

Modeling Geochemical datasets for Source Apportionment: Comparison of Least square regression and Inversion approaches

Gyana Ranjan Tripathy^{1,3,*} and Anirban Das²

¹AIRIE Program, Department of Geosciences, Colorado state university, Fort Collins, USA

²Pandit Deendayal Petroleum University, Gandhinagar, India

³National Institute of Oceanography, Dona Paula, Goa-403004, India

*Corresponding Author: grtripathy@gmail.com

Running Title: *Modeling geochemical datasets for source apportionment*

Abstract

Mathematical modeling of geochemical datasets finds frequent applications in Earth Sciences, particularly in areas of source apportionment and provenance studies. In this work, source apportionment modeling have been considered based on two commonly used methods, the Least Square Regression (LSR) and Inverse Modeling (IM), to determine the contributions of (i) solutes from different sources to *global river* water, and (ii) various rocks to a glacial till. The purpose of this exercise is to compare the results from the two mathematical methods, infer their merits and drawbacks and indicate approaches to enhance their reliability.

The application of the LSR and IM approaches to determine the source contributions to *global river* water using the same *a-priori* end member compositions yielded divergent results; the LSR analysis giving impossibly negative values of Na contribution from one of the sources (evaporites), in contrast to the IM approach which yield reliable estimates of source contributions, and a set of *a-posteriori* source compositions and associated uncertainties. Interestingly, the use of the *a-posteriori* composition derived from the IM approach in the LSR analysis as an input for end-member composition gave source contributions that were consistent with those derived from IM. Calculations based on the IM show that $46 \pm 8\%$ of Na in *global river* is derived from silicate weathering, consistent with some of the earlier reported estimates.

In case of the glacial till, the source contributions based on both the approaches were similar, however even in this case better agreement between the two approaches is obtained when the *a-posteriori* composition data of end members derived from the IM is used as input in the LSR model. These comparisons demonstrate that the IM is better suited for source apportionment studies among the two models, as it requires only rough estimates of end member composition, unlike the LSR that needs source composition to be better constrained. In addition, the IM also provides uncertainties in the source contributions and best estimates of their composition.

Keywords: Least Square Regression; Inverse Modeling; Source Apportionment; Geochemistry.

1. Introduction

Computational simulation methods have become important tools over the last few decades in providing quantitative information for many Earth science problems (Zhao et al., 2009 and references therein). For instance, the computational simulation method has been used to solve not only a wide range of ore forming problems within the upper crust of the Earth (Hobbs et al., 2000; Gow et al., 2002; Ord et al., 2002; Sorjonen-Ward and Zhang, 2002; Zhang et al., 2003, 2008; Liu et al., 2005, 2008, 2011; Lin et al., 2006; Schmidt Mumm et al., 2010; Ju et al., 2011), but also a broad range of other types of Earth science problems (Lin et al., 2003, 2008, 2009; Yan et al., 2003; Xing and Makinouchi, 2008; Zhao et al., 2008a, 2010; Zhang et al., 2011). Since three basic models, i.e. geological, mathematical and numerical simulation models, are involved in the computational simulation method, mathematical modeling plays an indispensable role in the computational simulation of Earth science problems (Zhao et al., 2008b, Zhao, 2009 and references therein).

Geochemical approaches, as an important part of the computational geoscience discipline (Zhao et al., 2009 and references therein), have found extensive applications in the field of Earth sciences to study and infer about various geological processes. Chemical, mineralogical and isotopic compositions of geological samples hold clues to the sources contributing to them and their mixing proportions and thus, provide useful insights on the processes and factors responsible for their mobilization and sequestration. Source-identification and source-apportionment of elements in different earth system components is, however, not straight forward; they rely largely on the mathematical modeling of compositional data. Many mathematical models are available to apportion source contributions in a mixture. These include the principal component analysis (PCA), positive matrix factorization (PMF), least square regression (LSR), forward and inverse modeling (IM) (Bryan et al., 1969; Beus and Grigorian, 1977; Allegre et al., 1983; Negrel et al., 1993; Paatero, 1997; Weltje, 1997; Krishnaswami et al., 1999; Makinen and Gustavsson, 1999; Bickle et al., 2005; Sohn, 2005; Wu et al., 2005; Huang and Conte, 2009; Tripathy et al., 2014). Among these, the PCA and factor analysis are often used to quantify source contribution to mixtures such as river water, sediments and aerosols. The merit of these two approaches is that knowledge on the number of sources and their compositions is not a prerequisite for their application to derive information on the source contribution to a mixture. The applications of both PCA and factor analysis, however, are limited as at times these approaches can provide non-positive results for source contributions (Larsen and Baker, 2003; Sofowote et al., 2008). This problem of the PCA and factor analysis methods is addressed in the PMF model, which has recently gained attention in

aerosol studies (Paatero and Tapper, 1993; Paatero, 1997; Kim et al., 2004; Sudheer and Rengarajan, 2012). Among the other mathematical models, least square regression has been widely used in geochemical problems (Bryan et al., 1969; Wright and Doherty, 1970; Le Maitre, 1979). For example, Makinen and Gustavsson (1999) used Chebyshev's series solution of linear programming to decipher the relative contribution of amphibolites, granitoids, quartzite and sand to a till (mixture). Quantification of source contribution to mixtures (river water) has also been carried out either through the forward model (Krishnaswami et al., 1999; Das et al., 2005; Moon et al., 2007; Tripathy et al., 2010) using a suite of mass balance equations and pre-assigned source composition or through inverse modeling, based on the composition of mixture and approximate information on the composition of their possible sources. Allegre et al. (1983) successfully used the inverse model to derive the contribution from various reservoirs to basalts to explain their chemical and isotopic composition. Subsequently, the inverse model has found extensive application to apportion the chemical and isotopic composition of rivers among its sources that include weathering of major lithologies present in their basins and atmospheric deposition (Negrel et al., 1993; Millot et al., 2003; Wu et al., 2005; Moon et al., 2007; Tripathy and Singh, 2010). More recently, the IM has also been extensively used to quantify various oceanic processes (Rahaman and Singh, 2012; Singh et al., 2012).

These models although used widely for source apportionment and quantification, only a few studies have attempted to inter-compare the results yielded from various models. Such exercises are needed to enhance the confidence in the application of these models and interpret their quantitative results to address source apportionment and their geochemical significance. Morandi et al. (1991), based on inter-comparison of results from two different models (*viz.* modified version of factor analysis/multiple regression and regression on absolute principal components), observed that although the results showed an overall consistency, they also had some discernible differences. This led the authors to suggest the need for using more than one model to accurately quantify source contribution from datasets. More recently, the results of Tripathy and Singh (2010) on the inter-comparison of source contributions derived using the forward and inverse models of chemical and Sr isotopic composition of the Ganga headwaters showed that the results exhibit statistically significant differences. These observations highlight the importance of and need for inter-comparison of results from different mathematical approaches to obtain robust data on source contribution and inferences on associated geochemical processes. The present study is an attempt in this direction. It inter-compares the results on source apportionment obtained using two commonly employed mathematical techniques: the least

square regression and the inverse model. The study has been conducted on two mixtures: the *global* river water and a glacial till. The results have led to better understanding of the inherent merits and limitations of these models, as discussed in the paper.

2. Methods

2.1. Least Square approximation using QR decomposition

The source-apportionment models/programs are based on the mass balance approach that involves the formulation of suitable equations for the budget of each element in the mixture and solving the set of equations to determine the contribution from various sources. In the LSR method, the known parameters are the precisely constrained elemental abundances (or their ratios) in the mixture (*e.g.* river water or sediments) and in their various possible sources (end members). The unknown parameters, which need to be quantified, are the relative contribution from each source to the mixture.

A set of linear equations totaling the number of unknowns can provide a unique solution. However, geochemical studies focusing on source apportionment often have more number of equations (*i.e.*, number of geochemical parameters) than the number of unknowns (*i.e.*, relative contribution from different sources). Equation (1) presents such an over-determined linear system, where the measured chemical data ($b_i = 1$ to m) of the system are related to the chemical composition of its possible sources/end-members ($a_{ij}; i = 1$ to $n; j = 1$ to m) and their relative contributions to the mixture ($x_i; i = 1$ to m). In this case, the number of equations ($n+1$) is more compared to the number of unknowns (m).

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \dots + a_{1m}x_m = b_1 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2m}x_m = b_2 \\ \dots\dots\dots \\ \dots\dots\dots \\ a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nm}x_m = b_n \\ x_1 + x_2 + \dots + x_m = 1 \end{cases} \quad (1)$$

These over-determined set of mass balance equations does not have a unique solution. However, the use of least square regression approach can provide a ‘best estimate’ solution for the contribution from each end member to the mixture ($x_i; i=1, m$ of equation (1)) which can satisfy equation (1) with the least residual.

