 20%.

(a)



(b)

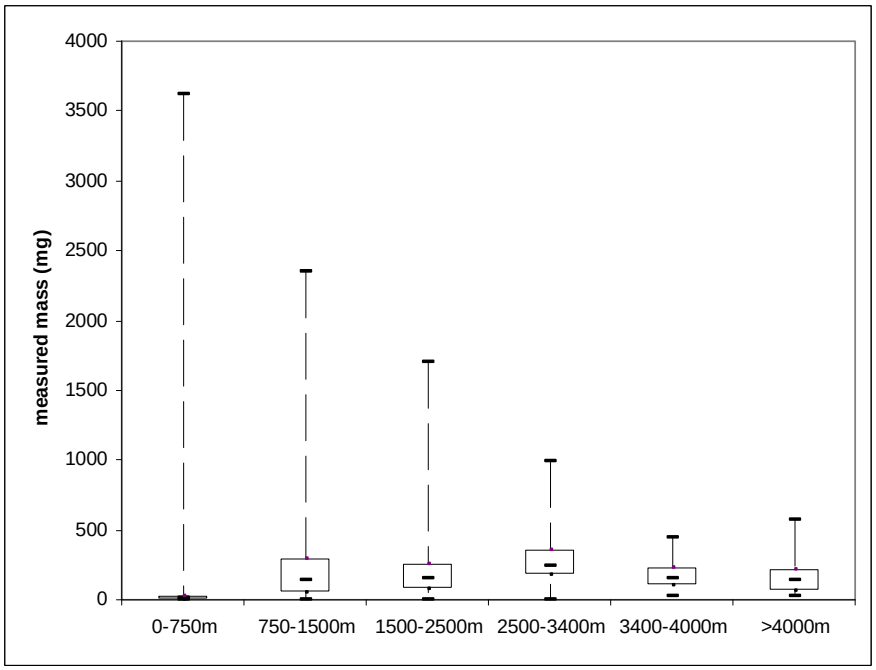
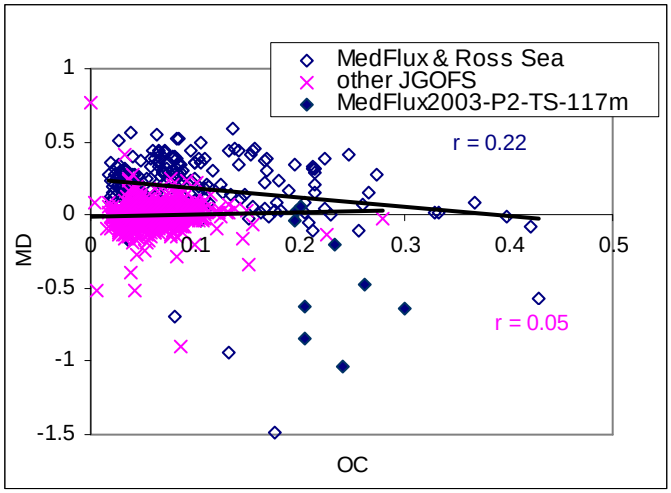
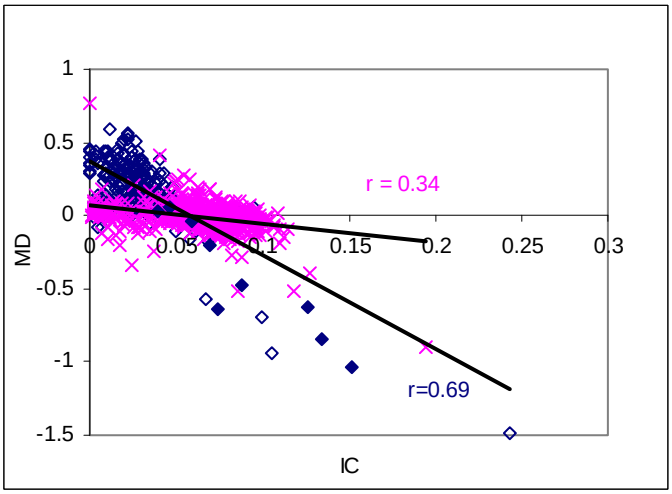


Figure 4.9 Data from all studies were divided into 6 groups based on their collection depth (see text for details). Box plots for (a) mass deficits, and (b) measured masses are shown for each group.

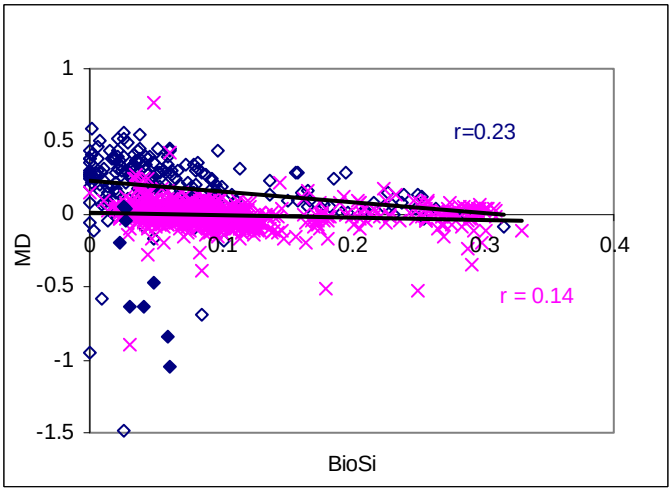
(a)


MedFlux & Ross Sea
other JGOFS
MedFlux2003-P2-TS-117m
r = 0.22
r = 0.05
MD
OC
1
0.5
0
-0.5
-1
-1.5
0
0.1
0.2
0.3
0.4
0.5

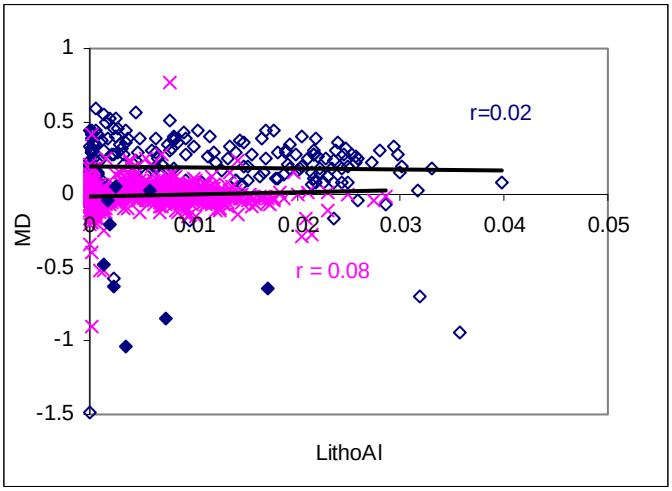
(b)


r = 0.34
r=0.69
MD
IC
1
0.5
0
-0.5
-1
-1.5
0
0.05
0.1
0.15
0.2
0.25
0.3

(c)


r=0.23
r = 0.14
MD
BioSi
1
0.5
0
-0.5
-1
-1.5
0
0.1
0.2
0.3
0.4

(d)



(e)

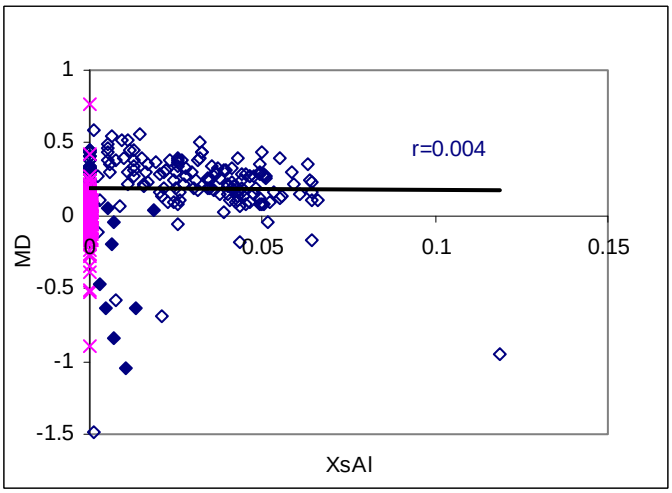


Figure 4.10 Regressions of mass deficit (MD) on (a) OC, (b) IC, (c) biogenic silica (BioSi), (d) lithogenic Al (LithoAl), and (e) excess Al (XsAl). Separate regression lines are drawn separately for (i) MedFlux and the Ross Sea, and (ii) for all of the other JGOFS datasets. Most samples of TS 117m in MedFlux 2003 P2 were outliers from the rest of MedFlux study, and were excluded in the regressions.

CHAPTER FIVE: Conclusions and future directions

1. Conclusions

Sinking particle is a key pathway connecting surface to the deep ocean, and plays an important role in global cycles of elements such as carbon. This thesis focused on the vertical dynamics of sinking particles, and their chemical compositions and degradation trajectories, using statistical approaches on the data from the MedFlux and US JGOFS projects. Overall, three key aspects of sinking particles were investigated: How fast do sinking particles sink? How to find out the degradation trajectories of sinking particles using PCA? And how to balance the mass of sinking particles between the direct measurement and the sum-up of individual components? Here I summarize the major findings.

I developed a new "benchmark" method to estimate the settling velocity based on fitting time-series sediment trap data into Fourier series. This method is based on the field collecting data rather than any laboratory observations. In addition, it has more advantages than any former benchmark methods, which relied more on direct observation of trap data (i.e., Honjo and Manganini, 1993; Berelson, 2002) In this protocol, first, the offset between cups of upper and lower traps could be any real numbers that are not restricted to integer multiples of cup rotation time. Second, the estimation of SVs for upper and lower pairs of traps at one site could proceed simultaneously. Furthermore,

this study shows that the accurate estimation of settling velocity depends upon the type of tracers and the resolution of data. The use of elemental ratios (OC/Al, OC/BioSi, OC/IC, etc.) gives less consistent settling velocities among the tracers than using single elements; the low resolution data are not sufficient for our method to pick out the benchmark pattern. The results, with the use of single element by high resolution data, show that the settling velocity for sinking particles at the DYFAMED and other US JGOFS study sites is 205 m/d by average. This value is on the upper end compared to well accepted records 50 – 200 m/d (Siegel and Deuser, 1997; Waniek et al., 2000). Moreover, I did not find significant difference on the estimation of settling velocity among total mass and different single tracer (i.e., OC, IC, BioSi, or Al). Comparing the settling velocity from sediment traps of three different depths, sinking particles do not seem to sink faster with depth in the open ocean, as claimed by Berelson (2002).

