

Space Invariant Independent Component Analysis and ENose for Detection of Selective Chemicals in an Unknown Environment

Tuan A. Duong and Margaret A. Ryan

Jet Propulsion Laboratory/California Institute of Technology

Pasadena, CA 91109

E-mail: tuan@brain.jpl.nasa.gov

Abstract: *In this paper, we present a space invariant topology to enable the Independent Component Analysis (ICA) to solve chemical detection from two unknown mixing chemical sources. The four sets of unknown paired mixture sources are collected via JPL 16-ENose sensor array in the unknown environment with, at most, 12 samples data collected. Per time invariant aspect, this appears to be an overcomplete case in ICA where the number of outputs (32) is larger than the number of inputs (16).*

Our space invariant topology along with the techniques of maximum entropy information by Bell and Sejnowski and natural gradient descent by Amari has demonstrated that it is effective to separate the two mixing unknown chemical sources with unknown mixing levels to the array of two original sources under insufficient sampled data. From separated sources, they can be identified by projecting them on the 11 known chemical sources to find the best match for detection.

In this paper, we present the results of our simulations. These simulations have shown that 100% correct detection could be achieved under the two cases: a) under-completed case where the number of input (mixtures) is larger than number of original chemical sources; and b) regular case where the number of input is the same as the number of sources while the time invariant topology approach may face the obstacles: overcomplete case, insufficient data and cumbersome topology.

Keywords: *space invariant independent component analysis, ENose, chemical detection*

I. Introduction

The need for low-power, miniature sensor devices that can monitor air quality in an enclosed space with multi-compound capability and minimum human operation has led to the development of polymer-carbon composite based electronic nose (ENose) at NASA's Jet Propulsion Laboratory (JPL) [1-3]. The sensor array in the JPL ENose consists of 32 conductometric sensors made from insulating polymer films loaded with carbon. In its current design, it has the capability to detect 10 common contaminants which may be released into the recirculated breathing air of the space shuttle or space station from a spill or a leak; target concentrations are based on the 1-hour Spacecraft Maximum Allowable Concentrations (SMAC) set by NASA (see Table I) [4], and are in the parts-per-million (ppm) range. The ENose was intended to fill the gap between an alarm which has little or no ability to distinguish among chemical compounds causing a response and an analytical instrument which can distinguish all compounds present but with no real-time or continuous event monitoring ability.

As in other array-based sensor devices, the individual sensor films of the ENose are not specific to any one analyte; it is in the use of an array of different sensor films that gases or gas mixtures can be uniquely identified by the pattern of measured response. The response pattern requires software analysis to deconvolute gas compounds and their concentrations.

An example sensor set is shown in Figure 1a, and the complete device on which the data was used in this study is shown in Figure 1b.

The specific analysis scenario considered for this development effort was one of leaks or spills of specific compounds. It has been shown in analysis of samples taken from space shuttle flights that, in general, air is kept clean by the air revitalization system and contaminants are present at levels significantly lower than the SMACs [5]; the JPL ENose has been developed to detect target compounds released suddenly into the breathing environment. A leak or a spill of a solvent or other compound would be an unusual event. Release of mixtures of more than two or three compounds would

be still more unusual; such an event would require simultaneous leaks or spills to occur from separate sources. Thus, for this phase of development, mixtures of more than two target compounds were not considered.

In this paper, we consider an approach to analysis of sensor responses to mixture so that use of the JPL ENose may be extended to detection of chemical compounds in an open and changing environment, such as a building or a geographical area where air exchange is not controlled and limited.

Table I Target Gases and Concentrations

Target Compound	Concentration Range Tested (ppm)	1 hr. SMAC (ppm)	Reference Number
Benzene	10 – 50	30	1
Ethanol	10 – 130	2000	2
Freon	50 – 525	50	3
Indole	.006 – 0.06	1	4
Methane	3000 – 7000	5300	5
Methanol	10 – 300	30	6
Propanol	75 – 180	400	7
Toluene	30 – 60	16	8
Ammonia	10 – 50	30	9
Formaldehyde	50 – 510	0.4	10
Medical Wipe	500–4000	--	13

As such operation in the open environment, the collected sensory data will be a mixing between the unknown chemicals with the unknown mixing levels (coefficient) between them. The identification of the chemical compounds among these mixing chemicals is a real challenge for real world applications.

To search for a chemical compound whether it exists in the operating environment, one of the most robust techniques is to recover the original chemicals. When done, the detection can be an easy step by finding the minimum phase between the predicted original chemicals and the target chemicals. More sophisticated way, the neural network approach [6-10] can be employed to capture the target chemicals in various conditions through learning e.g., concentration levels through the parameterized weight set, then the strongest correlation between parameterized weight and the predicted original can be used to identify the intended chemical.