As a first step to find the best-possible solution using the LSR, equation (1) can be rewritten as an equation of matrices, i.e.,

$$AX = B \quad (2)$$

where B and A are matrices containing chemical composition (n elements) of the mixture and all its possible sources (m), respectively and X lists the relative contribution from each of the sources. In order to solve equation (2) using the LSR approach, the matrices A and B need to be known accurately. To achieve this, least square approximation using the QR factorization of equation (2) was adopted in this study to find the solution of X . The QR decomposition of the matrix A is its factorization into an orthogonal matrix (Q) and a triangular matrix (R), i.e.

$$A = QR \quad (3)$$

where, Q is an orthogonal matrix, i.e. $Q^T Q = I$. I is the identity matrix and superscript ‘ T ’ stands for the transpose of the matrix. The QR factorization transforms a linear over-determined least square problem into a well-defined triangular system.

Mathematically, the QR decomposition of equation (2) can be understood as follows,

$$\begin{aligned} AX &= B \\ \Rightarrow A^T AX &= A^T B \\ \Rightarrow R^T Q^T QRX &= R^T Q^T B && \text{(using equation (3))} \\ \Rightarrow R^T RX &= R^T Q^T B && \text{(as, } Q^T Q = I) \\ \Rightarrow RX &= Q^T B \\ \Rightarrow X &= R^{-1} \cdot Q^T \cdot B && (4) \end{aligned}$$

To apply the QR decomposition of matrices to find solution of linear equations relating the geochemical datasets, a code using the subroutine of *L2QRSolve* from the library of *Mathematica 7*, is used in this study (cf. Appendix).

2.2. Inverse Model

The inverse model has been successfully used to apportion source contributions to various geological mixtures using data on their chemical composition. Mathematical description of the IM used in this study is presented briefly below. For more details reference is made to Tarantola (2005).

In geochemical studies, the observed geochemical datasets (d_{obs}) and their associated covariance matrix (C_D) form a probability density ($\rho_D(d)$) in the data (D) space which can be written as,

$$\rho_D(d) = const \times \exp\left(-\frac{1}{2}(d - d_{obs})^T C_D^{-1}(d - d_{obs})\right) \quad (5)$$

where superscripts ‘ T ’ and ‘ -1 ’ stand for transpose and inverse of the matrix respectively.

Similar to the observed datasets, a probability density ($\rho_M(m)$) for the model (M) space with the model parameters (m_{priori}) and their associated uncertainties (C_M) can be written as,

$$\rho_M(m) = const \times \exp\left(-\frac{1}{2}(m - m_{priori})^T C_M^{-1}(m - m_{priori})\right) \quad (6)$$

The observed data for the geochemical system is related to the model parameters by a set of equations (same as equations 1 and 2 as those for the LSR) which are of the form:

$$d = g(m) \quad (7)$$

The observed data follow Gaussian distribution, whereas the model parameters are log-normal in nature; this makes the equation (7) a non-linear system. Combination of equations (5)-(7) forms *a-posteriori* probability density ($\sigma_M(m)$). The iterative method in the IM aims to find the maximum likelihood for data and model parameters within their covariances that maximizes the $\sigma_M(m)$. The Quasi-Newton optimization method for non-linear relations, which relies on the iterative algorithm that starts the iteration from the *a-priori* information on the model parameters, has been used in this study. The computational code for the IM used in this study is the same as that applied successfully by Tripathy and Singh (2010) for apportioning the sources of solutes of the Ganga headwaters.

3. Geochemical Perspectives

Prior to mathematical modeling of datasets involving geochemical mixtures (e.g., river water, glacial till etc.) to infer about source contributions, it is essential to understand the processes and factors that regulate their chemical composition. Such information for river water (Garrels and Mackenzie,

1971; Berner and Berner, 1996; Drever, 1997; Brantley and White, 2009; Maher, 2010) and sedimentary systems (Nesbitt and Markovics, 1997; Gaillardet et al., 1999a; Das and Krishnaswami, 2006; Tripathy et al., 2011; Lupker et al., 2012) have been detailed in various earlier studies. Hence, they are only briefly discussed in this section.

The chemical composition of river water is controlled by relative supply of solutes from rain water, and chemical weathering of bedrocks present in their basins. In recent years, the chemistry of many rivers is also impacted by anthropogenic contributions. This effect, however, is not considered in subsequent discussion. Weathering of bedrocks present in drainage basins, composed mainly of silicates, carbonates and evaporites, dominates the water chemistry (Berner and Berner, 1996). Among these, carbonates and evaporites undergo congruent weathering and have faster dissolution kinetics than the silicates. On the contrary, chemical weathering of silicate rocks generally follows incongruent weathering and therefore, forms soils and secondary minerals. These secondary minerals formed during chemical weathering contribute to the formation and development of soils. Similar to river water, the sediments hold clues to their sources (provenances/lithologies/rock types) and transport processes. For example, ocean sediment is a mixture of materials carried by different rivers, and biogenic materials produced in the sea and atmospheric deposition. Similarly, during glacier grinding, different rock types from the basin are broken down, and homogenized to form a till, which again stores quantitative information about source materials. Field observations and mineralogical studies of the mixtures can provide important information about their potential sources. However, quantitative data on their mixing proportions have to depend on modeling of their chemical and isotopic composition. The success of these approaches prerequisites distinctly different chemical and isotope compositions of the potential sources.

4. Application of the LSR and inverse models

In this section, applications of the LSR and IM to apportion sources of elements in the *global river* water and a glacial till are discussed. In these case studies the results obtained from both the models are compared using an error-weighted statistical approach. In case of the glacial till, rough estimates of uncertainties associated with the results from the LSR approach have also been inferred, for which information is sparse.

4.1. Case study 1: *Global river*

Chemical weathering in river basins plays an important role in regulating the global carbon budget. Among the various lithologies in a river basin, weathering of silicates is a long term sink for atmospheric CO₂; this makes the determination of abundance of silicate derived cations in rivers an important exercise (Walker et al., 1981; Berner et al., 1983). Several geochemical studies have been reported from small to large river basins to quantify the CO₂ drawdown by silicate weathering. This study attempts to apportion the source contributions that regulate the *global river* water chemistry based on the commonly used least square regression and inverse modeling approaches. The goal of this exercise is to inter-compare the results derived from the two models and assess their relative merits and limitations.

The discharge-weighted chemical composition of the *global river* is computed in this study using the available data on the dissolved chemical composition of 60 major rivers (Gaillardet et al. 1999b). In order to minimize the effect of anthropogenic contributions on the river chemistry, rivers with Na⁺ < Cl⁻ have been excluded from the dataset. This implicitly assumes that rivers with Cl⁻ > Na⁺ is anthropogenically impacted. Table 1A lists the chemical composition of the pristine *global river* estimated in this work. The chemical data for the rivers used in this study constitutes ~46% of the global water discharge and 41% of the global drainage area. The average *global river* water chemistry reported in this study mostly consistent with that reported earlier (Meybeck, 2003). Interestingly, the estimated SO₄⁻² concentration (102 μM; Table 1A) for the *global river* in this study is lower compared to that reported recently (156 μM; Miller et al., 2011). The relatively higher SO₄⁻² concentration reported by Miller et al. (2011) is attributable to influence of recent anthropogenic input to rivers.

4.1.1. Source Apportionment

The dissolved constituents in the *global river* are assumed to be derived mainly from four sources: rain, weathering of silicates, carbonates and evaporites. The mass balance equations employed in the model to quantify the relative contribution of sources to the *global river* are based on the Na-normalized elemental ratios of the *global river* water and its possible sources. These equations are presented in the Appendix. Application of the inverse model to apportion the source contributions to the *global river* water chemistry requires rough estimates (*a-priori* information) of the chemical composition of all its potential sources. The *a-priori* chemical composition of the four sources (*i.e.*, rain, silicates, carbonates and evaporites) used in this study are listed in Table 1B (Gaillardet et al., 1999b;

Millot et al., 2003; Moon et al., 2007). Inversion modeling of these datasets indicates that the supply of dissolved Na to the *global river* is dominated by weathering of silicates ($46 \pm 8\%$) and rain ($36 \pm 10\%$) with sub-ordinate contributions from evaporites ($16 \pm 10\%$) and carbonates ($3 \pm 1\%$).