The principles of PCA and how it is applies to organic geochemical datasets, especially those of sinking particles, are described and discussed. From several organic compositional datasets, the comparison of 3D and 2D PCA plots on the chemical composition data of sinking particles shows that using any one dimensional principal component (PC) as the degradation index is oversimplified; a 3D PCA plot offers more information than a 2D PCA plot on the degradation trajectory for sinking particles. However, a 2D PCA plot works as well as a 3D plot if the first two PCs have explained more than approximately 60% of the variance and the third eigenvalue is not comparative to the second one. Thus, this study sets up a criterion on how to use PCA more efficiently for the data mining.

Dauwe et al. (1998; 1999) first applied PCA on compositional data of amino acids in organic matter subjected to different degree of degradation. Their PC1 explained the most variance of the data and was defined as the degradation index (DI) since degradation might be the major reason for the variance. This works well for samples having very different diagenetic histories. However, for samples with relatively similar diagenetic status, site scores along PC1 often do not solely represent the DI, as sources or diagenetic pathways may also be included. For example, Ingalls et al. (2003) found that PC1 also includes the source information of organic matter in sinking particles samples from the Southern Ocean. In this study, we showed that the coordination can be re-orientated so that the new axis can be lined up to the direction that mostly represents the degradation index. This technique is useful to sort out the information we need, especially when dealing with a complicated dataset.

The third major finding of this thesis is that the sum of major chemical compositions (OM, $CaCO_3$, opal, lithogenic minerals, and excess Al hydrate) is significantly less than the measured total mass in sediment traps in the Mediterranean and Ross Seas, where sinking particles were collected at depths less than 800 m. This mass deficit, however, not found in deeper (> 800 m) traps from other US JGOFS studies. It should be noted that the sample size in IRS sediment traps (usually less than 100 mg in each cup) is one or two orders of magnitude lower relative to the cone-shape traps, so that the analytical errors could be relatively larger. But these errors could not explain the mass deficit in a systematic way. The multiple regression results suggest that there might be water bound with or remaining in organic molecules and ballast minerals, especially for samples collected from shallow sediment traps. Wakeham et al (1997) pointed out

that there is a dramatic difference on biogeochemical properties as particles sink through the twilight zone. This suggests that the bound water might be very different for sinking particles with depth. This part of water was not counted in total mass when summing up the individual components, but was in the measured total mass. The mechanism of how the water bounded with OM and ballast minerals is still not clear.

2. Future work

Understanding the biogeochemistry of sinking particles is an important but challenging topic in the field of marine chemistry. Based on this thesis, I use the perspective of mathematics to suggest several future directions.

Estimation of settling velocity of sinking particles is a key if we are to model the biological pump. Obviously, the "in-situ" estimation based field-observation is more reliable due to the minimum sample handling and indeed represent in-situ settling compared to any laboratory measurements. The results by our developed "benchmark" method suggest that the cups in the sediment traps should be designed to open within 5 days in order to get more accurate estimation on settling velocity considering the SV is about 200 m/d on average and the traps are deployed 1000 m apart. To get more distributions of SVs in the open ocean in future, more field collections of sinking particles are suggested at different open oceans, and at every 1000 m in depth.

PCA is a powerful tool for analyzing a large-size dataset, and its potential in the field of marine geochemistry has only begun to be explored. For example, we mostly rely on the semi-quantitative property of PCA to visualize the degradation trajectories,

but rarely quantify the similarity or differences among samples. The distance between samples on a 3D plot could be studied in the future to represent the similarity index, which is useful for grouping samples. In addition, the current method for the determination of the degradation direction in a 3D PCA is very subjective and case dependent. For example, amino acids serine, glutamic acid, threonine, and aspartic acid are usually biomarkers of relatively fresh organic matter. To simplify problem, we only used the coordinate of serine for the determination of degradation direction in one of the case studies in Chapter 3. How to use a mathematical method for finding the degradation index, which combines most of the biomarker information, is a challenge for the future.

In this thesis, I hypothesized that the water bound with OM and ballast minerals is different in the sinking particles collected between shallow and deep water, and this difference could be used to explain the mass deficit in the Mediterranean Sea and Ross Sea. Sinking particles should be further collected in shallow waters from other open oceans sites other than the Mediterranean and Ross Seas to test and confirm this result. If this phenomenon is universal, in addition to solving the issue of mass deficit, it may also provide mechanistic insight on how sinking particles undergo diagenetic change from surface to ocean interior. For example, the bound water in OM and/or ballast minerals may be important in preserving organic matter in surface water, and organic matter is degraded efficiently after the bound water is lost. To evaluate the bound water, we could change the sample pretreatment procedure by introducing chemical or physical reactions, to see whether that will affect the measurement of total mass. We can also use modern microscopic techniques such as atom force microscopy to trace the fate of the bound water during a decomposition experiment. Nevertheless, this study suggests that a clear

definition of mass of an oceanic sample is needed. In other words, whether the bound water should be included in the "true mass" or not? Or at least, we should advocate a standard protocol for sample pretreatment procedure in the future.

References

Berelson, W. M., 2002. Particle settling rates increase with depth in the ocean. *Deep-Sea Research II* **49**, 237-251.

Dauwe, B. and Middelburg, J. J., 1998. Amino acids and hexosamines as indicators of organic matter degradation state in North Sea sediments. *Limnology & Oceanography* **43**, 782-798.

Dauwe, B., Middelburg, J. J., Herman, P. M. J., and Heip, C. H. R., 1999. Linking diagenetic alteration of amino acids and bulk organic matter reactivity. *Limnology & Oceanography* **44**, 1809-1814.

Honjo, S. and Manganini, S. J., 1993. Annual biogenic particle fluxes to the interior of the north-Atlantic ocean - studied at 34-degrees-N 21-degrees-W and 48-degrees-N 21-degrees-W. *Deep-Sea Research II* **40**, 587-607.

Ingalls, A. E., Lee, C., Wakeham, S. G., and Hedges, J. I., 2003. The role of biominerals in the sinking flux and preservation of amino acids in the Southern Ocean along 170 degrees W. *Deep-Sea Research II* **50**, 713-738.

Siegel, D. A. and Deuser, W. G., 1997. Trajectories of sinking particles in the Sargasso Sea: modeling of statistical funnels above deep-ocean sediment traps. *Deep-Sea Research I* **44**, 1519-1541.

Wakeham, S. G., Lee, C., Hedges, J. I., Hernes, P. J., and Peterson, M. L., 1997. Molecular indicators of diagenetic status in marine organic matter. *Geochimica et Cosmochimica Acta* **61**, 5363-5369.

Waniek, J., Koeve, W., and Prien, R. D., 2000. Trajectories of sinking particles and the catchment areas above sediment traps in the northeast Atlantic. *Journal of Marine Research* **58**, 983-1006.

Bibliography

Adams, S. J., Hawkes, G. E., and Curzon, E. H., 1991. A solid-state Si-29 nuclear-magnetic-resonance study of opal and other hydrous silicas. *American Mineralogist* **76**, 1863-1871.

Alldredge, A. L. and Gotschalk, C. C., 1989. Direct observations of the mass flocculation of diatom blooms - characteristics, settling velocities and formation of diatom aggregates. *Deep-Sea Research* **36**, 159-171.

Anderson, R. F., 2003. Chemical tracers of particle transport. *Treatise on Geochemistry*, Holland HD and Turekian KK. Elsevier Ltd. New York. pp 247-273.

Armstrong, R. A., Lee, C., Hedges, J. I., Honjo, S., and Wakeham, S. G., 2002. A new, mechanistic model for organic carbon fluxes in the ocean based on the quantitative association of POC with ballast minerals. *Deep-Sea Research II* **49**, 219-236.

Armstrong. R.A., Peterson, M.L., Lee, C., and Wakeham, S.G. Settling velocity spectra and the ballast ratio hypothesis. *Deep-Sea Research II* (submitted)

Bacon, M. P., Cochran, J. K., Hirschberg, D., Hammar, T. R., and Fleer, A. P., 1996. Export flux of carbon at the equator during the EqPac time-series cruises estimated from Th-234 measurements. *Deep-Sea Research II* **43**, 1133-1153.

Berelson, W.M., 2002. Particle settling rates increase with depth in the ocean. *Deep-Sea Research II* **49**, 237-251.

Bienfang, P. K., 1980. Phytoplankton sinking rates in oligotrophic waters off Hawaii, USA. *Marine Biology* **61**, 69-77.

Billett, D. S. M., Lampitt, R. S., Rice, A. L., and Mantoura, R. F. C., 1983. Seasonal sedimentation of Phytoplankton to the deep-sea benthos. *Nature* **302**, 520-522.

Bishop, J. K. B., Collier, R. W., Kettens, D. R., and Edmond, J. M., 1980. The chemistry, biology, and vertical flux of particulate matter from the upper 1500m of the Panama basin. *Deep-Sea Research* **27**, 615-640.

Bishop, J. K. B., Schupack, D., Sherrell, R. M., and Conte, M., 1984. A multiple-unit large-volume in situ filtration system for sampling oceanic particulate matter in mesoscale environments. *Advances in Chemistry Series* **209**, 155-175.

Boehme, J., Coble, P., Conmy, R., and Stovall-Leonard, A., 2004. Examining CDOM fluorescence variability using principal component analysis: seasonal and regional

modeling of three-dimensional fluorescence in the Gulf of Mexico. *Marine Chemistry* **89**, 3-14.

Bruland, K. W. and Silver, M. W., 1981. Sinking rates of fecal pellets from gelatinous Zooplankton (Salps, Pteropods, Doliolids). *Marine Biology* **63**, 295-300.