Recently, Independent Component Analysis (ICA) has proven that it not only de-correlates the second order statistics of the signals but also reduces the higher order statistical dependencies. ICA transforms an observed signal vector into a set of signals that are as statistically independent as possible. Theoretically, ICA is an information-theoretic approach, which exploits concepts from information theory such as entropy and mutual information.

The ICA roots in the early work of Herault and Jutten [11] who first introduced an adaptive algorithm in a simple feedback architecture that was able to separate several unknown independent sources. ICA was further developed by [12-17]. Amari et al. [18] have used natural gradient descent based on the Riemannian metric tensor to optimize the curvature of a particular manifold in n dimensional space. This technique is employed to apply to the Infomax [16] to simplify the learning rule that is used in this paper. ICA has applications for feature extraction in speech recognition systems, in communication systems, in medical signal processing, and in image processing.



Figure 1: a) a JPL ENose sensor substrate on the left with eight polymer-carbon composite sensors; b) the complete JPL ENose on the right.

II. Technical approach

In this study, we based on the approximation of the set up topology shown in Figure 2 below:

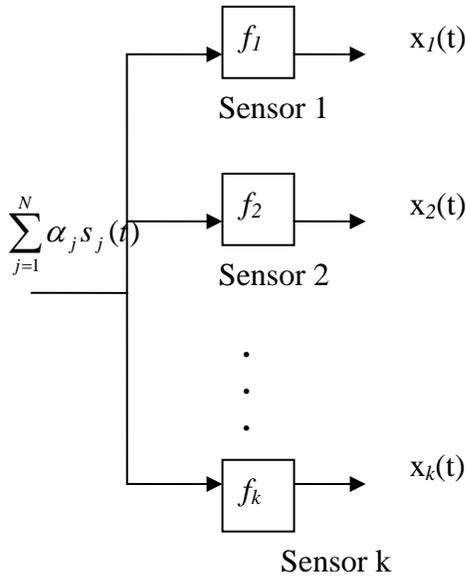


Figure 2: The system model and set up of the ENose array

In Figure 2, the collected sensing data $x_i(t)$ consists of changes in electrical resistance corresponding to sensor response to the unknown mixture of chemical sources s_j and their densities (or concentrations) at the time t .

Due to the small separation between the sensors themselves, the input of each sensor resistance is assumed to be uniquely distributed. The sensory data can be modeled as follows:

$$x_i(t) = f_i\left(\sum_{j=1}^N \alpha_j s_j(t)\right) \quad i = \overline{1-k} \quad (1)$$

Where f_i is the unknown non linear activation, α_j is the unknown mixing coefficient of chemical source j , and i is the index of sensor number and N as number of sensor and s_j is original source.

From equation (1), using the first order of a Taylor expansion, it is rewritten:

$$x_i = a_i + \sum_{j=1}^N \beta_{ij} s_j \quad (2)$$

$$\text{Where } a_i = f_i\left(\sum_{j=1}^N \alpha_j s_j^o\right) + \sum_{j=1}^N \left. \frac{\partial f_i}{\partial s_j} \right|_{s_j^o} \alpha_j s_j^o$$

$$\text{and } \beta_{ij} = \left. \frac{\partial f_i}{\partial s_j} \right|_{s_j^o} \alpha_j \quad (3)$$

And s_j^o is an operating point of the source s_j .

For each sampling data point in time for the same sensor i , x_i fluctuates around its operating point a_i and it can be considered as a common bias for $x_i(t)$ with $t \in [t, t + k\Delta t]$. From this argument, Equation (2) can be simplified to:

$$Y = \begin{bmatrix} x_1 - a_1 \\ x_2 - a_2 \\ \dots \\ x_k - a_k \end{bmatrix} \quad Y = \Gamma S \quad (4)$$

Where Y is unbiased mixture data, Γ is the unknown mixing matrix and S is the chemical source signal.

The learning rule based on the maximum entropy algorithm [11] is given by:

$$y = g(u) = g(Wx) \quad (5)$$

Where g is a non linear function *e.g.* the logistic function or hyperbolic tangent function. The update weight can be calculated as:

$$\Delta W = W^{-T} + \Phi(u)x^T \quad (6)$$

Where W^{-T} is an inverse transport of the $N \times N$ weight matrix W , x^T is a mixing input vector (observed vector), and

$$\Phi(u) = [\phi_i(u_i)] = \begin{bmatrix} \frac{\partial y_i}{\partial u_i} \\ \frac{\partial y_i}{\partial u_i} \end{bmatrix} \quad (7)$$

To simplify equation (5) using a natural gradient descent technique by Amari [13], the learning rule can be:

$$\Delta W = (I + \Phi(u)u^T)W \quad (8)$$

With $u = Wx$

The details can be found in [14].