Reliability of the results obtained from the IM depends on the ability of the optimization (Quasi-Newton) method to converge to the best-possible combination of data and model parameters that can solve equation (1) with least residual. The robustness of the IM, therefore depends on how well it converges to nearly the same *a-posteriori* values, independent of the *a-priori* inputs provided for the model parameters. Tripathy and Singh (2010) assessed this by making large (few orders of magnitude) changes in *a-priori* values of elemental ratios for the sources. Their results showed that the *a-posteriori* Ca/Na and Mg/Na ratios for carbonates obtained from inverse modeling of geochemical datasets of the Ganga headwater converge to similar values despite orders of magnitude (45-3000 for Ca/Na and 25-1500 for Mg/Na) differences in their *a-priori* inputs. Further, these authors (Tripathy and Singh, 2010) authenticated the robustness of the IM model calculations by comparing the results obtained from their model with that of an earlier study (Millot et al., 2003). In addition to these results, another sensitivity test for the IM is also carried out in this study. The effectiveness of the model to converge to nearly the same fractional contribution of sodium from silicates ($f_{\text{sil}}(\text{Na})$) was tested by changing the *a-priori* input by about an order of magnitude (Fig. 1). The results of this test converged to nearly identical values, within a narrow range of 0.45 ± 0.01 . Further, the IM converges to the same fractional contributions from various sources to the dissolved Na of the *global river*, even though the *a-priori* Sr isotopic composition for silicates in the IM is changed widely from 0.74 ± 0.04 to 0.721 ± 0.01 (global silicate rock compositions; Allegre et al., 2010). These results confirm that the inverse model used in this study yields consistent results both for the chemical composition and fractional contribution of sources even if the *a-priori* inputs of model parameters are changed by an order of magnitude, or the isotopic ratio ($^{87}\text{Sr}/^{86}\text{Sr}$ in the present study) is varied significantly.

In addition to the inversion method, the LSR was also used (*cf.* section 2) to apportion the sources of dissolved solutes to the *global river* using two sets of data for source composition, in the first case the *a-priori* compositions of sources (Table 1B) and in the second case the IM-derived *a-posteriori* values was used as the input. In the first case the chemical data for the *global river* (Table 1A) and its possible sources (Table 1B) are the same for both the approaches and therefore allow direct inter-comparison of their results. The LSR approach yielded the fraction of dissolved Na supplied by rain (68%), silicates (57%), carbonates (3%) and evaporites (-24%), significantly different from that of the

IM and absurd as one of the sources (evaporites) has negative contribution. Further, the contribution of Na from silicates and rain obtained from the LSR is significantly higher compared to that obtained from the IM. The estimated non-positive contribution from one of the sources using the LSR underscores the limitation of the method. A likely cause for this difference is the use of imprecise information on the source compositions as these results depend critically on them. Such results have also been documented in some of the earlier studies based on the factor analysis and PCA (Larsen and Baker, 2003; Sofowote et al., 2008).

In the second case, the use of IM-derived *a-posteriori* results, which are expected to be more representatives of the source compositions, estimated the Na contribution of 46%, 36%, 15% and 3% respectively from silicates, rain, evaporites and carbonates to the *global river*. These estimates are nearly the same as those obtained from the IM. Error-weighted regression fit of the results on the source contributions obtained from these two models shows that they are statistically the same (Fig. 2), with a slope of 1.00 ± 0.15 and intercept of near zero. Note that the LSR model does not provide errors on its results.

Further, to assess the reliability of the model calculations, summation of the square of residuals and square of relative residuals (of all mass balance equations) are estimated from both the models. The details on these residuals are provided in the Appendix. The residual values for the LSR are 0.42 and 7.3×10^{-2} , respectively, similar to the corresponding values of 0.44 and 7.5×10^{-2} obtained for the IM. These findings suggest that, if accurate information on the composition of sources is available, the LSR and IM approaches converge to the same result. The results from the two case studies further suggest that the IM is more versatile compared to the LSR approach as the former can provide meaningful estimates on source contributions even by using rough values as inputs for source composition.

4.2. Case study 2: Sedimentary Mixture

In another case study, both the least square regression and inversion approaches were used to estimate contributions from various sources/components to a sedimentary mixture, and the results obtained are compared. For the mathematical framework, the chemical composition of a sedimentary mixture/till, reported by Makinen and Gustavsson (1999), has been used. The till is composed of four components: amphibolites (AFB), granitoids (GR), quartzites (QRTZ) and sands (S). The available information on the till and its four components are the abundances of major elements (Al, Ca, K, Mg, Na, Si and Ti) oxides and trace element (Cr, Ni, Pb, Rb, Sr, V and Y). The aim is to find the proportion

of the individual components contributing to the till by solving the mixing equations (*cf.* Appendix). Makinen and Gustavsson (1999) apportioned the contribution of components using the Chebyshev's method and therefore, selecting this particular geochemical dataset provides an additional scope to compare results obtained from our study using the LSR and IM with those reported previously using the Chebyshev's method.

4.2.1. Results from the Least square regression and Inversion approaches

The solutions for the mass balance equations of the till are obtained using (i) the LSR approach with both the *a-priori* and *a-posteriori* source compositions from the IM; (ii) the inversion method following the Quasi-Newton approach.

The results on the contribution of various sources to the till based on the two models (LSR and IM) and the use from the Chebyshev's method reported by Makinen and Gustavsson (1999) are presented in Table 2. The proportions of the four components (AFB, GR, QRTZ and S) contributing to the till as determined by Makinen and Gustavsson (1999) are 9%, 26%, 3% and 62%, respectively. These estimates compare well with the corresponding values of 8%, 20%, 6% and 65% obtained from the LSR, using the same *a-priori* data as those used for the IM. The LSR results change to 11%, 22%, 10% and 57% when the IM-derived *a-posteriori* data are used as the input for the source composition. These results agree well with the results from the IM, $13 \pm 3\%$, $24 \pm 6\%$, $11 \pm 3\%$ and $52 \pm 9\%$ respectively. These comparisons suggest that the IM model results on source contributions are in better agreement with the results of the LSR model when the *a-posteriori* source compositions from the IM is used as the input for the later. Further, these results are also in overall agreement with that obtained using the Chebyshev's method (Table 2; Fig. 3).

The computed sum of the squares of residuals and relative residuals from the models are also listed in Table 2. It shows that values obtained using the IM are intermediate between the values from the LSR and the Chebyshev's approaches, with the LSR providing the least values amongst the computer codes. This observation is also evident from Fig. 4, which shows a comparison of the ratios of the model predicted abundances (Q_c) of major element oxides and trace elements in the till with the measured values (Q_m). It is observed from the Fig. 4 that in majority of the cases, (i) the agreement between the model derived and measured values is better (within $\pm 2\%$) for the LSR and IM, while the computed values using the Chebyshev's model often deviate by more than 2%; (ii) the values obtained using the LSR model matches closest to the measured values. For the LSR, although the Q_c values used

in Fig. 4 are the results obtained using *a-priori* inputs, the above observations remain similar even if the results with *a-posteriori* inputs are considered.

Similar to the results obtained for the *global river* water, the LSR and IM models provided statistically similar estimates when the IM-derived *a-posteriori* are used for source composition in the LSR. The slope of the error-weighted regression relation between the LSR results (with *a-posteriori* inputs) and the IM ones is found to be 0.89 ± 0.19 , indistinguishable from 1.0 within associated uncertainties. This value decreases to 0.70 ± 0.15 , different than unity, when the LSR modeling was done with *a-priori* values of the sources. Possibly, this could have arisen because of difference in the methodology (*a-priori* vs. *a-posteriori*) adopted for the LSR approach and/or may be due to lack of information on the uncertainty associated with the estimates obtained using the LSR. The limitation of the LSR approach is that it does not provide uncertainty associated with its results (*i.e.*, source contributions). In order to gain information on the magnitude of errors associated with the LSR, the results obtained from the LSR and IM models are statistically analyzed by assigning incremental values of uniform relative error varying between 5% and 30% on each of the LSR obtained values. The results of this exercise show that with increase in errors associated with the LSR values, the slope of the (Williamson bi-directional error-weighted) regression line approached towards unity. The slope of the correlation line (0.73 ± 0.26 ; Fig 5) became close to unity within uncertainties when the relative errors assigned to the LSR results are considered to be $\sim 25\%$. This comparison suggests that the estimates from the LSR are likely to be associated with uncertainty of $\sim 25\%$, similar to the errors associated with the estimates obtained from the IM. It is noted that error estimate on each of the LSR results may not be uniform ($\sim 25\%$), and therefore the errors derived from the above indirect approach should be considered as a first order assessment for uncertainty associated with the LSR results.

5. Merits and limitations of the models

The case studies presented above demonstrate that the LSR approach can provide quantitative information on the relative contribution of sources to a mixture by solving a set of over-determined equations. The use of the LSR for source apportionment studies is often limited as it does not provide covariance associated with the results. Further, the robustness and accuracy of the results depend on how well the source compositions are constrained. This requirement can often be met in mixtures of solid components, whereas in mixtures such as the river water, source compositions are generally poorly constrained and have to be inferred by other independent approaches. The observed large difference

between the LSR and IM results for the source apportionment of the *global river* water (cf. section 4.1.) is, therefore, most likely attributable to limited information on their source compositions and incongruent weathering of bedrocks in the basin.