Buesseler, K. O., Benitez-Nelson, C. R., Moran, S. B., Burd, A., Charette, M., Cochran, J. K., Coppola, L., Fisher, N. S., Fowler, S. W., Gardner, W., Guo, L. D., Gustafsson, O., Lamborg, C., Masque, P., Miquel, J. C., Passow, U., Santschi, P. H., Savoye, N., Stewart, G., and Trull, T., 2006. An assessment of particulate organic carbon to thorium-234 ratios in the ocean and their impact on the application of Th-234 as a POC flux proxy. *Marine Chemistry* **100**, 213-233.

Burd, A. B. and Jackson, G. A., 2002. Modeling steady-state particle size spectra. *Environmental Science & Technology* **36**, 323-327.

Burnham, K.P., and Anderson, D.R., 1998. *Model selection and inference: a practical information-theoretic approach*. Springer, New York.

Cattell, R. B., 1966. The scree test for the number of factors. *Multivariate Behavioral Research* **1**, 245-276.

Chou, L. and Wollast, R., 1997. Biogeochemical behavior and mass balance of dissolved aluminum in the western Mediterranean Sea. *Deep-Sea Research II* **44**, 741-768.

Clegg, S. L. and Whitfield, M., 1990. A generalized-model for the scavenging of trace-metals in the open ocean .1. Particle cycling. *Deep-Sea Research* **37**, 809-832.

Clegg, S. L. and Whitfield, M., 1991. A generalized-model for the scavenging of trace-metals in the open ocean .2. Thorium scavenging. *Deep-Sea Research* **38**, 91-120.

Cochran, J. K., Buesseler, K. O., Bacon, M. P., and Livingston, H. D., 1993. Thorium isotopes as indicators of particle dynamics in the upper ocean - results from the JGOFS North-Atlantic Bloom Experiment. Deep-Sea Research I **40**, 1569-1595.

Collier, R., Dymond, J., Honjo, S., Manganini, S., Francois, R., and Dunbar, R., 2000. The vertical flux of biogenic and lithogenic material in the Ross Sea: Moored sediment trap observations 1996-1998. *Deep-Sea Research II* **47**, 3491-3520.

Dauwe, B. and Middelburg, J. J., 1998. Amino acids and hexosamines as indicators of organic matter degradation state in North Sea sediments. *Limnology & Oceanography* **43**, 782-798.

Dauwe, B., Middelburg, J. J., Herman, P. M. J., and Heip, C. H. R., 1999. Linking diagenetic alteration of amino acids and bulk organic matter reactivity. *Limnology & Oceanography* **44**, 1809-1814.

de la Rocha, C. L., 2003. The biological pump. *Treatise on Geochemistry*, Holland HD and Turekian KK. Elsevier Ltd. New York. pp 83-111.

Deuser, W. G., 1986. Seasonal and interannual variations in deep-water particle fluxes in the Sargasso Sea and their relation to surface hydrography. *Deep-Sea Research* **33**, 225-246.

Deuser, W. G., Ross, E. H., and Anderson, R. F., 1981. Seasonality in the supply of sediment to the deep Sargasso Sea and implications for the rapid transfer of matter to the deep ocean. *Deep-Sea Research* **28**, 495-505.

Diercks, A. R. and Asper, V. L., 1997. In situ settling speeds of marine snow aggregates below the mixed layer: Black Sea and Gulf of Mexico. *Deep-Sea Research I* **44**, 385-398.

Dittmar, T., 2004. Evidence for terrigenous dissolved organic nitrogen in the Arctic deep sea. *Limnology & Oceanography* **49**, 148-156.

Dymond, J., Collier, R., McManus, J., Honjo, S., and Manganini, S., 1997. Can the aluminum and titanium contents of ocean sediments be used to determine the paleoproductivity of the oceans? *Paleoceanography* **12**, 586-593.

Edwards, A. W. F., 1992. *Likelihood*. Johns Hopkins University Press, Baltimore, MD, USA.

Eisma, D., 1986. Flocculation and de-flocculation of suspended matter in estuaries. *Netherlands Journal of Sea Research* **20**, 183-199.

Eppley, R. W. and Peterson, B. J., 1979. Particulate organic matter flux and planktonic new production in the deep ocean. *Nature* **282**, 677-680.

Formenti, P., Elbert, W., Maenhaut, W., Haywood, J., and Andreae, M. O., 2003. Chemical composition of mineral dust aerosol during the Saharan Dust Experiment (SHADE) airborne campaign in the Cape Verde region. *Journal of Geophysical Research* **108**, No. D18, 8576, doi:10.1029//2002JD002648.

Fowler, S. W. and Knauer, G. A., 1986. Role of large particles in the transport of elements and organic-compounds through the oceanic water column. *Progress in Oceanography* **16**, 147-194.

François, R., Honjo, S., Krishfield, R., and Manganini, S., 2002. Factors controlling the flux of organic carbon to the bathypelagic zone of the ocean. *Global Biogeochemical Cycles* **16**:GB1087.

Gaffey, S. J., 1995. H_2O and OH in echinoid calcite - a spectroscopic study. *American Mineralogist* **80**, 947-959.

Goni, M. A., Yunker, M. B., Macdonald, R. W., and Eglinton, T. I., 2000. Distribution and sources of organic biomarkers in arctic sediments from the Mackenzie River and Beaufort Shelf. *Marine Chemistry* **71**, 23-51.

Gonzalez-Silvera, A., Santamaria-Del-Angela, E., Garcia, V. M. T., Garcia, C. A. E., Millan-Nunez, R., and Muller-Karger, F., 2004. Biogeographical regions of the tropical and subtropical Atlantic Ocean off South America: classification based on pigment (CZCS) and chlorophyll-a (SeaWiFS) variability. *Continental Shelf Research* **24**, 983-1000.

Goutx, M., Wakeham, S. G., Lee, C., Duflos, M., Guigue, C., Liu, Z. F., Moriceau, B., Sempere, R., Tedetti, M., and Xue, J. H., 2007. Composition and degradation of marine particles with different settling velocities in the northwestern Mediterranean Sea. *Limnology & Oceanography* **52**, 1645-1664.

Haake, B., Ittekkot, V., Rixen, T., Ramaswamy, V., Nair, R. R., and Curry, W., 1993. Seasonality and interannual variability of particle fluxes to the deep Arabian Sea. *Deep-Sea Research I* **40**, 1323-1344.

Hedges, J. I., Baldock, J. A., Gelinas, Y., Lee, C., Peterson, M. L., and Wakeham, S. G., 2002. The biochemical and elemental compositions of marine plankton: A NMR perspective. *Marine Chemistry* **78**, 47-63.

Hilborn, R., and Mangel, M., 1997. *The ecological detective: confronting models with data*. Princeton University Press, Princeton, NJ, USA.

Hill, P. S., 1998. Controls on floc size in the sea. *Oceanography* **11**, 13-18.

Hill, P. S., Syvitski, J. P., Cowan, E. A., and Powell, R. D., 1998. In situ observations of floc settling velocities in Glacier Bay, Alaska. *Marine Geology* **145**, 85-94.

Honjo, S. and Manganini, S. J., 1993. Annual biogenic particle fluxes to the interior of the north-Atlantic ocean - studied at 34-degrees-N 21-degrees-W and 48-degrees-N 21-degrees-W. *Deep-Sea Research II* **40**, 587-607.

Honjo, S. and Manganini, S. J., 1993. Annual biogenic particle fluxes to the interior of the north-Atlantic ocean - studied at 34-degrees-N 21-degrees-W and 48-degrees-N 21-degrees-w. *Deep-Sea Research II* **40**, 587-607.

Honjo, S., 1982. Seasonality and interaction of biogenic and lithogenic particulate flux at the Panama basin. *Science* **218**, 883-884.

Honjo, S., 1996. Fluxes of particles to the interior of the open ocean. *Particle flux in the ocean*. Ittekkot, V. John Wiley & Sons, Chichester ; New York, pp.91-154.

Honjo, S., Dymond, J., Collier, R., and Manganini, S. J., 1995. Export production of particles to the interior of the equatorial Pacific-Ocean during the 1992 EqPac Experiment. *Deep-Sea Research II* **42**, 831-870.

Honjo, S., Dymond, J., Prell, W., and Ittekkot, V., 1999. Monsoon-controlled export fluxes to the interior of the Arabian Sea. *Deep-Sea Research II* **46**, 1859-1902.

Honjo, S., Francois, R., Manganini, S., Dymond, J., and Collier, R., 2000. Particle fluxes to the interior of the Southern Ocean in the Western Pacific sector along 170 degrees W. *Deep-Sea Research II* **47**, 3521-3548.

Honjo, S., Manganini, S. J., Krishfield, R. A., and Francois, R., 2008. Particulate organic carbon fluxes to the ocean interior and factors controlling the biological pump: A synthesis of global sediment trap programs since 1983. *Progress in Oceanography* **76**, 217-285.

Hu, J. F., Zhang, H. B., and Peng, P. A., 2006. Fatty acid composition of surface sediments in the subtropical Pearl River estuary and adjacent shelf, Southern China. *Estuarine Coastal and Shelf Science* **66**, 346-356.

Ingalls, A. E., Lee, C., Wakeham, S. G., and Hedges, J. I., 2003. The role of biominerals in the sinking flux and preservation of amino acids in the Southern Ocean along 170 degrees W. *Deep-Sea Research II* **50**, 713-738.

Jackson, G. A. and Burd, A. B., 1998. Aggregation in the marine environment. *Environmental Science & Technology* **32**, 2805-2814.

Jackson, G. A., 1990. A model of the formation of marine algal flocs by physical coagulation processes. *Deep-Sea Research part A* **37**, 1197-1211.

Jacobs, G. A., Emery, W. J., and Born, G. H., 1993. Rossby Waves in the Pacific-ocean extracted from geosat altimeter data. *Journal of Physical Oceanography* **23**, 1155-1175.

King, J. R. and Jackson, D. A., 1999. Variable selection in large environmental data sets using principal components analysis. *Environmetrics* **10**, 67-77.