III. Space Invariant Architecture

The most common ICA approach is that the number of variables and the number of sources are the same. However, in this study we face two obstacles: 1) there are 12 or less samples (mixing chemical compounds) from each sensor and the total number of sensor is 16 and they do not have sufficient data set (at least 16 data samples required); and 2) the number of variables is 16 as number of sensors while the number of compounds in a mixture is 2 and it is considered over complete case.

For the time invariant approach, the data that will flow orthogonal (dotted arrow) with the time invariant direction as shown in Figure 3 will require 32 outputs (16 channels for each chemical compound). The topology is 16 inputs, 32 outputs and 12 or less sample data which may not be a solvable problem.

Instead of using a time invariant approach, we use the space invariant approach (Figure 3 with dashed and dotted arrow) which allows us to have more data points and enable the square mapping matrix (the dimension of mixing sources and sensors are the same). This approach is feasible due to the mathematical model based on equations (2-4).

The architectures are shown below:

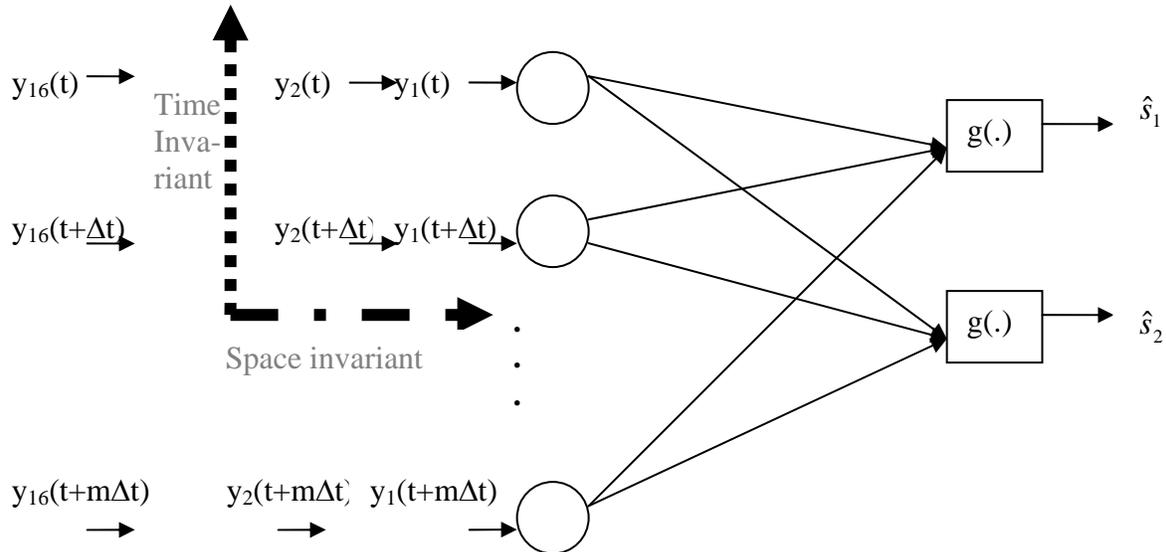


Figure 3: The space invariant ICA architecture (m=2-12)

In Figure 3, the unbiased input $Y_i(t)$ ($i=1-k$) is based on temporal mixture data and the sensory data are spatially invariant.

From laboratory set up, we have collected a set of single spectra of 11 chemicals using 16 elements in the ENose sensor array; it is averaged and shown in Figure 4 below.