The merit of the inverse model is that it requires only rough estimates of *a-priori* source composition, making it more flexible compared to the use of the LSR. Through iterative approaches, this model finds solutions which are best-fit to the mass balance equations of all the parameters. The inversion approach converges to nearly the same results, even when the *a-priori* composition of the sources is varied over orders-of-magnitude-differences (Tripathy and Singh, 2010), but it may not be able to converge to the global optimum of the joint probability distribution of data- and model- space, when the *a-priori* data are varied far more widely, thereby yielding results with large uncertainties. The quantification of the source contribution with uncertainties is the major merit of the IM model as geochemical significances of both processes and sources contributing to the elements in a mixture depend on how reliable estimates can be made.

Although this study compares the LSR and the IM results for two case studies involving geochemistry of river water and glacial tills, outcomes of this study can be extrapolated to other Earth science problems, such as dynamics of ore-forming processes, magma evolution, crustal contamination and geo-environmental problems. Results from this study indicate that the inversion method compared to the LSR can provide more accurate information for wide spectrum of Earth science problems.

6. Conclusions

In Earth science studies, source apportionment of geological samples is achieved through suitable modeling of their composition along with those of their potential sources. In this study, we present a detailed comparative study of source apportionment based on the least square regression and the inversion models on two mixtures: the *global river* water and glacial till. The source contribution estimated using the LSR model with the same end member compositions as the IM-derived *a-posteriori* values yielded nearly identical results as those of the inverse model. In contrast, when the calculations are made with the LSR and IM models using the same end-member compositions as the *a-priori* values of the IM model, the results diverge significantly. Analogous to the *global river* analysis, in the case of till also, better agreement between the two models is observed when the *a-posteriori* information derived from the IM is used as the input for the LSR. Further, estimates of uncertainties in the LSR model is made by uniformly increasing the input end-member composition starting with the *a-priori*

values used in the IM. This exercise seems to indicate that for the two models to yield results which are the same within errors, the relative errors associated with the LSR has to be $\sim 25\%$. This inter-comparison exercise underscores the need to have accurate knowledge of end member compositions to obtain reliable results from the LSR, unlike the IM model which is more versatile and applicable to systems where source compositions are less constrained.

Acknowledgements:

We thank Tarak Nath Dey and Subimal Deb for their help in computer programming. Discussions with Sunil K Singh were helpful. S. Krishnaswami is thankfully acknowledged for the constructive comments that helped improving the manuscript. Constructive suggestions of Chongbin Zhao and two anonymous reviewers are thankfully acknowledged. This is NIO contribution no. xxxx.

References

- Allegre, C., Hart, S., Minister, J., 1983. Chemical structure and evolution of the mantle and continents determined by inversion of Nd and Sr isotopic data, I. Theoretical methods. *Earth and Planetary Science Letters* 66, 177-190.
- Allegre, C., Louvat, P., Gaillardet, J., Meynadier L., Rad, S., Campas, F., 2010. The fundamental role of island arc weathering in the oceanic Sr isotope budget. *Earth and Planetary Science Letters* 292, 51-56.
- Beus, S. V., Grigorian, A. A. 1977. *Geochemical exploration methods for mineral deposits*. Applied publishing ltd., Wilmette, USA, 287 pp.
- Berner R. A., Lasaga A. C., Garrels R. M., 1983. The carbonate-silicate geochemical cycle and its effect on atmospheric carbon-dioxide over the past 100 million years. *American Journal of Science* 284, 641-683.
- Berner, E. K., Berner, R. A., 1996. *Global Environment: Water, Air and Geochemical Cycles*, Prentice Hall, Upper Saddle River, N. J., 376 pp.
- Bickle, M. J., Chapman, H. J., Bunbury, J., Harris, N. B. W., Fairchild, I. J., Ahmad, T., Pomies, C., 2005. Relative contributions of silicate and carbonate rocks to riverine Sr fluxes in the headwaters of the Ganges. *Geochimica et Cosmochimica Acta* 69, 2221-2240.
- Brantley, S., White, A., 2009. Approaches to modeling weathered regolith. *Reviews in Minerology and Geochemistry* 70, 435-484.
- Bryan, W.B., Finger, L.W., Chayes, F., 1969. Estimating proportions in petrographic mixing equations by least-squares approximation. *Science* 163, 926-927.

- Das, A., Krishnaswami, S., Pande, K., 2005. Chemical weathering in the Krishna basin and western ghats of the Deccan traps, India : Rates of basalt weathering and their controls. *Geochimica et Cosmochimica Acta* 69, 2067-2084.
- Das, A., Krishnaswami, S. 2007. Elemental geochemistry of river sediments from the Deccan Traps, India: Implications to sources of elements and their mobility during basalt–water interaction. *Chemical Geology* 242, 232-254.
- Drever, J. I., 1997. *The geochemistry of natural waters: surface and groundwater environments*. Prentice Hall, Upper saddle, 436 pp.
- Gaillardet J., Dupré, B., Allègre, C. J. 1999a. Geochemistry of large river suspended sediments: silicate weathering or recycling tracer? *Geochimica et Cosmochimica Acta* 63, 4037-4051.
- Gaillardet, J., Dupre, B., Louvat, P., Allegre, C., 1999b. Global silicate weathering and CO₂ consumption rates deduced from the chemistry of large rivers. *Chemical Geology* 159, 3-30.
- Garrels, R. M., Mckenzie, F. T. 1971. *Evolution of sedimentary rocks*. W. W. Norton and Company, N. Y., 397 pp.
- Gow, P., Upton, P., Zhao, C., Hill, K., 2002. Copper-Gold mineralization in the New Guinea: numerical modeling of collision, fluid flow and intrusion-related hydrothermal system. *Australian Journal of Earth Science* 49, 753-771.
- Hobbs, B.E., Zhang, Y., Ord, A., 2000. Application of coupled deformation, fluid flow, thermal and chemical modelling to predictive mineral exploration. *Journal of Geochemical Exploration* 69-70, 505-509.
- Huang, S., Conte, M., 2009. Source/process apportionment of major and trace elements in sinking particles in the Sargasso sea. *Geochimica et Cosmochimica Acta* 73, 65-90.
- Ju, M., Dai, T., Yang, J., 2011. Finite element modeling of pore-fluid flow in the Dachang ore district, Guangxi, China: Implications for hydrothermal mineralization. *Geoscience Frontiers* 2(3), 463-474.
- Kim, E., Hopke, P.K., Larson, T.V., Maykut, N.N., Lewtas, J., 2004. Factor analysis of Seattle fine particles. *Aerosol Science and Technology* 38, 724–738.
- Krishnaswami, S., Singh, S. K., Dalai, T. K., 1999. Silicate weathering in the Himalaya: Role in contributing to major ions and radiogenic Sr to the Bay of Bengal. In: Somayajulu, B. L. K. (Eds.), *Ocean Science, Trends and Future Directions*, Indian National Science Academy and Akademia International, pp. 23–51..
- Larsen, R., Baker, J., 2003. Source apportionment of polycyclic aromatic hydrocarbons in the urban atmosphere: a comparison of three methods. *Environmental Science and Technology* 37, 1873-1881.