Klaas, C. and Archer, D. E., 2002. Association of sinking organic matter with various types of mineral ballast in the deep sea: Implications for the rain ratio. *Global Biogeochemistry Cycle* **16**, 1-16.

Kryc, K. A., Murray, R. W., and Murray, D. W., 2003. Al-to-oxide and Ti-to-organic linkages in biogenic sediment: relationships to paleo-export production and bulk Al/Ti. *Earth & Planetary Science Letters* **211**, 125-141.

Lattin, J. M., Carroll, J. D., Green, P. E., and Green, P. E., 2003. *Analyzing multivariate data*. Thomson Brooks/Cole, Pacific Grove, CA.

Lee, C. and Cronin, C., 1982. The vertical flux of particulate organic nitrogen in the sea - decomposition of amino-acids in the Peru upwelling area and the equatorial Atlantic. *Journal of Marine Research* **40**, 227-251.

Lee, C. and Cronin, C., 1984. Particulate Amino-Acids in the Sea - Effects of primary productivity and biological decomposition. *Journal of Marine Research* **42**, 1075-1097.

Lee, C., Peterson, M.L., Wakeham, S.G., Armstrong, R.A., Cochran, J.K., Miquel, J.-C., Fowler, S., Hirschberg, D., Beck, A., Xue, J., Particulate organic matter and ballast fluxes measured using in time-series and settling velocity sediment traps in the northwestern Mediterranean Sea. *Deep-Sea Research II* (submitted).

Lee, C., Wakeham, S. G., and Arnosti, C., 2004. Particulate organic matter in the sea: The composition conundrum. *Ambio* **33**, 565-575.

Lee, C., Wakeham, S. G., and Hedges, J. I., 2000. Composition and flux of particulate amino acids and chloropigments in equatorial Pacific seawater and sediments. *Deep-Sea Research I* **47**, 1535-1568.

Liu, Z. F. and Lee, C., 2006. Drying effects on sorption capacity of coastal sediment: The importance of architecture and polarity of organic matter. *Geochimica et Cosmochimica Acta* **70**, 3313-3324.

Liu, Z. F., Lee, C., and Wakeham, S. G., 2006. Effects of mercuric chloride and protease inhibitors on degradation of particulate organic matter from the diatom Thalassiosira pseudonana. *Organic Geochemistry* **37**, 1003-1018.

Loh, A. N. and Bauer, J. E., 2000. Distribution, partitioning and fluxes of dissolved and particulate organic C, N and P in the eastern North Pacific and Southern Oceans. *Deep-Sea Research I* **47**, 2287-2316.

Mann, K. H. and Lazier, J. R. N., 1991. *Dynamics of marine ecosystems : biological-physical interactions in the oceans*. Blackwell Scientific Publications, Boston.

Mariano, A. J., Hitchcock, G. L., Ashjian, C. J., Olson, D. B., Rossby, T., Ryan, E., and Smith, S. L., 1996. Principal component analysis of biological and physical variability in a Gulf Stream meander crest. *Deep-Sea Research I* **43**, 1531-1565.

Massel, S. R., 2001. Wavelet analysis for processing of ocean surface wave records. *Ocean Engineering* **28**, 957-987.

McCave, I. N., 1975. Vertical flux of particles in ocean. *Deep-Sea Research* **22**, 491-502.

Moncheva, S., Gotsis-Skretas, O., Pagou, K., and Krastev, A., 2001. Phytoplankton blooms in Black Sea and Mediterranean coastal ecosystems subjected to anthropogenic eutrophication: Similarities and differences. *Estuarine Coastal and Shelf Science* **53**, 281-295.

Moreda-Pineiro, A., Marcos, A., Fisher, A., and Hill, S. J., 2001. Evaluation of the effect of data pre-treatment procedures on classical pattern recognition and principal components analysis: a case study for the geographical classification of tea. *Journal of Environmental Monitoring* **3**, 352-360.

Mortlock, R. A. and Froelich, P. N., 1989. A simple method for the rapid-determination of biogenic opal in pelagic marine-sediments. *Deep-Sea Research* **36**, 1415-1426.

Passow, U., 2002. Transparent exopolymer particles (TEP) in aquatic environments. *Progress in Oceanography* **55**, 287-333.

Peterson, M. L., Wakeham, S. G., Lee, C., Askea, M. A., and Miquel, J. C., 2005. Novel techniques for collection of sinking particles in the ocean and determining their settling rates. *Limnology & Oceanograpgy-Method* **3**, 520-532.

Peterson, M.L., Fabres, J., Wakeham, S.G., Lee, C., Miquel. J.-C., Sampling the vertical particle flux in the upper water column using a large diameter free-drifting NetTrap adapted to an Indented Rotating Sphere sediment trap. *Deep-Sea Research II* (submitted)

Philip, G. and Watson, D., 1988. Determining the representative composition of a set of sandstone samples. *Geological Magazine* **125**, 267-272.

Prahl, F. G., Dymond, J., and Sparrow, M. A., 2000. Annual biomarker record for export production in the central Arabian Sea. *Deep-Sea Research II* **47**, 1581-1604.

Railsback, L. B., 1999. Patterns in the compositions, properties, and geochemistry of carbonate minerals. *Carbonate Evaporite* **14**, 1-20.

Rama, K. M. and Goldberg, E. D., 1961. Lead-210 in natural waters. *Science* **134**, 98-99.

Ruiz, J. M., McAdon, M. H., and Garces, J. M., 1997. Aluminum complexes as models for Broensted acid sites in zeolites: Structure and energetics of [Al(OH)(4)](-), [Al(H_2O)(6)](3+), and intermediate monomeric species [Al(OH)(x)(H_2O)(n-x)center dot mH(2)O](3-x) obtained by hydrolysis. *Journal of Physical Chemistry B* **101**, 1733-1744.

Sarmiento, J. L. and Gruber, N., 2006. *Ocean biogeochemical dynamics*. Princeton University Press, Princeton, N.J.

Schrimm, M., Buscail, R., and Adjeroud, M., 2004. Spatial variability of the biogeochemical composition of surface sediments in an insular coral reef ecosystem: Moorea, French Polynesia. *Estuarine Coastal and Shelf Science* **60**, 515-528.

Sheldon, R. W., Sutcliff.Wh, and Prakash, A., 1972. Size distribution of particles in ocean. *Limnology & Oceanography* **17**, 327-340.

Sheridan, C. C., Lee, C., Wakeham, S. G., and Bishop, J. K. B., 2002. Suspended particle organic composition and cycling in surface and midwaters of the equatorial Pacific Ocean. *Deep-Sea Research I* **49**, 1983-2008.

Siegel, D. A. and Deuser, W. G., 1997. Trajectories of sinking particles in the Sargasso Sea: modeling of statistical funnels above deep-ocean sediment traps. *Deep-Sea Research I* **44**, 1519-1541.

Siezen, R. J. and Mague, T. H., 1978. Amino-acids in suspended particulate matter from oceanic and coastal waters of Pacific. *Marine Chemistry* **6**, 215-231.

Small, L. F., Fowler, S. W., and Unlu, M. Y., 1979. Sinking rates of natural copepod fecal pellets. *Marine Biology* **51**, 233-241.

Smayda, T. J., 1969. Some measurements of sinking rate of fecal pellets. *Limnology & Oceanography* **14**, 621-625.

Stewart, G., Cochran, J. K., Xue, J. H., Lee, C., Wakeham, S. G., Armstrong, R. A., Masque, P., and Miquel, J. C., 2007. Exploring the connection between Po-210 and organic matter in the northwestern Mediterranean. *Deep-Sea Research I* **54**, 415-427.

Sutton, R. and Sposito, G., 2005. Molecular structure in soil humic substances: The new view. *Environmental Science & Technology* **39**, 9009-9015.

Takahashi, K., Fujitani, N., Yanada, M., and Maita, Y., 2000. Ling-term biogenic particle fluxes in the Bering Sea and the central subarctic Pacific Ocean, 1990-1995. *Deep-Sea Research I* **47**, 1723-1759.

Taylor, S. R. and Mclennan, S. M., 1995. The geochemical evolution of the continental-crust. *Reviews of Geophysics* **33**, 241-265.

Trull, T.W., Bray, S.G., Buesseler, K.O., Lamborg, C.H., Manganini, S., Moy, C., Valdes, J. In-situ measurement of mesopelagic particle sinking rates and the control of

carbon transfer to the ocean interior during the Vertical Flux in the Global Ocean (VERTIGO) voyages in the North Pacific. *Deep-Sea Research II* (in press).

Vandenberg, C. M. G., Boussemart, M., Yokoi, K., Prartono, T., and Campos, M. L. A. M., 1994. Speciation of aluminum, chromium and titanium in the NW Mediterranean. *Marine Chemistry* **45**, 267-282.

Waite, A., Fisher, A., Thompson, P. A., and Harrison, P. J., 1997. Sinking rate versus cell volume relationships illuminate sinking rate control mechanisms in marine diatoms. *Marine Ecology-Progress Series* **157**, 97-108.

Wakeham, S. G. and Lee, C., 1993. Production, transport, and alteration of particulate organic matter in the marine water column. In: Macko, M. E. a. S. (Ed.), *Organic Geochemistry*.

Wakeham, S. G., Lee, C., Farrington, J. W., and Gagosian, R. B., 1984. Biogeochemistry of particulate organic-matter in the oceans - results from sediment trap experiments. *Deep-Sea Research* **31**, 509-528.

Wakeham, S. G., Lee, C., Hedges, J. I., Hernes, P. J., and Peterson, M. L., 1997. Molecular indicators of diagenetic status in marine organic matter. *Geochimica et Cosmochimica Acta* **61**, 5363-5369.

Wakeham, S. G., Lee, C., Peterson, M. L., Liu, Z., Szlosek, J., Putnam, I., and Xue, J. Organic compound composition and fluxes in the Twilight Zone - time series and settling velocity sediment traps during MEDFLUX. *Deep-Sea Research II* (submitted)

Waniek, J., Koeve, W., and Prien, R. D., 2000. Trajectories of sinking particles and the catchment areas above sediment traps in the northeast Atlantic. *Journal of Marine Research* **58**, 983-1006.