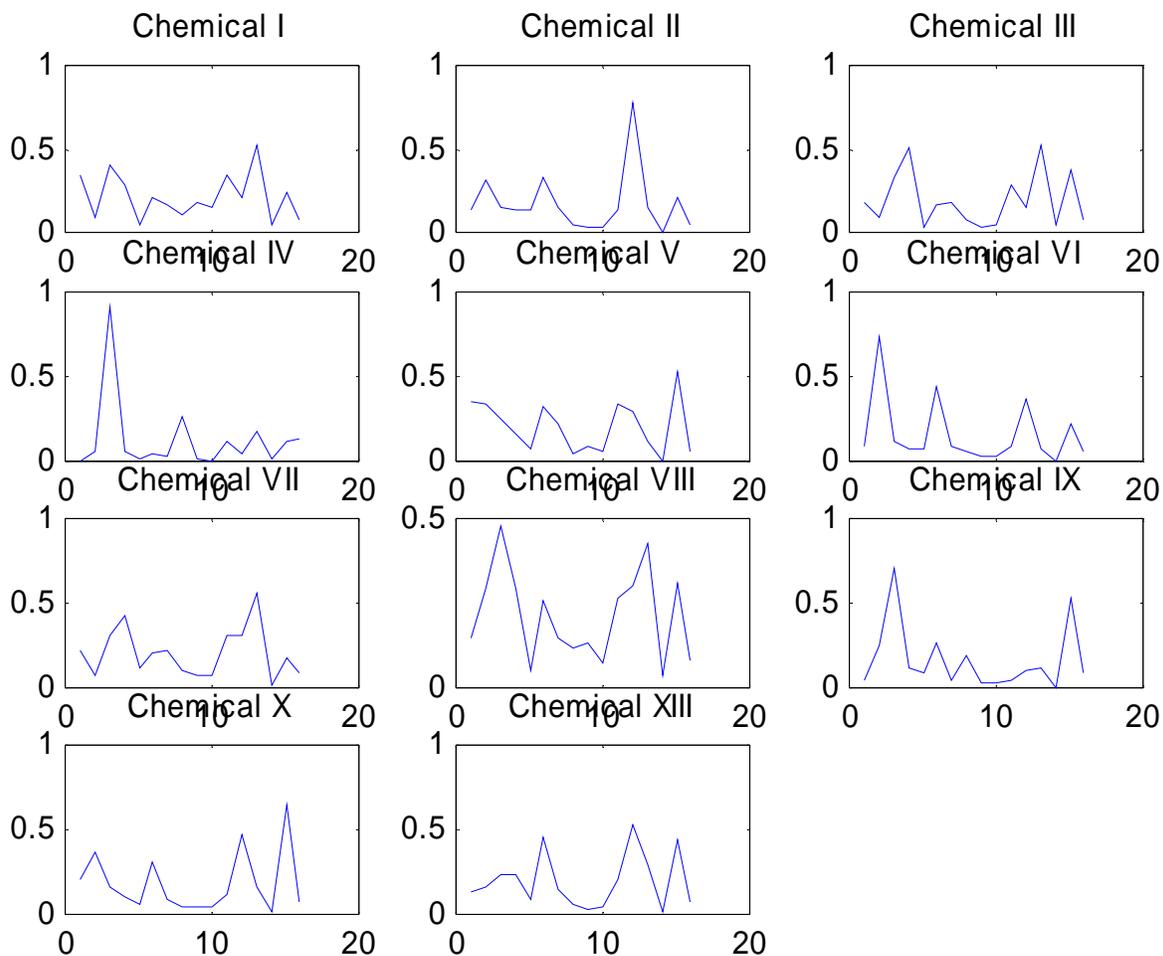


Figure 4: Classes of 11 chemical sources

In this study, the number of sensors used is 16 and mixing chemical sources is 2; we will examine four sets of data, as shown in Table II below:

Table II: *Parameters of chemical mixture*

Data set	Number of variables (sensors)	Mixture	Number of samples available
1	16	1 & 7	12
2	16	3 & 10	9
3	16	2 & 13	8
4	16	2 & 5	8

IV. Simulation results

Based on the data available provided in Table II, we divided this study into two experiments:

- 1) Over complete case: in this case we used complete sample data (see column 4 in Table II) as input to the network and the output size is 2 as two original sources recovered.
- 2) Squared case: this is a straight forward case with 2 inputs and 2 outputs when the data was rearranged so that each input is from the same sensor with non-overlapped consecutive sampling times $(t+ i*\Delta t)$ and $t+(i+1)*\Delta t$.

Experiment 1

In this case, we have studied the four data sets in Table II and we used all data vectors available (The maximum number of data vectors is 12), which is less than the number of sensors (16).

For data set 1, there are 12 mixtures of chemicals 1 and 7 and the data are shown in Figure 5a.