- Le Maitre, R., 1979. A new generalised petrological mixing model. *Contributions to Mineralogy and Petrology* 71, 133-137.
- Lin, G., Hobbs, B.E., Ord, A., Muhlhaus, H.B., 2003. Theoretical and numerical analyses of convective instability in porous media with temperature-dependent viscosity. *Communications in Numerical Methods in Engineering* 19, 787-799.
- Lin, G., Zhou, Y., Wei, X., 2006. Structure controls on fluid flow and related mineralization in the Xiangshan uranium deposit, Southern China. *Journal of Geochemical Exploration* 89, 231-234.
- Lin, G., Hobbs, B.E., Zhang, L., Zhou, Y., 2008. Potential effects of upward throughflow on thermal structure models within the continental lithospheric mantle-crust. *Chinese Journal of Geophysics* 51(2), 393-401.
- Lin, G., Peng, M., Zhang, L., Zhang, D., Liu, S., 2009. Numerical analysis and simulation experiment of lithospheric thermal structures in the South China Sea and the Western Pacific. *Journal of Earth Science* 20, 85-94.
- Liu, L., Yang, G., Peng, S., 2005. Numerical modelling of coupled geodynamical processes and its role in facilitating predictive ore discovery: An example from Tongling, China. *Resource Geology* 55, 21-31.
- Liu, L., Shu, Z., Wan, C., Cai, A., Zhao, Y., 2008. The controlling mechanism of ore formation due to flow-focusing dilation spaces in skarn ore deposits and its significance for deep-ore exploration: Examples from the Tongling-Anqing district. *Acta Petrologica Sinica* 24, 1848-1856.
- Liu, L., Wan, C., Zhao, Y., 2011. Geodynamic constraints on orebody localization in the Anqing orefield, China: Computational modeling and facilitating predictive exploration of deep deposits. *Ore Geology Reviews* 43, 249-263.
- Lupker, M., France-Lanord, C., Galy, V., Lave, J., Gaillardet, J., Gaujrel, A. P., Guilmette, C., Rahman, M., Singh, S. K., Sinha, R., 2012. Predominant floodplain over mountain weathering of Himalayan sediments (Ganga basin). *Geochimica et Cosmochimica Acta* 84, 410-432.
- Maher, K., 2010. The dependence of chemical weathering rates on fluid residence time. *Earth and Planetary Science Letters* 294, 101-110.
- Makinen, J., Gustavsson, N., 1999. Estimating proportions in geochemical mixing equations by Chebyshev's method. *Applied Geochemistry* 14, 85-90.
- Meybeck, M., 2003. Global Occurrence of major elements in rivers. in: Drever, J. I. (Eds.), *Surface and Ground water, weathering and soils*, Vol. 5, *Treatise on geochemistry*, H. D. Holland and K. K. Turekian (Eds.), Oxford: Elsevier-Pergamon, 207-223.
- Miller, C. A., Peucker-Ehrenbrink, B., Walker, B. D., Marcantonio, F., 2011. Re-assessing the surface cycling of molybdenum and rhenium. *Geochimica et Cosmochimica Acta* 75, 7146-7179.

- Millot, R., Gaillardet, J., Dupré, B., Allègre, C. J., 2003. Northern latitude chemical weathering rates: clues from the Mackenzie River Basin, Canada. *Geochimica et Cosmochimica Acta* 67, 1305-1329.
- Moon, S., Huh, Y., Qin, J., van Pho, N., 2007. Chemical weathering in the Hong (Red) River basin: Rates of silicate weathering and their controlling factors. *Geochimica et Cosmochimica Acta* 71, 1411-1430.
- Morandi, M., Liou, P., Daisey, J., 1991. Comparison of two multivariate modeling approaches for the source apportionment of inhalable particulate matter in Newark, NJ. *Atmospheric Environment* 25A, 927-937.
- Negrel, P., Allegre, C. J., Dupre, B., Lewin, E., 1993. Erosion sources determined by inversion of major and trace element ratios and strontium isotopic ratios in river water: the Congo Basin case. *Earth and Planetary Science Letters* 120, 59-76.
- Nesbitt, H. N., Markovics, G., 1997. Weathering of granodioritic crust, long-term storage of elements in weathering profiles, and petrogenesis of siliciclastic sediments. *Geochimica et Cosmochimica Acta* 61, 1653-1670.
- Ord, A., Hobbs, B.E., Zhang, Y., Broadbent, G.C., Brown, M., Willetts, G., Sorjonen-Ward, P., 2002. Geodynamic modeling of the century deposit, Mt Isa Province, Queensland. *Australian Journal of Earth Sciences* 49, 1011-1039.
- Paatero, P., Tapper, U., 1993. Analysis of different modes of factor analysis as least squares fit problems. *Chemometrics and Intelligent Laboratory Systems* 18, 183-194.
- Paatero, P., 1997. Least squares formulation of robust non-negative factor analysis. *Chemometrics and Intelligent Laboratory Systems* 37, 23-35.
- Rahaman, W., Singh, S. K., 2012. Sr and $^{87}\text{Sr}/^{86}\text{Sr}$ in estuaries of western India: Impact of submarine groundwater discharge. *Geochimica et Cosmochimica Acta* 85, 275-288.
- Schmidt Mumm, A., Brugger, J., Schacht, U., 2010. Fluids in geological processes: The present state and future outlook. *Journal of Geochemical Exploration* 106, 1-7.
- Singh S. P., Singh, S. K., Goswami, V., Bhushan, R., V. Rai, 2012. Spatial distribution of dissolved neodymium and ϵ_{Nd} in the Bay of Bengal: Role of particulate matter and mixing of water masses. *Geochimica et Cosmochimica Acta* 94, 38-56.
- Sohn, A., 2005. A general inversion for end-member ratios in binary mixing systems. *Geochemistry Geophysics Geosystem* 6, 11, Q11007.
- Sofowote, U., McCarry, B., Marvin C., 2008. Source apportionment of PAH in Hamilton Harbour suspended sediments: Comparison of two factor analysis methods. *Environmental Science and Technology* 42, 6007-6014.

- Sorjonen-Ward, P., Zhang, Y., 2002. Numerical modeling of orogenic processes and mineralization in the south eastern part of the Yilgarn Craton, Western Australia, *Australian Journal of Earth Sciences* 49, 935-964.
- Sudheer, A. K., Rengarajan, R., 2012. Atmospheric mineral dust and trace metals over urban environment in western India during winter. *Aerosol and Air quality Research* 12, 923-933.
- Tarantola, A., 2005. The least square criterion. in: *Inverse problem theory and methods for model parameter estimation*. Society for Industrial and Applied Mathematics, Philadelphia, pp. 68-72.
- Tripathy, G. R., Singh, S. K., 2010. Chemical erosion rates of river basins of the Ganga system in the Himalaya: Reanalysis based on inversion of dissolved major ions, Sr, and $^{87}\text{Sr}/^{86}\text{Sr}$. *Geochemistry Geophysics Geosystem* 11, Q03013. doi: 10.1029/2009gc002862.
- Tripathy, G. R., Goswami, V., Singh, S. K., Chakrapani, G. J., 2010. Temporal variations in Sr and $^{87}\text{Sr}/^{86}\text{Sr}$ of the Ganga headwaters: estimates of dissolved Sr flux to the mainstream. *Hydrological Processes* 24, 1159-1171.
- Tripathy, G. R., Singh, S. K., Bhushan, R., Ramaswamy, V., 2011. Sr-Nd isotope composition of the Bay of Bengal sediments: Impact of climate on erosion in the Himalaya. *Geochemical Journal* 45, 175-186.
- Tripathy, G. R., Singh, S. K., Ramaswamy, V., 2014. Major and trace element geochemistry of Bay of Bengal sediments: Implications to provenances and their controlling factors. *Palaeogeography Palaeoclimatology Palaeoecology* 397, 20-30.
- Walker, J. C. G., P. B. Hays, J. F. Kasting, 1981. A negative feedback mechanism for the longterm stabilization of Earth's surface temperature. *Journal of Geophysical Research* 86, 9776-9782.
- Weltje, G., 1997. End-member modeling of compositional data: Numerical-statistical algorithms for solving the explicit mixing problem. *Mathematical Geology* 29, 503-549.
- Wright, T.L., Doherty, P.C., 1970. A linear programming and least squares computer method for solving petrologic mixing problems. *Geological Society of America Bulletin* 81, 1995-2008.
- Wu, L., Huh, Y., Qin, J., Du, G., van Der Lee, S. 2005. Chemical weathering in the Upper Huang He (Yellow River) draining the eastern Qinghai-Tibet Plateau. *Geochimica et Cosmochimica Acta* 69, 5279-5294.
- Xing, H.L., Makinouchi, A., 2008. Three-dimensional finite element simulation of large-scale nonlinear contact friction problems in deformable rocks. *Journal of Geophysics and Engineering* 5, 27-36.
- Yan, Y., Lin, G., Wang, Y., Guo, F., Li, Z., Li, X., 2003. Apatite fission track age of Mesozoic sandstones from Beipiao basin, eastern China: Implications for basin provenance and tectonic evolution. *Geochemical Journal* 37, 377-389.

- Zhang, L., Li, Z., Lin, G., Guo, H., 2011. Numerical simulation of effects of upward throughflow on thermal structure and the thickness of the continental lithosphere. *Journal of Geophysics and Engineering* 8, 322-329.
- Zhang, Y., Hobbs, B.E., Ord, A., Barnicoat, A., Lin, G., 2003. The influence of faulting on host-rock permeability, fluid flow and ore genesis of gold deposits: a theoretical 2D numerical model. *Journal of Geochemical Exploration* 78-79, 279-284.
- Zhang, Y., Schaub, P.M., Ord, A., Hobbs, B.E., Barnicoat, A., 2008. Fault-related dilation, permeability enhancement, fluid flow and mineral precipitation patterns: numerical models, in: Wibberley, C.A.J., Kurz, W., Imber, J., Holdsworth, R.E., Collettini, C. (Eds.), *The Internal Structure of Fault Zones: Implications for Mechanical and Fluid-flow Properties*. Geological Society, vol. 299. Special Publications, London, pp. 239–255.
- Zhao, C., Hobbs, B.E., Hornby, P., Ord, A., Peng, S., Liu, L., 2008a. Theoretical and numerical analyses of chemical-dissolution front instability in fluid-saturated porous rocks. *International Journal for Numerical and Analytical Methods in Geomechanics* 32, 1107-1130.
- Zhao, C., Hobbs, B.E., Ord, A., 2008b. *Convective and Advective Heat Transfer in Geological Systems*. Springer, Berlin, 229pp.
- Zhao, C., 2009. *Dynamic and Transient Infinite Elements: Theory and Geophysical, Geotechnical and Geoenvironmental Applications*, Springer, Berlin, 259pp.
- Zhao, C., Hobbs, B.E., Ord, A., 2009. *Fundamentals of Computational Geoscience: Numerical Methods and Algorithms*. Springer, Berlin, 323 pp.
- Zhao, C., Hobbs, B.E., Ord, A., 2010. Theoretical analyses of nonaqueous-phase-liquid dissolution induced instability in two-dimensional fluid-saturated porous media, *International Journal for Numerical and Analytical Methods in Geomechanics* 34, 1767-1796.