Xue, J. and Armstrong, R. A. An improved “benchmark” method for estimating particle settling velocities from time-series sediment trap fluxes. *Deep-Sea Research II* (submitted)

Yamashita, Y. and Tanoue, E., 2003. Distribution and alteration of amino acids in bulk DOM along a transect from bay to oceanic waters. *Marine Chemistry* **82**, 145-160.

Yunker, M. B., Belicka, L. L., Harvey, H. R., and Macdonald, R. W., 2005. Tracing the inputs and fate of marine and terrigenous organic matter in Arctic Ocean sediments: A multivariate analysis of lipid biomarkers. *Deep-Sea Research II* **52**, 3478-3508.

Yunker, M. B., Macdonald, R. W., Veltkamp, D. J., and Cretney, W. J., 1995. Terrestrial and marine biomarkers in a seasonally ice-covered Arctic estuary - integration of multivariate and biomarker approaches. *Marine Chemistry* **49**, 1-50.

Zimmerman, A. R. and Canuel, E. A., 2001. Bulk organic matter and lipid biomarker composition of Chesapeake Bay surficial sediments as indicators of environmental processes. *Estuarine Coastal and Shelf Science* **53**, 319-341.

Appendix A: Mechanics of principal components analysis (PCA)

This appendix contains a self-contained explanation of the mechanics of PCA. All that is assumed is that the readers know how to multiply matrices, and have some basic feeling for sample means and variances.

A1.1. Data matrices

Consider a set of data records $X_1, X_2, \cdots X_j, \cdots, X_n$, $j = 1, \cdots n$, where each record is written as a *row vector* with *m* elements, each of which represents a measurement (variable) in the record:

$$\begin{aligned} X_1 &= [x_{11} \quad x_{12} \quad ... \quad x_{1k} \quad ... \quad x_{1m}]; \\ X_2 &= [x_{21} \quad x_{22} \quad ... \quad x_{2k} \quad ... \quad x_{2m}]; \\ &\cdots \\ X_j &= [x_{j1} \quad x_{j2} \quad ... \quad x_{jk} \quad ... \quad x_{jm}]; \\ &\cdots \\ X_n &= [x_{n1} \quad x_{n2} \quad ... \quad x_{nk} \quad ... \quad x_{nm}]. \end{aligned} \tag{A.1}$$

The total data set can be written as a *column vector* of these data records which, by expanding each row vector, becomes the *data matrix X*:

$$X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_j \\ \vdots \\ X_n \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1k} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2k} & \cdots & x_{2m} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{j1} & x_{j2} & \cdots & x_{jk} & \cdots & x_{jm} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nk} & \cdots & x_{nm} \end{bmatrix}. \tag{A.2}$$

Alternately, the data set can be written as a column vector of row vectors x_k, each of which is a list of the values of variable *k* ($k = 1, \cdots m$) in all data records:

$$x_k = \begin{bmatrix} x_{1k} \\ x_{2k} \\ \vdots \\ x_{jk} \\ \vdots \\ x_{nk} \end{bmatrix} = \begin{bmatrix} x_{1k} & x_{2k} & \cdots & x_{jk} & \cdots & x_{nk} \end{bmatrix}^T, \tag{A.3}$$

where T is the *transpose* operation that transforms a column vector into a row vector or vice versa. The data matrix can therefore also be written:

$$X = \begin{bmatrix} x_1 & x_2 & \cdots & x_k & \cdots & x_m \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1k} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2k} & \cdots & x_{2m} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{j1} & x_{j2} & \cdots & x_{jk} & \cdots & x_{jm} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nk} & \cdots & x_{nm} \end{bmatrix} . \qquad \text{(A.4)}$$

A1.2. Summary statistics

Recall that the *sample mean* of a set of measurements of variable *k* is given by

$$\bar{x}_k = \sum_{j=1}^{n} x_{jk} \quad , \qquad \text{(A.5)}$$

The *variance* of a random variable is a measure of dispersion among samples; it is defined as the average squared distance of sample values from the sample mean:

$$\text{var}(x_k) = \frac{1}{n-1} \sum_{j=1}^{n} (x_{jk} - \bar{x}_k)^2 \ . \qquad \text{(A.6)}$$

The *covariance* between two variables *k* and *l*, which is a measure of how much two variables change together across data records, is given by

$$\text{cov}(x_k, x_l) = \frac{1}{n-1} \sum_{j=1}^{n} (x_{jk} - \bar{x}_k)(x_{jl} - \bar{x}_l) \quad . \qquad \text{(A.7)}$$

If the two variables tend to vary together, in the sense that *l* tends to be greater than its mean when *k* is greater than its mean, then the covariance between them will be positive; otherwise, it will be negative. A variance is a special case of covariance when the two variables are identical.

Finally, variances and covariances can be combined to give a *correlation coefficient r*:

$$r_{kl} = \frac{\text{cov}(x_k, x_l)}{\sqrt{\text{var}(x_k) \cdot \text{var}(x_l)}} . \qquad \text{(A.8)}$$

Dividing by the square root of the variances "normalizes" the covariances, yielding "correlations" that are all numbers between -1 and 1, which, like covariances, are measures of the relationship between two variables. A correlation coefficient of 0 means these two variables are independent and uncorrelated.

A.1.3. Standardizing the data and the correlation matrix

The statistics are of key importance in PCA and, since PCA is usually written in terms of matrices, understanding how to express variances, covariances, and correlation coefficients in vector notation makes PCA easier to understand.

Data sets used in PCA need to be *standardized* before being used because the magnitudes, and even the units, of variables might be very different; standardization makes all variables dimensionless and scales each variable to some measure of its variation. (For further explanation, see section 2.2.1). A particularly useful way to standardize data is to subtract the mean of each variable (Eq. A.5) and divide by its standard deviation $\sigma_k = \sqrt{\text{var}(x_k)}$ (Eq. A.6):

$$a_{jk} = \frac{x_{jk} - \bar{x}_k}{\sigma_k} \quad . \tag{A.9}$$

In Eq. (A.9) each raw data element x_{jk} is standardized into a_{jk}; the original $n \times m$ data matrix X is thereby transformed into a standardized dataset A:

$$A = \begin{bmatrix} A_1 \\ A_2 \\ \vdots \\ A_j \\ \vdots \\ A_n \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & \cdots & a_k & \cdots & a_m \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1k} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2k} & \cdots & a_{2m} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{j1} & a_{j2} & \cdots & a_{jk} & \cdots & a_{jm} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nk} & \cdots & a_{nm} \end{bmatrix}. \tag{A.10}$$

A remarkable statistical property of standardizing using Eq. (A.9) is that in the resulting dataset A, the correlation coefficient between two variables is the same as their covariance:

$$r_{kl} = \text{cov}(a_k, a_l). \tag{A.11}$$

A vector is often defined as a geometric objective that has both magnitude and direction. The $a_1, a_2, \cdots a_k$ are all column vectors of dimension $n \times 1$; their transposes (Eq. A.3) are all vectors of dimension $1 \times n$. Multiplying a $1 \times n$ matrix by a $n \times 1$ matrix yields a 1×1 matrix, the determinant (det) of which is a scalar whose value is that of the single element in the 1×1 matrix. Using these facts, variances and covariances of the columns of dataset A (i.e., Eqs. A.3 & A.4) can also be written as

$$\text{var}(a_k) = \frac{1}{n-1} \det[a_k{}^T a_k] \quad , \tag{A.12}$$

and

$$\mathrm{cov}(a_k, a_l) = \frac{1}{n-1}\det[a_k{}^T a_l] = \frac{1}{n-1}\det[a_l{}^T a_k]. \tag{A.13}$$

For the data matrix A ($n \times m$), the *covariance matrix* (= *correlation matrix* if the original data matrix X has been standardized using Eq. A.9) is a matrix of size $m \times m$, where each element represents the covariance between two corresponding columns (the row and column number of that element):

$$\mathrm{cov}(A) = \frac{1}{n-1}A^T A = \begin{bmatrix} 1 & \mathrm{cov}(a_1,a_2) & \cdots & \mathrm{cov}(a_1,a_k) & \cdots & \mathrm{cov}(a_1,a_m) \\ \mathrm{cov}(a_2,a_1) & 1 & \cdots & \mathrm{cov}(a_2,a_k) & \cdots & \mathrm{cov}(a_2,a_m) \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \mathrm{cov}(a_k,a_1) & \mathrm{cov}(a_k,a_2) & \cdots & 1 & \cdots & \mathrm{cov}(a_k,a_m) \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \mathrm{cov}(a_m,a_1) & \mathrm{cov}(a_m,a_2) & \cdots & \mathrm{cov}(a_m,a_k) & \cdots & 1 \end{bmatrix} \tag{A.14}$$

The correlation matrix is the basic entity upon which a PCA works.

A1.4. What PCA does

Each sample X_j, as shown in A1.1, is a m-dimensional vector, where m is the number of measurements. X_j could simply be seen as a vector in a m-dimensional basis, each measurement represent one of the axis for this basis, and all of the axes are perpendicular to each other. This basis could be generated as an $m \times m$ identical matrix with all diagonal elements 1 and all the others 0:

$$\begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

Each sample could be seen as a linear combination of the measurements with their unit length of basis vectors:

$$X_j = \begin{bmatrix} x_{j1} & x_{j2} & \cdots & x_{jk} & \cdots & x_{jm} \end{bmatrix} \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} = \begin{bmatrix} x_{j1} & x_{j2} & \cdots & x_{jk} & \cdots & x_{jm} \end{bmatrix}$$

Principal component analysis (PCA) is to find an idea basis to express data so that the variances are maximized on the new axes.

A dot product of two vectors u and v, is a linear combination of their components:

$$\langle u, v \rangle = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n . \tag{A.15}$$

A subset $\{v_1, v_2, \cdots, v_{k,} \cdots v_m\}$ of matrix V is called orthonomal if and only if $\langle v_i, v_i \rangle = 1$ and $\langle v_i, v_j \rangle = 0$ when $i \neq j$. In particular, if V is orthonormal, any two vectors in V are independent and uncorrelated, and each vector in V has unit length. So this orthonomal matrix is a basis for a space.