Data Set 1

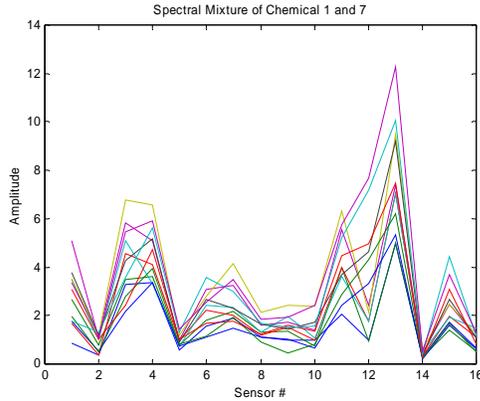


Figure 5a: The mixture data of chemical 1 & 7

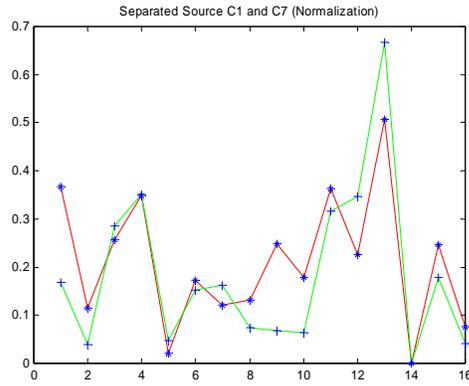


Figure 5b: Separated signal sources 1 and 7 via space invariant ICA of 12 inputs

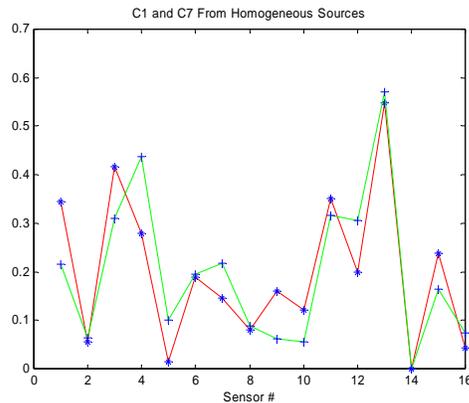


Figure 5c: Original chemical sources 1 and 7 by averaging technique

Using the space invariant ICA approach, the recovered signal sources (chemical 1 and 7 sources) are shown in Figure 5b and the average of the single chemical source 1 and 7 are shown in Figure 5c.

Figures 5b and 5c, show a strong correlation between the chemical 1 and separated chemical named 1 (green traces) and chemical 7 and separated signal named 7 (red).

To confirm its performance, we projected the separated sources 1 and 7 by ICA technique on the known 11 chemical sources shown in Figure 4, the results are provided in Table III.

Table III: The Projection of the separated chemical sources 1 and 7 on the original chemical

Single chemical source	Separated chemical 1	Separated chemical 7
1	0.8457	0.9601
2	0.4772	0.6291
3	0.7325	0.9369
4	0.6286	0.5258
5	0.7433	0.6996
6	0.4872	0.4304
7	0.7944	0.9876
8	0.7938	0.9014
9	0.6679	0.5737
10	0.5271	0.6051
13	0.6223	0.7896

As can be inferred from Table III, the single source of chemical 1 has the greatest overlap with the separated source, labeled separated chemical 1. Similarly, single source chemical 7 has the greatest overlap with separated chemical 7 from the mixture shown in Figure 5a.

Data Set 2

Data set 2, a mixture of chemicals 3 and 10, is plotted in Figure 6a below.

For the mixture from data set 2, the performance of space invariant ICA has demonstrated its effective capability to separate the mixture of chemicals 3 and 10 as shown in Figures 6b as compared with the original sources in Figure 6c.

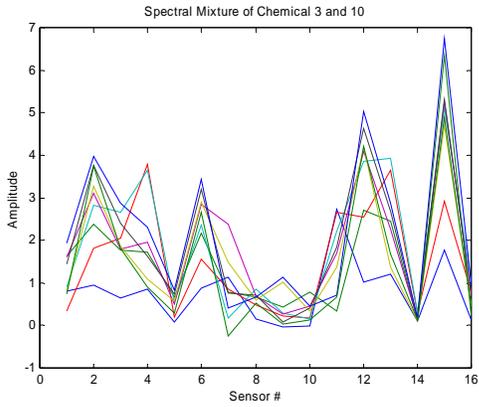


Figure 6a: Mixture data of chemicals 3 & 10

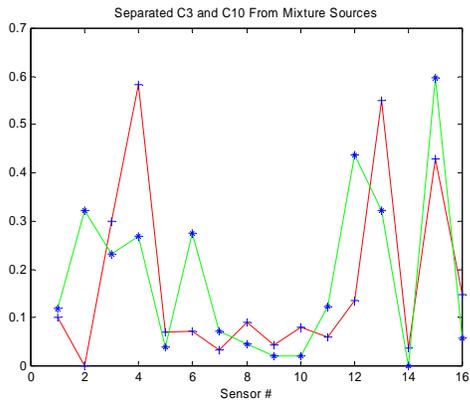


Figure 6b: separated signal sources 3 and 10 via space invariant ICA with 9 inputs and 2 outputs

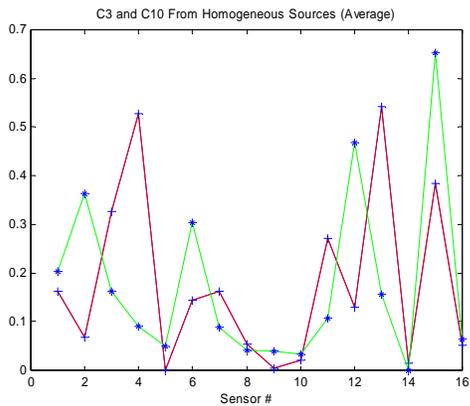


Figure 6c: Original chemical sources 3 and 10 by averaging technique.

Table IV shows the correlation between the original and separated sources of chemical 3 and 10.