Appendix

A. Computational Programming

A-1: Least Square Regression

A computational code for least square regression of a suite of mass balance equations is developed using the *Mathematica* software. The mathematical details are provided in the text (*cf.* Section 2.1); this program consists of the following steps:

1. Two matrices of chemical data for the mixtures (B) and its possible sources (A) are formed.
2. Available datasets for these two matrices are imported.
3. The relation between these two matrices is the mass balance equations 1 and 2 of section 2.1. These equations that relate matrices A and B are included in the code.
4. The subroutine *L2QRSolve* from the *Mathematica* software is used to perform the least square regression of the defined linear system (equation 1; section 2.1). This subroutine “*L2QRSolve*” (<http://reference.wolfram.com/mathematica>) is based on the following algorithm,

```
L2QRSolve[a_, b_] :=  
Module[  
  {Q, R, m, n},  
  {m, n} = Dimensions[a];  
  {Q, R} = QRDecomposition[a];  
  LinearSolve[R, Q.b]  
]
```

5. This program estimates the relative contribution from each of the sources to the mixture (*i.e.* the matrix (X)) using least square approximation through QR decomposition (*cf.* Section 2.1).

A-2: Inverse Modeling

The mathematical details of the inversion method used in this study are presented in the text (*cf.* Section 2.2); the approach adopted to develop this program is detailed below:

1. The computational code defines two set of matrices (d_{obs} and C_D ; m and C_M). The d_{obs} and C_D matrices are composed of chemical data and associated covariance of the mixture respectively. The m and C_M are the matrices that include *a-priori* chemical composition and associated covariance of all possible sources of the mixture.
2. Available data for these matrices are imported.
3. The inversion of the matrices, C_D and C_M , are calculated for the available data and termed as *ICD* and *ICM* respectively.

4. The mass balance equations $g(m)$ relating d_{obs} and model parameters are defined.
5. A matrix, $dCalmat$ is defined as $dCalmat = G_{\alpha}^i = \left(\frac{\partial g^i}{\partial m^{\alpha}}\right)_{m_{prior}}$.
6. Using these defined matrices, an iterative algorithm for estimating least misfit (S) function starting from an initial misfit value of $Sold$ is developed in *Mathematica* and is provided below,

```

Sold=0;
While[
  Abs[(Sold-S/.refresh)/S/.refresh]>10^(-5),
  Sold=S/.refresh;
  Print["mcurr = ",mcurrent," S = ",S/.refresh];
  dataresiduals=(dCalmat-dobs)/.refresh;
  modelresiduals=(mcurrent-mprior)/.refresh;
  g=G/.refresh;gt=GT/.refresh;
  gradient=gt.ICD.dataresiduals+ICM.modelresiduals;
  direction=Inverse[gt.ICD.g+ICM].gradient;
  mnew=mcurrent-μ.direction;
  mcurrent=mnew;
]

```

7. This iterative program converges to best-fit for source composition and their respective contributions to the mixtures. The inverse model also provides the covariance associated with all the resulting parameters.

B. Mass Balance Equations:

B-1: River water system

For the source apportionment of solutes of the *global river* water, the mass balance equations for various elemental (Na-normalized) ratios can be written as:

$$\sum_{i=1}^4 \left(\frac{X}{Na}\right)_i \times f_i(Na) = \left(\frac{X}{Na}\right)_{riv} \quad (A-1)$$

$$\sum_{i=1}^4 \left(\frac{{}^{87}Sr}{{}^{86}Sr}\right)_i \times \left(\frac{Sr}{Na}\right)_i \times f_i(Na) = \left(\frac{Sr}{Na}\right)_{riv} \times \left(\frac{{}^{87}Sr}{{}^{86}Sr}\right) \quad (A-2)$$

$$\sum_{i=1}^4 f_i(Na) = 1 \quad (A-3)$$

where $f_i(Na)$ refers to the fraction of Na contributed by source i ; $\left(\frac{X}{Na}\right)_{riv}$ and $\left(\frac{X}{Na}\right)_i$ are the Na-normalized molar ratios of element X (Cl^- , Ca^{+2} , Mg^{+2} , Sr^{+2} , HCO_3^-) measured in the river water and in source i respectively. Source $i=1, 2, 3, 4$ refers to rain, silicates, carbonates and evaporites.

B-2: Sedimentary system

In case of sedimentary mixtures (e.g. glacial till in this study), the general set of mass balance equations is expressed as:

$$\sum_{j=1}^4 X_j \times f_j = X_{till} \quad (A-4)$$

$$\sum_{j=1}^4 f_j = 1 \quad (A-5)$$

where $X = Al_2O_3, CaO, K_2O, MgO, Na_2O, SiO_2, TiO_2, Cr, Ni, Pb, Rb, Sr, V$ and $Y; j=1, 2, 3$ and 4 refers to Amphibolite, Granitoid, Quartz and Sands and f_j refers to the (mass) fractional contribution of component i to the sedimentary mixture. *Till* refers to the ‘glacial till’.

C. Residual Calculations

The residual is defined as the difference between the measured and the model values of a chemical parameter. The general expressions of square of residuals and square of relative residuals for the case studies discussed earlier, respectively, are $\sum (Q_m - Q_c)^2$ and $\sum [(Q_m - Q_c) / Q_m]^2$.

$$\text{For the global river: } Q_c = \sum_{i=1}^4 \left(\frac{X}{Na}\right)_i \times f_i(Na) \text{ and } Q_m = \left(\frac{X}{Na}\right)_{riv} \quad (A-6)$$

$$\text{For the sedimentary mixture: } Q_c = \sum_{j=1}^4 X_j \times f_j \text{ and } Q_m = X_{till} \quad (A-7)$$

where Q_m and Q_c are the measured and computed values; the meanings of other symbols are discussed earlier.

Generally, the numerical solutions are compared based on sum of the squares of the residuals, calculated for all the mass balance equations. However, we propose the use of sum of square of relative residual if measured quantities are not expressed in the same units (e.g., in sedimentary mixture the abundances vary from percentage to ppm). Caution should be taken not to compare results on the basis of square of residuals unless all the measured components are expressed in the same units.

Figure Captions:

Figure 1: Sensitivity test of inverse model: a-posteriori values of $f_{Si}(Na)$ are treated as a function of variations in a-priori $f_{Si}(Na)$. The results show that *a-posteriori* values of $f_{Si}(Na)$ converges 0.45 ± 0.01 for about an order of magnitude difference in its *a-priori* input.

Figure 2: Regression analysis of relative contribution (%) of four sources to the dissolved Na budget of the *global river* water estimated using the least square regression and inverse models: The straight line in the figure represents the linear fit of the data which shows that both the models yield nearly the same values for source contributions.

Figure 3: Comparison of relative contribution from the four components to the sedimentary mixture (till) estimated based on three different mathematical models, viz. the LSR and inverse model (present study) and Chebyshev's solution reported by Makinen and Gustavsson (1999): The LSR model relies on two sets of source contribution data: (i) *a-priori* values used in IM and (ii) *a-posteriori* information derived from IM.

Figure 4: Plot of ratios of the computed values of major element oxides and of trace elements (Q_C) in the till based on the three models to the measured values (Q_M): The bold line parallel to X-axis represents $Q_C = Q_M$ and the dashed lines represent values within $\pm 2\%$ of the ratios.

Figure 5: Scatter diagram of the results obtained from the least square regression and inversion models on the source contributions to the sedimentary mixture (Makinen and Gustavsson; 1999): Williamson bidirectional error-weighted linear fit (dashed line) of the datasets indicates that the results obtained from the LSR are likely to be associated with uncertainty of $\sim 25\%$ to provide statistically similar results to those obtained from the inversion approach. The solid line represents 1:1 fit of the datasets.