In a 2-dimentional sample dataset, i.e. Fig. 3.1a, the idea basis has Z_1 and Z_2 as axes as these two axes show the maximum variances among data. PCA is to re-express these data points in the new basis. In this example, the new basis is $\pi / 6$ angles anti-clockwise of the old basis. The basis matrix for the new basis is $\begin{bmatrix} \frac{\sqrt{3}}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{\sqrt{3}}{2} \end{bmatrix}$, which is orthonomal. So the sample that has measurments $\begin{bmatrix} \frac{\sqrt{3}}{2} & \frac{1}{2} \end{bmatrix}$ in the old basis should be re-expressed as $\begin{bmatrix} 1 & 0 \end{bmatrix}$ in the new basis. PCA is to first find out the new basis for these data points, and then re-express these data points in the new basis.

Remember each sample could be seen as a linear combination of the measurements with their unit length of basis vectors, we multiply sample $[\frac{\sqrt{3}}{2} \quad \frac{1}{2}]$ by the new basis:

$$[\frac{\sqrt{3}}{2} \quad \frac{1}{2}] \begin{bmatrix} \frac{\sqrt{3}}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{\sqrt{3}}{2} \end{bmatrix} = [1 \quad 0]$$

and find the result is the same as the re-expressed data point in the new basis. So do the other samples in Fig. 3.1a. The new basis matrix is therefore the rotation of transforming old data into the new basis.

In the m-dimensional measurements, with an assumption that the re-expressed data set (Y) by PCA is a linear combination of the new basis (V), there is

$$Y = AV \tag{A.16}$$

where A is the original dataset which has been standardized. V is not only the basis matrix for the new basis, but also the rotation for transforming A into Y. PCA is

simplified as finding the new basis for the dataset so that maximum variances among dataset are along with axes.

A1.5. How to find the new basis

In the following Eq, $A_1, A_2, A_3, \cdots, A_n$ represent each of the samples with all measurements, $v_1, v_2, v_3, \cdots, v_m$ are the column vectors for the new basis V. Geometrically, V is the rotation which transforms the dataset A into a re-expressed dataset Y in the new basis V.

$$Y = AV = \begin{bmatrix} A_1 \\ A_2 \\ A_3 \\ \vdots \\ A_n \end{bmatrix} \begin{bmatrix} v_1 & v_2 & v_3 & \cdots & v_m \end{bmatrix} = \begin{bmatrix} A_1 v_1 & A_1 v_2 & A_1 v_3 & \cdots & A_1 v_m \\ A_2 v_1 & A_2 v_2 & A_2 v_3 & \cdots & A_2 v_m \\ A_3 v_1 & A_3 v_2 & A_3 v_3 & \cdots & A_3 v_m \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ A_n v_1 & A_n v_2 & A_n v_3 & \cdots & A_n v_m \end{bmatrix} \quad \text{(A.17)}$$

$$Cov(Y) = \frac{1}{n-1} Y^T Y \quad \text{(A.18)}$$

The covariance $Cov(Y)$ is a square symmetric matrix ($m \times m$), the diagonal terms of $Cov(Y)$ are the variance of particular measurement, and the off-diagonal terms of $Cov(Y)$ are the covariance between two measurements. To make the variance for the vectors maximized, the optimized $Cov(Y)$ should be diagonal with all of off-diagonal terms are zero. Therefore, PCA is to find an orthonormal matrix V to make $Cov(Y)$ a diagonal matrix.

$$Cov(Y) = \frac{1}{n-1} Y^T Y = \frac{1}{n-1} (A^T V^T A V) = V^T \frac{A^T A}{n-1} V = V^T Cov(A) V \quad \text{(A.19)}$$

Notice that $Cov(A)$ is the covariance of A, and is a symmetric matrix with size $m \times m$. There are always two matrixes U and D for a symmetric matrix, U is an orthonomal and D is diagonal matrix, and they satisfy the following Eq:

$$Cov(A)U = UD, \quad \text{(A.20)}$$

where D is called the eigenvalue and U the eigenvector of $Cov(A)$.

$$U' Cov(A) U = U' U D. \quad \text{(A.21)}$$

Since eigenvectors U is orthonomal, the dot product of any two columns is zero, and the dot product of one particular column with itself is one. $U'U$ is proved to be an identity matrix:

$$U'U = I$$

$$U'Cov(A)U = U'UD = D \tag{A.22}$$

After a comparison between Eqs. A.19 and A.22, it follows that U and D are the solutions for V and $Cov(Y)$. To diagonalize $Cov(A)$, therefore, is simply to find eigenvectors V for $Cov(A)$.

The performing of PCA for an original dataset X could be simplified as: (1) standardize dataset X into A; (2) find the eigenvector V for $Cov(A)$; (3) calculate the re-expressed dataset (principal component) by $Y = AV$. Several properties are summarized for this PCA eigenvector-eigenvalue calculation: (1) The eigenvectors V contain m solutions: $v_1, v_2, v_3, \cdots, v_m$. It is an orthonormal matrix, *i.e.* any two columns are vertical to each other and, all columns ($v_1, v_2, \cdots, v_m$) have magnitude 1; (2) The re-expressed dataset Y is orthogonal, with any two columns uncorrelated to each other; (3) The diagonal elements of $Cov(Y)$ are eigenvalues, the exact amounts of the variance accounted for by corresponding vectors in the new dataset Y.

A1.6. what are site scores and loadings in the new basis

Site scores (Y) are the positions of the samples in the new space, i.e., the re-expressed dataset, and are calculated by $Y = AV$. A is the standardized dataset, V is the matrix for the new basis. Site scores are also called the principal components of PCA.

Loadings (V) are the orthonomal basis for the re-expressed data. The values of loadings are in the matrix of eigenvector. For example, the loadings of each variable on the first PC are shown by the first eigenvector (the first column in the matrix V). Loadings give correlations of the re-expressed dataset (Y) with the original data (A). The loadings on the first column of eigenvector give the correlations of the first PC with all of the variables in the original data A, i.e., all of the column vectors in A. The higher loading a variable has along the first PC, the more correlated the first PC is with A on that specific variable.

Appendix B: How to find a single direction instead of using any PC that representing a specific character well among dataset in a 3D PCA result?

In a PCA, each single PC is a linear combination of all the variables, and the first PC (PC1), as the most important one among PC's, explains the maximum variance of the data. This, however, does not necessarily mean that PC1 or any other single PC could represent a specific geochemical process that is expected. For example, during sample degradation, the mole percentages of amino acids (BALA, GABA) usually become more abundant while those of the fresh ones (SER, GLU, ASP, GLY, THR, *et al.*) decrease. Samples might be more distinct by their sources instead of degradation, and these amino acids that mostly represent degradation information did not show maximum loadings along PC1. As discussed in text part, the degradation trajectories of OM are not even a simple straight line in a 3D coordination. **In this appendix, we are trying to find an approximate single direction that largely differentiates samples by their degradation character.** The finding of the approximate degradation direction is useful for a straight visualizing the degradation status among samples. There must be some angle between PC1 and degradation direction defined by the degradation biomarker, GABA and BALA in a 3D coordinate. Here we introduce a new axis by rotating the PC1, so that the representative amino acids would have highest correlation with the new degradation direction (*i.e.*, these amino acids have highest projections on the degradation index) (Yunker et al., 1995).

The first three PCs are remained in the 3D figure for discussion. PC1, PC2, and PC3 are along with axes x, y, and z, respectively. As shown in Fig. B.1a, the degradation direction for amino acids is approximately from SER to GABA, which is obviously not aligned along PC1. Assume the projection of SER-GABA direction on x-y geometric plane surface has a clockwise angle α with x axis; the projection of SER-GABA direction on x-z geometric plane surface has a clockwise angle β with x axis. We now change the orientation of axes by first z axis anti-clockwise rotating an angle α (alpha), so that SER-GABA are right on x-z geometric plane surface; and then by y axis anti-clockwise rotating an angle β, so that SER-GABA are right aligned along the new x axis and their loadings on it are maximized (Fig. B.1b). The new loading for samples, $[x' \quad y' \quad z']$, is calculated by the product of old loading $[x \quad y \quad z]$, and a matrix representing rotating:

$$[x' \quad y' \quad z'] = [x \quad y \quad z] \times \begin{bmatrix} \cos\alpha & -\sin\alpha & 0 \\ \sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} \cos\beta & 0 & -\sin\beta \\ 0 & 1 & 0 \\ \sin\beta & 0 & \cos\beta \end{bmatrix} \qquad \text{(B.1)}$$