Table IV: The Projection of the separated chemical sources 3 and 10 on the original chemical

Single chemical source	Separated chemical 3	Separated chemical 10
1	0.8250	0.7723
2	0.4760	0.8282
3	0.9464	0.8254
4	0.5235	0.4420
5	0.6134	0.8935
6	0.3414	0.7670
7	0.8681	0.7829
8	0.8165	0.8887
9	0.6561	0.7803
10	0.6125	0.9619
13	0.7200	0.9444

Data Set 3

Data set 3, a mixture of chemicals 2 and 13, is plotted in Figure 7a.

For the mixture of chemicals 2 and 13, space invariant ICA has demonstrated its capability to separate the mixture of chemical 2 and 13 as shown in Figure 7b as compared with the original sources in Figure 7c.

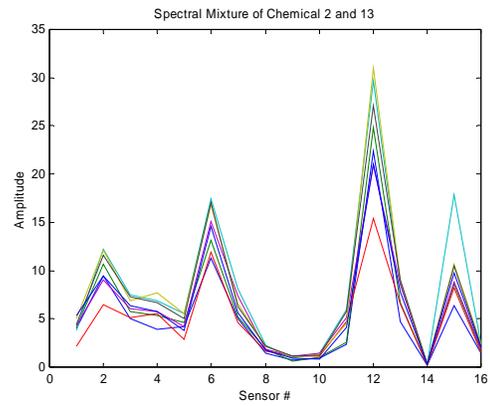


Figure 7a: Mixture data of chemical 2 & 13

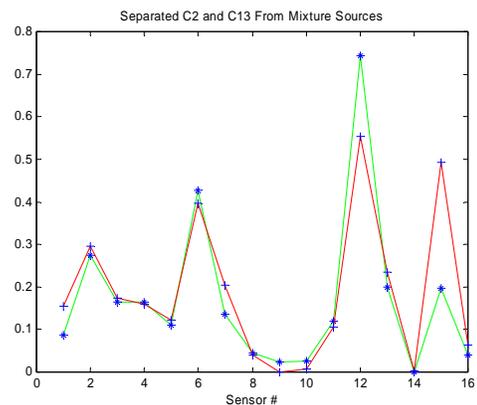


Figure 7b: separated signal sources 2 and 13 via space invariant ICA with 8 inputs and 2 outputs.

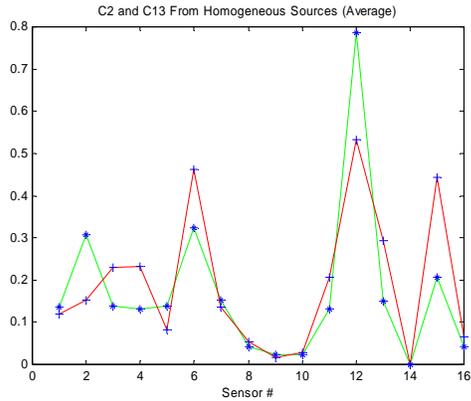


Figure 7c: Original chemical sources 2 and 13 by averaging technique

In Table V it shows the correlation between the original and separated source of chemicals 2 and 13.

Table V: The projection of the separated chemical sources on the original chemical

Single chemical source	Separated chemical 2	Separated chemical 11
1	0.6194	0.6916
2	0.9893	0.9225
3	0.5733	0.6970
4	0.3166	0.3574
5	0.7677	0.9007
6	0.8016	0.8116
7	0.7037	0.7350
8	0.7643	0.8269
9	0.5654	0.7117
10	0.8331	0.9598
13	0.9257	0.9709

Data Set 4

Data set 4, a mixture of chemicals 2 and 5, is plotted in Figure 8a

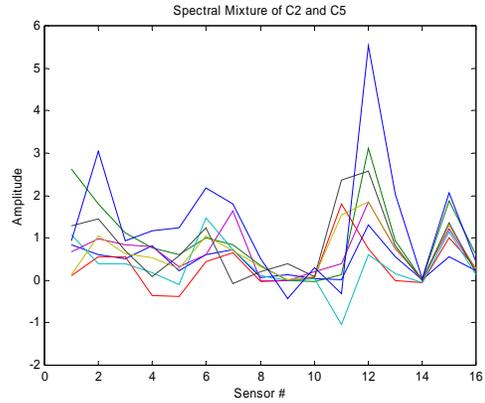


Figure 8a: Mixture data of chemicals 2 & 5

For mixture from data set 4 shown in Figure 8a, space invariant ICA has confirmed its powerful capability to separate the mixture of chemicals 2 and 5 as shown in Figure 8b as compared with the original sources in Figure 8c.

Table V shows the correlation between the original and separated source of chemical 2 and 5.