Table 1A: Discharge-weighted chemical composition and Sr isotopic data of the *global river water* obtained from composition data of major rivers in the world. Also reported are data from an earlier study. Rivers, which are possibly influenced by anthropogenic input (with Na<Cl), have not been considered for obtaining the global average values of major ions. Source of data for the major rivers are from Gaillardet et al. (1999b).

	Na ⁺	K ⁺	Ca ²⁺	Mg ²⁺	Cl ⁻	SO ₄ ²⁻	HCO ₃ ⁻	SiO ₂	Sr	⁸⁷ Sr/ ⁸⁶ Sr
	μM								nM	
Present Study	244	36	351	143	144	102	908	127	854	0.7118
Meybeck (2003)	240	44	297	123	167	88	798	145	-	-

Table 1B: *A-priori* elemental ratios (molar) used in the inverse model to apportion the sources of solutes of the *global river water*.

	Rain	Silicates	Carbonates	Evaporites
Cl/Na	1.15 ± 0.2	0.001 ± 0.001	0.001 ± 0.001	1.0 ± 0.2
Ca/Na	0.02 ± 0.02	0.35 ± 0.15	50 ± 20	0.17 ± 0.09
Mg/Na	0.11 ± 0.1	0.24 ± 0.12	20 ± 8	0.02 ± 0.01
HCO ₃ /Na	0.004 ± 0.004	2 ± 1	100 ± 40	0.3 ± 0.3
1000×Sr/Na	0.19 ± 0.1	3 ± 1	75 ± 25	3 ± 2
⁸⁷ Sr/ ⁸⁶ Sr	0.708 ± 0.005	0.74 ± 0.04	0.71 ± 0.05	0.708 ± 0.004

Data Source: Millot et al., 2003; Moon et al., 2007

Table 2: Comparison of contribution of four components (amphibolites (AFB), granitoids (GR), quartzite (QRTZ) and sand (S)) in the mixture (till) estimated from various mathematical models. The measured compositions of the mixture (till) and its four components reported by Makinen and Gustavsson (1999) are also reported along the computed chemical composition of the till obtained from the different models.

		Fractional contribution to Till (%)			
		Chebyshev*	LSR [§] (a-priori)	LSR [§] (a-posteriori)	Inverse [§]
Component	AFB	9	8	11	13 ± 3
	GR	26	20	22	24 ± 6
	QRTZ	3	6	10	11 ± 3
	S	62	65	57	52 ± 9
Residuals	$\Sigma(Q_M - Q_C)^2$	229	8	23	60
	$\Sigma((Q_M - Q_C)/Q_M)^2$	16×10 ⁻³	2×10 ⁻³	3×10 ⁻³	5×10 ⁻³

		Computed values for Till (Q _C)				Measured Values*				
		Chebyshev*	LSR [§] (a-priori)	LSR [§] (a-posteriori)	Inverse [§]	Till (Q _M)	AFB	GR	QRTZ	S
Al₂O₃		14.71	14.22	14.16	14.15	14.24	15.5	16.3	7.15	14.3
CaO		3.47	3.37	3.32	3.31	3.35	5.63	2.69	0.42	3.63
K₂O		2.77	2.68	2.67	2.66	2.68	1.98	3.42	1.73	2.67
MgO	Wt. %	2.40	2.33	2.34	2.36	2.34	5.54	1.57	0.15	2.4
Na₂O		3.14	3.01	2.94	2.90	2.99	1.99	3.83	0.9	3.13
SiO₂		65.31	65.39	66.56	66.54	65.38	51.8	66.9	88.1	65.5
TiO₂		0.93	0.91	0.89	0.89	0.9	1.58	0.58	0.11	1.02
Cr		88	87	87	86	86	155	56	30	95
Ni		31	30	31	31	31	82	13	8	32
Pb		32	32	32	31	32	28	33	24	33
Rb	ppm	79	75	76	76	74	76	100	42	72
Sr		296	284	278	275	282	225	359	79	291
V		136	133	134	136	134	316	76	31	140
Y		30	30	29	29	30	34	21	9	35