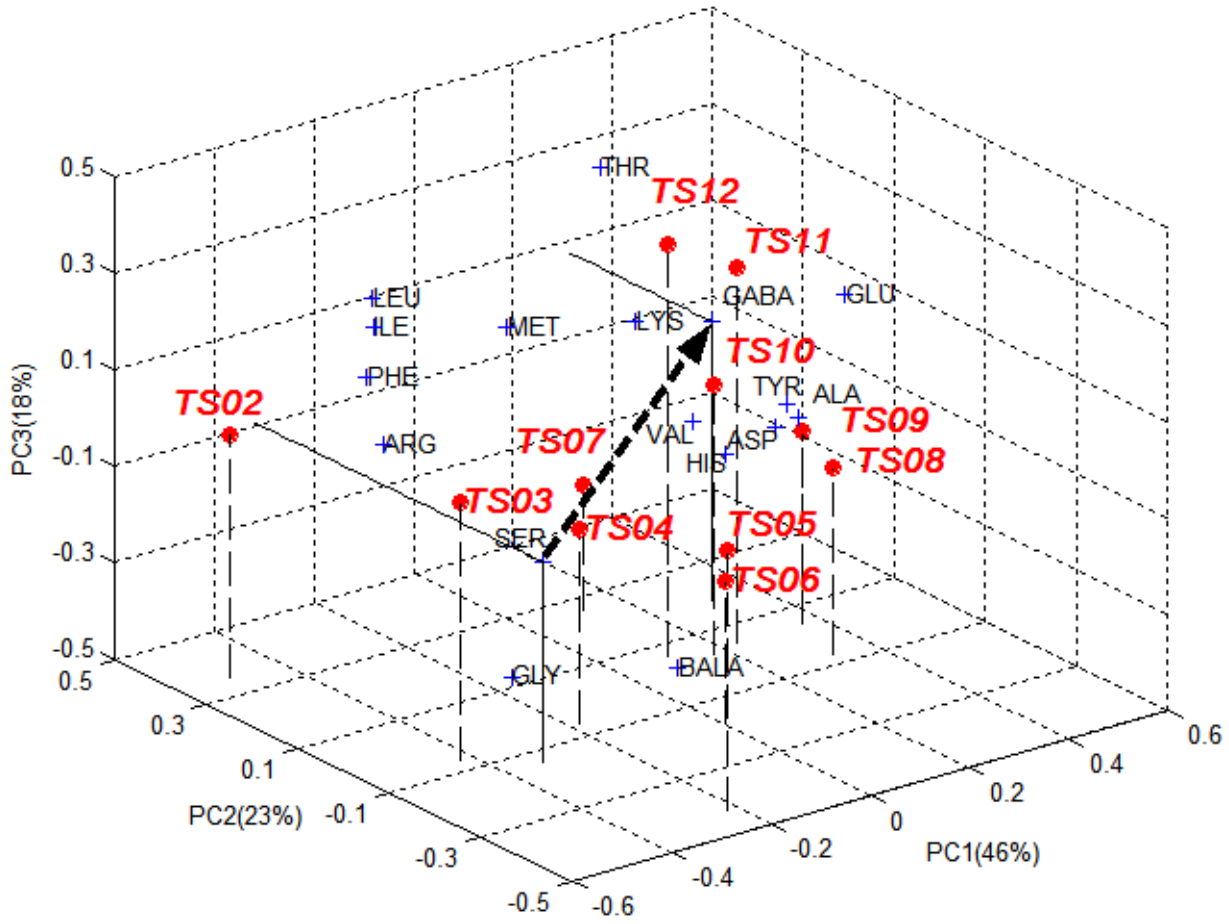
Here, *x*, y and *z* are single matrix for PC1, PC2, and PC3, respectively. The angle α and β in this study are calculated by the loadings of SER and GABA in un-rotated 3D PCA. The projection of each sample on the new x axis is its degradation index. The advantage of rotation after PCA is that, samples are well separated through x

axis, which is explained only by degradation character. This is very useful if a specific index of either degradation or source needs to be quantified.

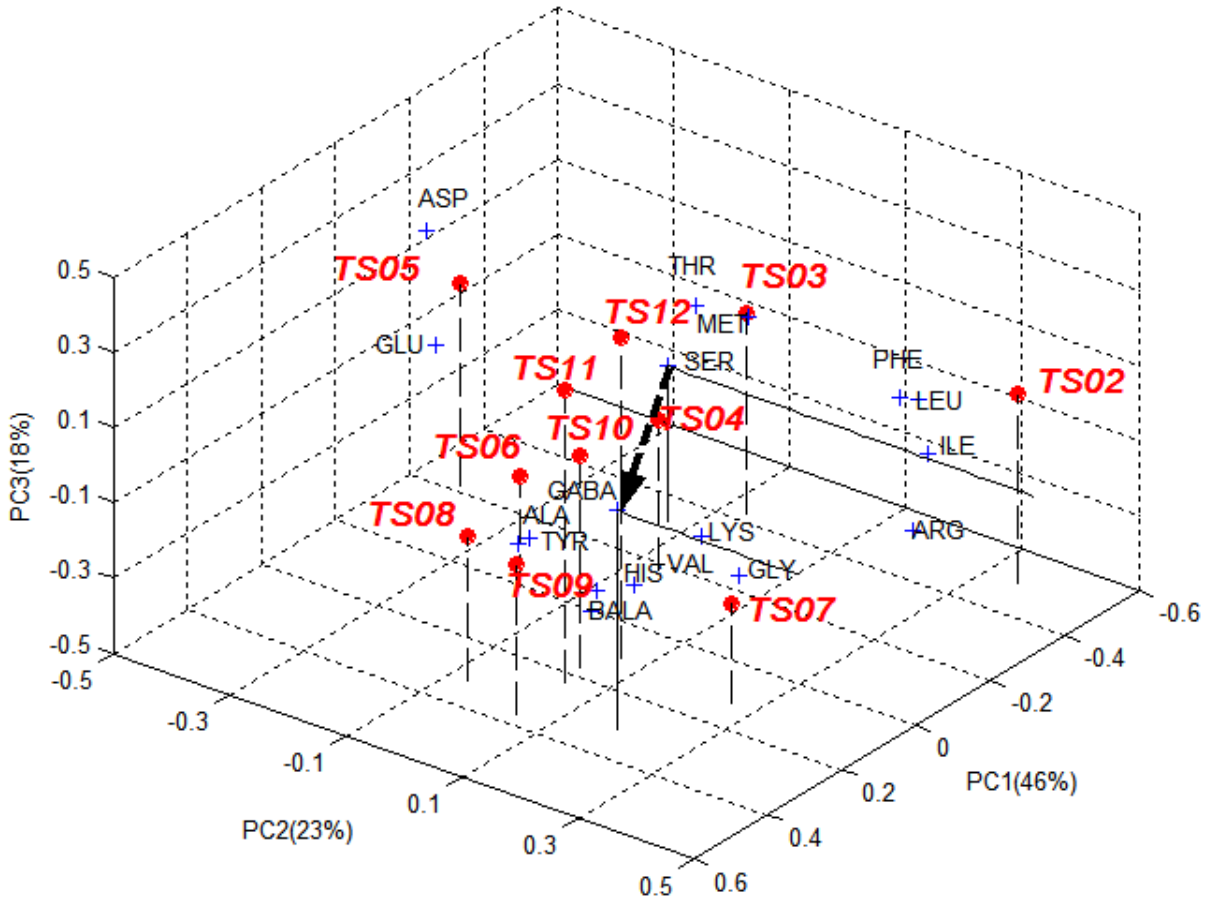
Fig. B.1a also shows that PC1, PC2 and PC3 account for 46%, 23% and 18% of the variance. Since it is an orthogonal rotation, the rotated result in Fig. B.1b still explains 87% of the variance in total, but will reallocate them in the new axes (40% for x axis, 27% for y axis, and 20% for z axis).

For finding the approximate single degradation direction, we could use the middle loading of BALA and GABA for degradation product. We also could consider ASP, GLU, GLY, THR, *et al.* as fresh phytoplankton biomarkers, which would obviously increase the complexity of loading degradation direction.

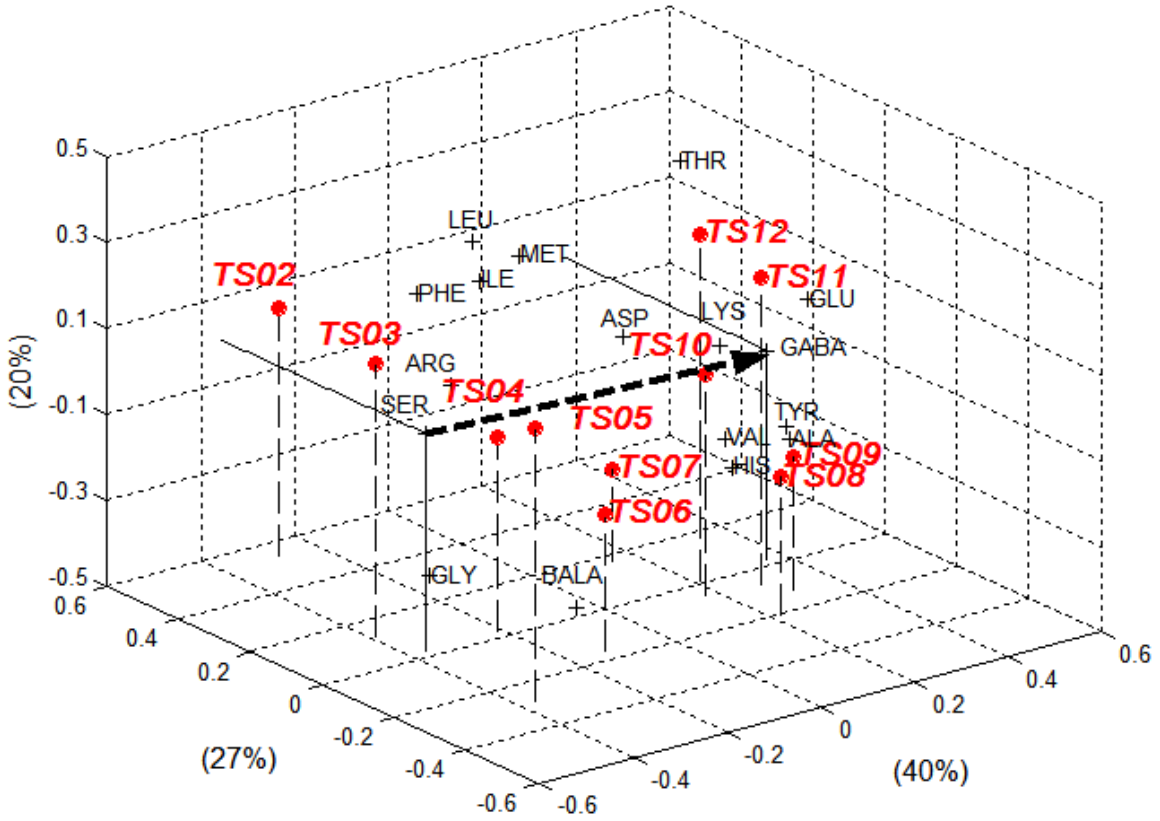
(a-1)


THR
TS12
TS11
GABA
GLU
LEU
ILE
MET
LYS
TS10
PHE
TYR
ALA
TS02
TS09
TS07
VAL
ASP
ARG
HIS
TS08
TS03
TS04
TS05
SER
TS06
GLY
BALA
PC3(18%)
PC2(23%)
PC1(46%)
0.5
0.3
0.1
-0.1
-0.3
-0.5
0.6
0.4
0.2
0
-0.2
-0.4
-0.6