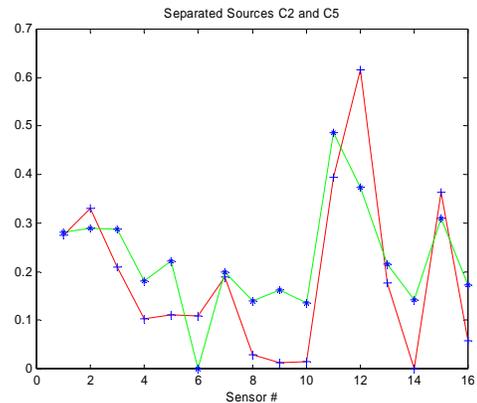


Figure 8b: separated signal sources 2 and 5 via space invariant ICA with 8 inputs and 2 outputs.

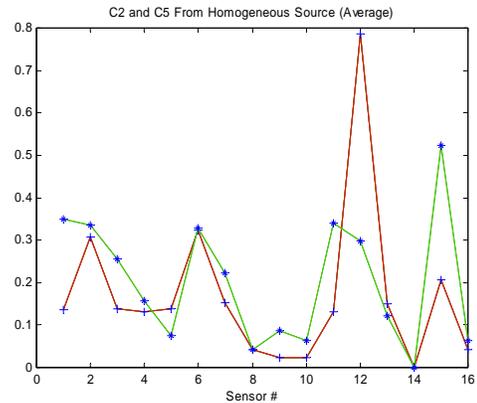


Figure 8c: Original chemical sources 2 and 5 by averaging technique

Table VI: The Projection of the separated chemical sources 2 and 5 on the original chemicals

Single chemical source	Separated chemical 2	Separated chemical 5
1	0.7311	0.8138
2	0.9017	0.7261
3	0.6623	0.7240
4	0.3824	0.5042
5	0.8992	0.8639
6	0.7301	0.6039
7	0.7450	0.7974
8	0.8098	0.8247
9	0.6086	0.6377
10	0.8753	0.7447
13	0.8680	0.7566

Experiment 2

In this experiment, we paired data set 1 (16x12) in columns to obtain the data set 96x2. From this conversion, data values in a single row are the data from the same sensor with consecutive sampling times ($t+i*\Delta t$ and $t+(i+1)*\Delta t$); this new data set allows the same number of mixing sources and of original sources.

Using this new data set, space invariant ICA has produced the results that were validated with the 11 known chemicals. Classification was 100% correct, based on the projection on 11 classes shown in Figure 4.

To simplify the results, we tabulated the mean and standard deviation of its projection (separated sources C1 and C7) on each single chemical source and the results are summarized Table VII below:

Table VII: The mean and standard deviation of its projection of separated sources C1 and C7 on each single chemical source

Single chemical source	Mean Separated chemical 7	Standard deviation	Mean Separated chemical 1	Standard deviation
1	0.8851	0.0830	0.8959	0.0890
2	0.5893	0.1405	0.5551	0.0792
3	0.8728	0.0784	0.8046	0.0397
4	0.4973	0.0723	0.5057	0.1541
5	0.6611	0.0599	0.7285	0.0742
6	0.4161	0.0970	0.4465	0.0903
7	0.9222	0.0765	0.8664	0.1104

8	0.8398	0.0604	0.8141	0.1099
9	0.5539	0.0460	0.5601	0.1031
10	0.5671	0.1103	0.5702	0.1141
13	0.7342	0.1096	0.6902	0.1101

Table VI is compact information to show that the separation sources from the mixture have successfully identified the original chemical sources.

V. Discussions

To separate two (2) mixing sources from a sixteen (16) element sensory data array, known as the over complete case, poses a challenges for mathematical model and network topology. The non overlapped paire-wise (*i.e.* sensor i and sensor $i+1$) or overlapped pair-wise (*i.e.* (sensor i and sensor $i+1$) and (sensor $i+1$ and sensor $i+2$) so on) may face cumbersome and ineffective techniques.

As shown above, the mathematical model has demonstrated space invariant ICA to be an effective topology to overcome insufficient data samples and the over complete case. Moreover, the chemical data itself is fuzzy and inconsistent, and the optimal topology is not answered in this study. By simulation we have shown that chemical source separation problem can be solved effectively with complete time sampling data ($k=12$) (under complete case) and two consecutive sampling data ($k=2$). Optimal topology may require a model of noise in order to determine the size of the sampling input. Moreover, space invariant ICA governed by equation (4) is only valid when the sampling time is sufficiently small. Hence, the sampling time also plays an important role to ensure that the model approach holds.

VI. Conclusions

We have provided a mathematical model to enable the space invariant ICA topology from which Informax and natural gradient descent technique can be applied and simulation has confirmed that our modeling is effective and sufficient to perform chemical source separation to enable the smart ENose.