*From Makinen and Gustavsson (1999); §Results from present study

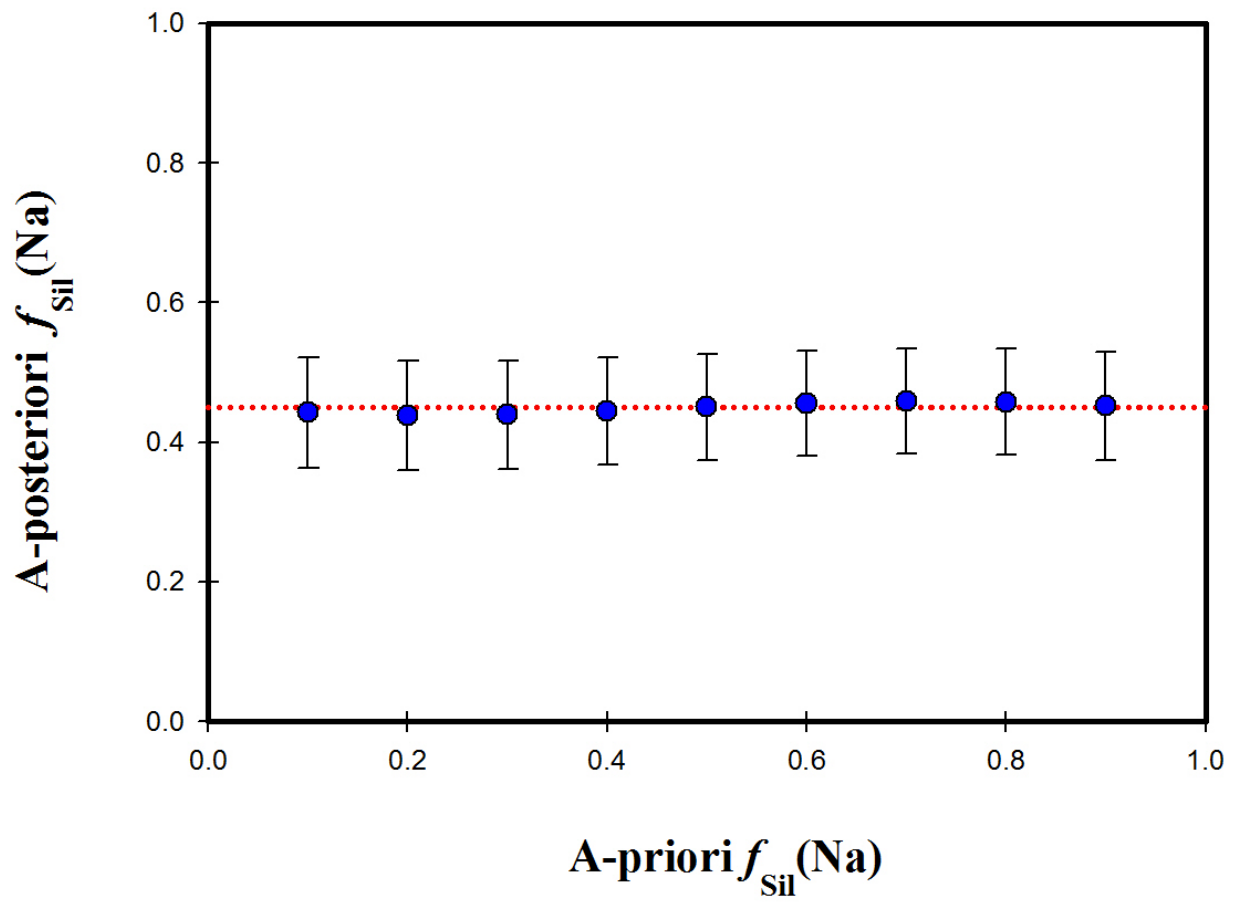


Figure 1

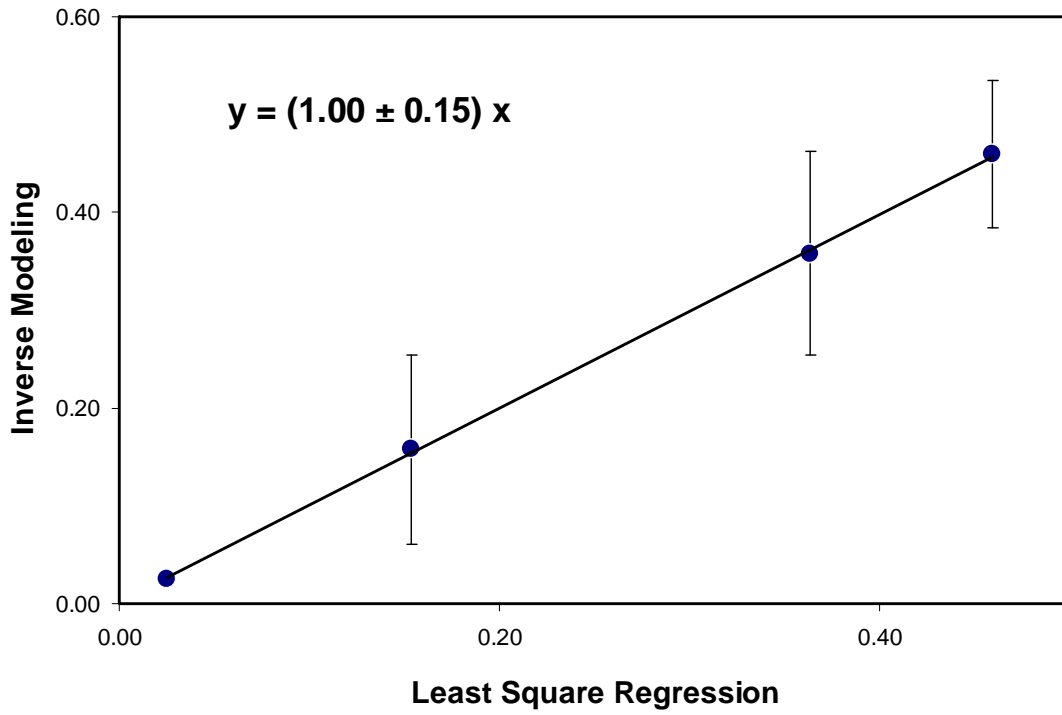


Figure 2

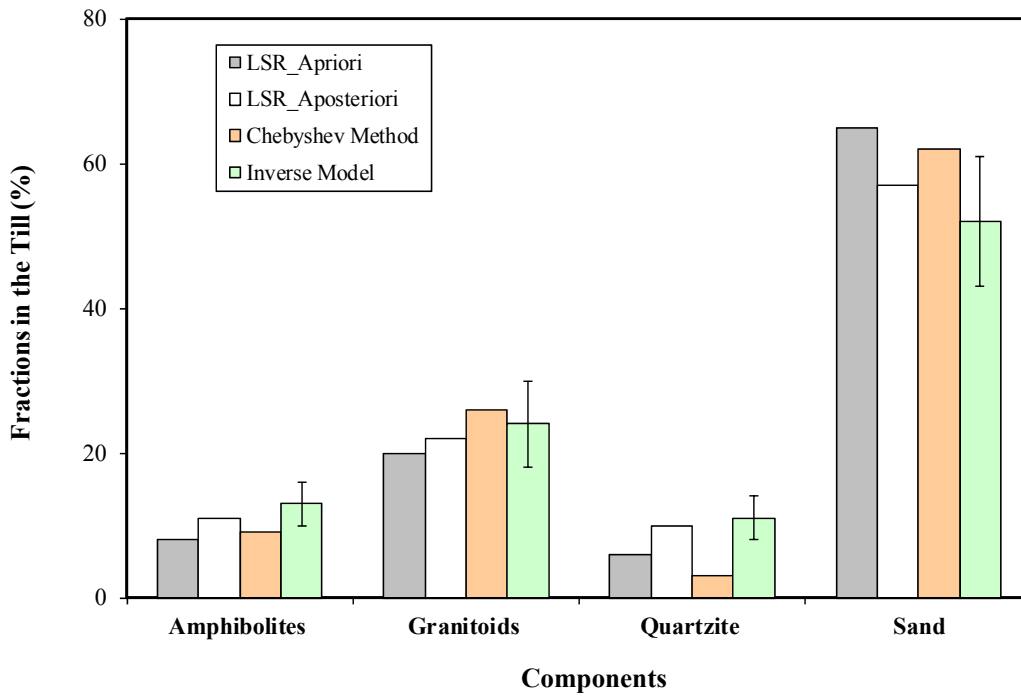


Figure 3

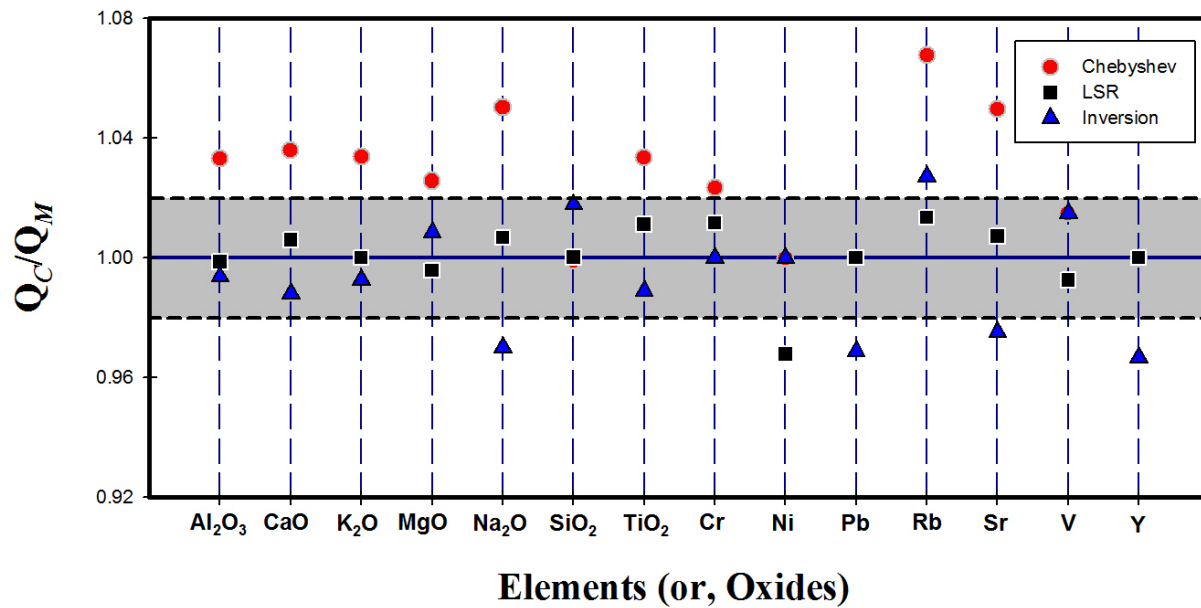


Figure 4

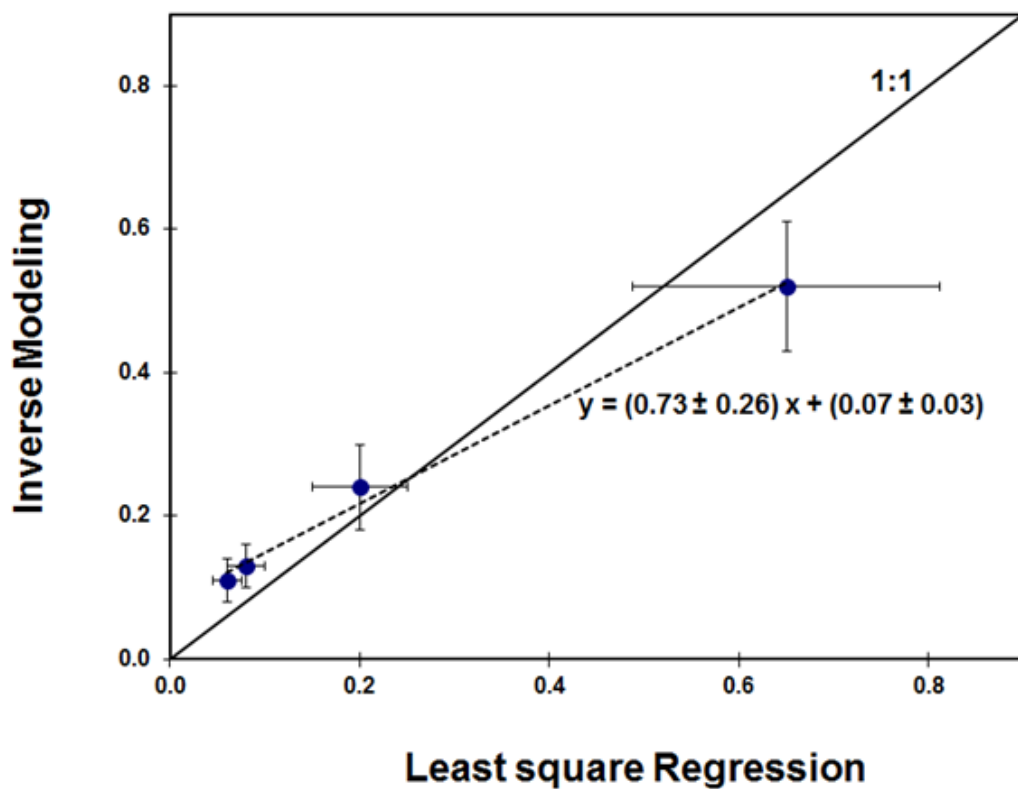


Figure 5