(a-2)


ASP
TS05
THR
TS03
TS12
MET
GLU
TS11
SER
PHE
TS02
LEU
TS10
TS04
TS06
ILE
GABA
TS08
ALA
LYS
ARG
TYR
VAL
GLY
HIS
TS09
BALA
TS07
PC3(18%)
PC2(23%)
PC1(46%)
0.5
0.3
0.1
-0.1
-0.3
-0.5
-0.5
-0.3
-0.1
0.1
0.3
0.5
0.6
0.4
0.2
0
-0.2
-0.4
-0.6

(b-1)


0.5
0.3
0.1
-0.1
-0.3
-0.5
(20%)
THR
LEU
MET
ILE
PHE
TS12
TS11
GLU
LYS
ASP
TS10
GABA
TS02
TS03
ARG
SER
TS04
TS05
TYR
VAL
ALA
TS09
TS07
HIS
TS08
TS06
GLY
BALA
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
(27%)
(40%)

(b-2)

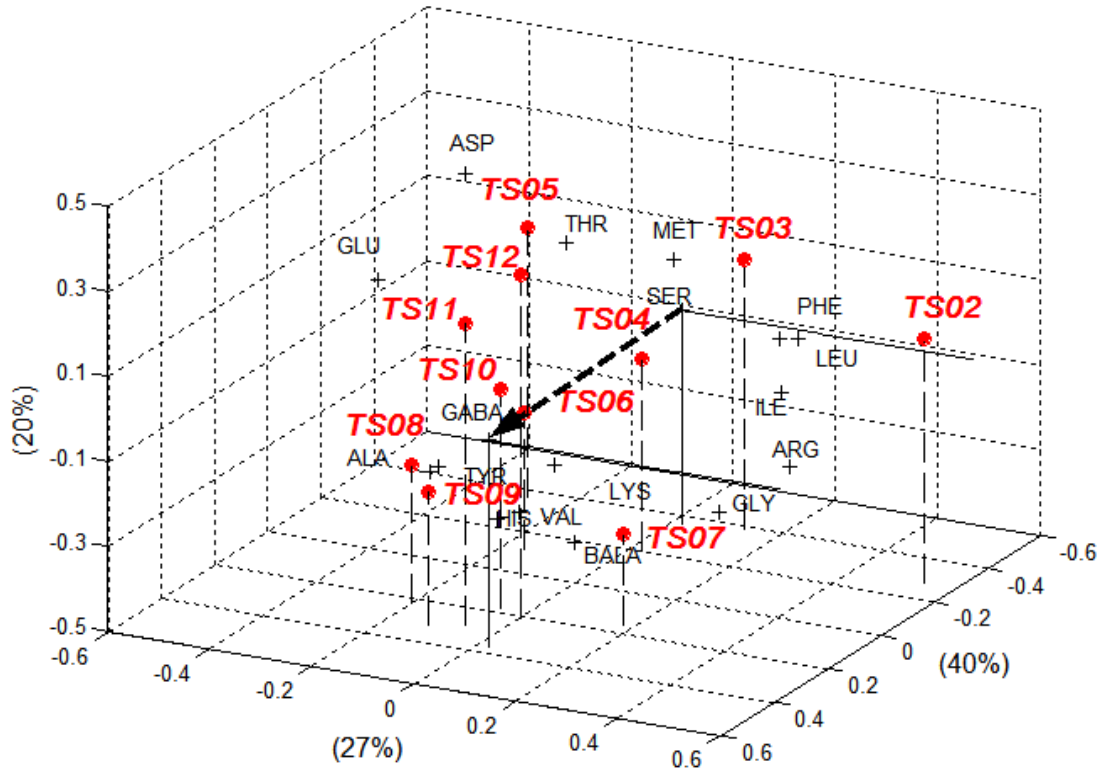


Figure B.1 A comparison on (a) the 3D plots (with two view angles, a-1 & a-2) of the first 3 principal components for amino acids dataset of time series sediment trap samples in March to May 2003, DYFAMED site Mediterranean Sea with (b) the one that has been rotated (with two view angles, b-1 & b-2) so that amino acids SER and GABA have biggest correlations with the new first axis, and are loading along that axis.

Appendix C: 3D PCA plot with the first three PCs

```
% PCA
% standardize data: demean and divided by standard deviation
[n,m]=size(X);
X=X-ones(n,1)*mean(X);
X=X./(ones(n,1)*std(X,1));

% find covariance for standardized data
c=cov(X);

% ftn svd to find eigenvector and eigenvalue for covariance c
[v,d,vt]=svd(c);

% Y is site core
Y=X*v;
D=diag(d)/sum(diag(d));

figure(1)
num=[1:1:n]';
label=num2str(num);
plot3(Y(:,1),Y(:,2),Y(:,3),'ro','MarkerSize',6,'MarkerFaceColor','r');
text(Y(:,1)+0.005,Y(:,2),Y(:,3),label,'FontSize',
16,'FontWeight','demi','FontAngle','italic');
xlabel('PC1');ylabel('PC2');
hold on;

% v is loading
Plot3(v(:,1),v(:,2),v(:,3);'+');
num=[1:1:m]';
label=num2str(num);
text(v(:,1)+0.003,v(:,2),v(:,3),label);
```

Appendix D: The rotation of 3D PCA plot

```
[n,m]=size(X);

% standardize data: demean and divided by standard deviation
X=X-ones(n,1)*mean(X);
X=X./(ones(n,1)*std(X,1));

% find covariance for standardized data
c=cov(X);

% ftn svd to find eigenvector and eigenvalue for covariance c
 [v,d,vt]=svd(c);

Y=X*v;
D=diag(d)/sum(diag(d));
```

$$T12=\begin{bmatrix} \cos\alpha & -\sin\alpha & 0 \\ \sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$T13=\begin{bmatrix} \cos\beta & 0 & -\sin\beta \\ 0 & 1 & 0 \\ \sin\beta & 0 & \cos\beta \end{bmatrix}$$

```
% first z axis anti-clockwise rotating an angle α (alpha), then by y axis anti-clockwise
% rotating an angle β
vv=[v(:,1)  v(:,2) v(:,3)]*T12*T13;
YY=[Y(:,1)  Y(:,2)  Y(:,3)]*T12*T13;

figure(1)
num=[1:1:n]';
label=num2str(num);
plot3(YY(:,1),YY(:,2),YY(:,3),'ro','MarkerSize',6,'MarkerFaceColor','r');
text(YY(:,1)+0.005,YY(:,2),YY(:,3),label,'FontSize',
16,'FontWeight','demi','FontAngle','italic'); hold on;

% vv is loading on new axes; YY is site score on new axes
Plot3(vv(:,1),vv(:,2),vv(:,3),'+');
text(vv(:,1)+0.003,vv(:,2),vv(:,3),label);
var(YY)/sum(diag(d))
```

Appendix E. Abbreviation table, cited from Goutx et al., (2007)

Biogenic silica	$BSiO_2$	Total sugars	TCHO
Organic carbon	OC	Fucose	fuc
Particulate organic carbon	POC	Rhamnose	rha
		Arabinose	ara
Total lipids*	TLip	Galactosamine	gal-am
hydrocarbons	HC	Glucosamine	glucosa
ketone	KET	Galactose	galact
wax esters	WE	Glucose	glucose
triacylglycerols	TG	Mannose	man
free fatty acids	FFA	Xylose	xyl
alcohols	ALC	Fructose	fru
sterols	ST	Ribose	rib
1,3-diglycerides	1,3 DG		
1,2-diglycerides	1,2 DG	Lipid biomarkers	
monoglycerides	MG	tetradecanoic acid	14 : 0
chloroplast lipids	CL	*iso*-pentadecanoic acid	i-15 : 0
pigments	PIG	*anteiso*-pentadecanoic acid	a-15 : 0
monogalactosyl-diglycerides	MGDG	Pentadecanoic acid	15 : 0
digalactosyl-diglycerides	DGDG	hexadecenoic acid	16 : 1
phosphatidylglycerides	PG	hexadecanoic acid	16 : 0
phosphatidylethanolamines	PE	octadecatetraenoic acid	18 : 4
phosphatidylcholines	PC	octadecadienoic acid	18 : 2
metabolites	METAB	oleic acid	18 : 1ω9
Total pigments	TPig	*cis*-vaccenic acid	18 : 1ω7
chlorophyll *a*	Chl *a*	octadecanoic acid	18 : 0
chlorophyll *b*	Chl *b*	eicosapentaenoic acid	20 : 5
phaeophorbide *a*	phide	eicosenoic acid	20 : 1
pyrophaeophorbide *a*	pyrophide	eicosanoic acid	20 : 0
phaeophytin *a*	phytin	docosahexaenoic acid	22 : 6
fucoxanthin	fuco	docosanoic acid	22 : 0
		Hexadecanol	16ROH
Total hydrolyzed amino acids	THAA	Octadecanol	18ROH
aspartic acid	ASP	Phytol	phytol
glutamic acid	GLU	cholesta-5,22-dien-3β-ol	27(5,22)
histidine	HIS	cholest-22-en-3β-ol	27(22)
serine	SER	cholest-5-en-3β-ol	27(5)
arginine	ARG	cholestan-3β-ol	27(0)
glycine	GLY	24-methylcholesta-5,22-dien-3β-ol	28(5,22)
threonine	THR	24-methylcholesta-5,24(28)-dien-3β-ol	28(5,24/28)
β-alanine	BALA	24-ethylcholesta-5,22-dien-3β-ol	29(5,22)
alanine	ALA	24-ethylcholesta-5-en-3β-ol	29(5)
tyrosine	TYR	4,23,24-trimethylcholest-22-en-3β-ol	30(22)
γ-aminobutyric acid	GABA	C_{37}-C_{39}alkenones	alken
methionine	MET		
valine	VAL		
phenylalanine	PHE		
isoleucine	ILE		
leucine	LEU		
lysine	LYS		

* TLip is the sum of all lipid classes except HC, which can be contaminated by anthropogenic compounds.