Further study will be conducted to validate its usefulness for the real world and open environment for chemical detection. In addition, the miniaturized, compact, light weight and low power hardware approach is also a driven force for NASA mission from which System-On-a-Chip approach will be our next focus based on this modeling approach.

Acknowledgements

The research described herein was performed by the Jet Propulsion Laboratory, California Institute of Technology under contract with the National Aeronautics and Space Administration (NASA). The authors would like to thank Drs Thomas Prince and Katherine Dumas for their support and Drs. Vu A. Duong and T. Daud for their insightful discussion and suggestions.

References

- [1] Ryan, M. A.; Homer, M. L.; Buehler, M.G.; Manatt, K. S.; Zee, F.; Graf, J. Monitoring the Air Quality in a Closed Chamber Using an Electronic Nose. *Proceedings of the 27th International Conference on Environmental Systems. Society of Automotive Engineers 1997*, 97-ES84.
- [2] Ryan, M. A.; Homer, M. L.; Buehler, M. G.; Manatt, K. S.; Lau, B.; Karmon D.; Jackson, S. Monitoring Space Shuttle Air for Selected Contaminants Using an Electronic Nose. *Proceedings of the 28th International Conference on Environmental Systems. Society of Automotive Engineers 1998*, 98156.
- [3] Ryan, M. A.; Homer, M. L.; Zhou, H.; Manatt, K. S.; Ryan V. S.; Jackson, S. P. Operation of an Electronic Nose Aboard the Space Shuttle and Directions for Research for a Second Generation Device. *Proceedings of the 30th International Conference on Environmental Systems Society of Automotive Engineers 2000*, 00ICES-259.
- [4] *Spacecraft Maximum Allowable Concentrations for Selected Airborne Contaminants*, National Academy Press:Washington, DC, **1994**; Vols. 1 & 2
- [5] James, J. T.; Limer, T.F.; Leano, H.J.; Boyd, J.F.; Covington, P.A. Volatile Organic Contaminants Found in the Habitable Environment of the Space Shuttle - STS-26 to STS-55. *Aviation Space Environ. Med.* **1994**, 65, 851-857.
- [6] Rumelhart, D. E., and McClelland, J. L. **1986**, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Vol. 1: Foundation*. MIT Press, Cambridge, MA.
- [7] Fahlman, S. E., Lebiere, C. **1990**, "The Cascade Correlation learning architecture," in *Advances in Neural Information Processing Systems II*, Ed: D. Touretzky, Morgan Kaufmann, San Mateo, CA, pp. 524-532.
- [8] Cohen, M. and Grossberg, S. **1983**, "Absolute stability of global pattern formation and parallel memory stage by competitive neural networks," *IEEE Trans. Systems, Man, Cybernetics*, vol. SMC-13, pp. 815-826.
- [9] Widrow, B. **1962**, "Generalization and information storage in networks of ADALINE neurons," Ed: G.T. Yovitt, "Self-Organizing Systems," *Spartan Books*, Washington DC.
- [10] T. A. Duong and Allen R. Stubberud, "Convergence Analysis Of Cascade Error Projection: An Efficient hardware learning algorithm", *International Journal of Neural System*, Vol. 10, No. 3, pp. 199-210, June **2000**.
- [11] J. Herault and C. Jutten. Space or time adaptive signal processing by neural network models. In Denker J. S. (ed), editor, *Neural network for computing: AIP Conference proceedings* 151, New York. American Institute for Physics, **1986**.
- [12] C Jutten and J. Herault. Blind separation of the sources, part I: an adaptive algorithm based on neuromimetic architecture. *Signal Processing*, 24:1-10, **1991**.
- [13] P. Common. Independent component Analysis – a new concepts? *Signal Processing*, 36(3):287-314, 1994.
- [14] D.-T. Pham, P. Garrat and C. Jutten. Separation of a mixtures of independent sources through a maximum likelihood approach. In *EUSIPCO* pages 771-774, **1992**.
- [15] J. Cardoso. High-order contrast for independent component analysis. *Neural Computation*, **1998(a)**.
- [16] A. J. Bell and T. J. Sejnowski. An Information-Maximization Approach to Blind Separation and Blind Deconvolution. *Neural Computation*, 7:1129-1159, **1995**.
- [17] Vu A. Duong, *Independent Component Analysis for Space-Time and Nonlinear Mixtures* Ph.D. Thesis UCI, **2004**.
- [18] S. I. Amari. Natural gradient works efficiently in learning. *Neural Computation*, **1998**.
- [19] A. Bell "Information Theory, Independent-Component Analysis and Applications" in *Unsupervised Adaptive Filtering, Vol. 1 Blind Source Separation* edited by Simon Haykin, JohnWiley & Son, Inc., Chapter 7, **2000**.