# Heterarchical Clustering in Mutually Constraining but Disjunct Feature Spaces

**Wolfgang Fink**
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, CA 91109-8099
*wolfgang.fink@.jpl.nasa.gov*

**Rebecca Castaño**
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, CA 91109-8099
*rebecca.castano@jpl.nasa.gov*

**Ashley Davies**
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, CA 91109-8099
*ashley.davies@jpl.nasa.gov*

**Eric Mjolsness**
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, CA 91109-8099
*mjolsness@jpl.nasa.gov*

## Abstract

We introduce a general type of optimization algorithm that infers data models relating two or more disjunct feature spaces, representing different aspects of a set of objects. The associated inference problem is solved by formulating a novel objective function, generalizing mixture model clustering to heterarchical classification. For the optimization of this objective function we use a clocked objective function update scheme. As a concrete example we apply the clustering algorithm to geological data (rocks) to infer the spatial as well as mineral relationships within a field geology model. We test the algorithm with synthetic data generated according to a particularly chosen probability distribution function.

## 1 Introduction

Clustering algorithms are dependent on the presence of features that distinguish among the data. These features can be mathematically represented as feature vectors. Clustering of feature vectors can be done in a variety of ways such as K-means [Duda, Hart, and Stork, 2000], EM for mixture models [Bishop, 1995], and hierarchical clustering algorithms [Williams, 2000]. Common to these algorithms is the fact that all features are treated equivalently in the joint feature space.

In this work we present a clustering method that allows operation on qualitatively disjunct feature spaces simultaneously and in a unified way [Fink, Castaño, Davies, and Mjolsness, 2001]. In particular we introduce a heterarchical clustering algorithm for mutually constraining heterogeneous features, applicable to a variety of classification problems. Without loss of generality we consider the case of feature vectors consisting of

two heterogeneous feature domains. We will show that this class of "mixed" feature vectors can be successfully clustered by our method where standard clustering algorithms, represented by EM for mixture models, fail.

In Section 2 we explain the underlying theory of our clustering algorithm by giving an example application. We derive the appropriate constrained optimization problem for inferring, in this case, geological relationships from observed data (about rocks) in the form of an objective function, and define the optimization procedure using a clocked objective function update scheme. In Section 3 we show the results of numerical simulations for the example application. We end with conclusions and a discussion about future work in Section 4.

## 2 Theory

The first step of our clustering method is to formulate an objective function which is to be optimized, describing mathematically the clustering problem. In order to optimize the objective function the derivatives with respect to each optimization variable are calculated. In our case clocked objective functions [Gold, Rangarajan, and Mjolsness, 1996] and Soft-assign techniques [Rangarajan, Gold, and Mjolsness, 1996; Rangarajan, Yuille, Gold, and Mjolsness, 1997; Rangarajan, Yuille, and Mjolsness, 1999] are employed for the optimization of the objective function. We will illustrate our algorithm by giving an example drawn from geological planetary surface exploration [Mjolsness, Davies, Castaño, Lou, and Fink, 2000].

### 2.1 Rock-Patch-Facies-Deposit Model

In our current study we look at selected geological processes in a Martian environment. Starting from what a rover can actually observe, we deploy a geological model of how rocks are dispersed from a few sources into many separate locations. Rocks from the same source (e.g., impact crater ejecta, volcanic deposition, flood channels) form deposits. The rocks within each deposit are spatially clustered into patches. In addition to belonging to a patch, each rock also belongs to a class based on its mineral composition and morphology, called a facies. A deposit contains one or more facies. Each patch within a deposit contains the same ratio of rocks from the facies represented in the deposit, e.g., 70% of facies A and 30% of facies B. These relationships between deposits, patches, facies, and individual rocks are summarized in the rock-patch-facies-deposit model (RPFD model), depicted in Figure 1. We stress that this model is a generalized form of classification, relating the distribution of individual clasts to each other to derive the compositional (facies) and spatial (patches) relationship within the deposits under investigation. A classification result can then be compared to other geologic distributions at other locations on Mars or elsewhere.

The distribution of patch locations in a deposit, rock locations in a patch, and mineral composition vectors within a facies are all taken to be Gaussian here for simplicity. Based on this scenario, an objective function can be derived (see 2.2). Observable parameters such as rock location, shape, clast size, and spectra can be used to invert the model, estimating the extent and composition of surface deposits and identifying the corresponding geological formation processes (fluvial, impact, volcanic, aeolian).

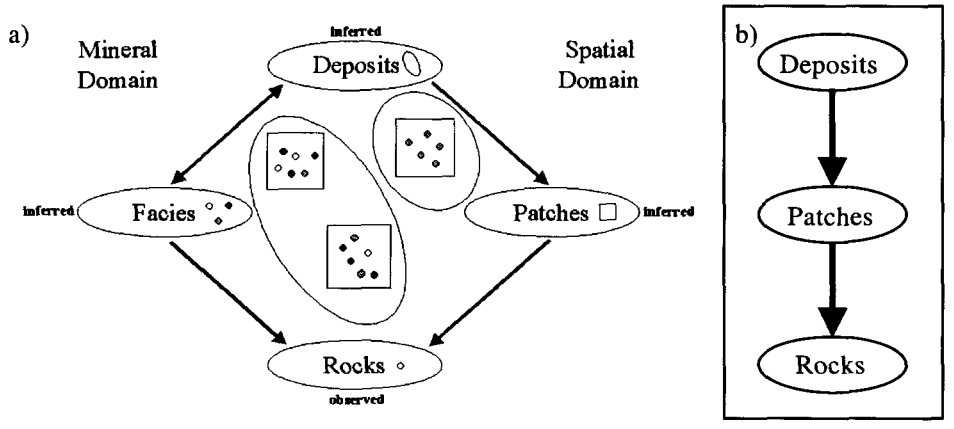


Fig. 1a: RPFD model, describing the relationships between deposits, patches (location), facies (mineral composition and morphology), and individual rocks (location + mineral composition and morphology). Single-headed arrows represent one-to-many relationships, while the double-headed arrow represents a many-to-many relationship. In the center of the figure two deposits are depicted: the first deposit consists of a single patch containing one facies, while the second deposit consists of two patches, containing three facies. Compare this heterarchical clustering to simple hierarchical clustering in Fig. 1b.

## 2.2 Objective Function

From the above described geological model we derive an appropriate constrained objective function for inferring geological relationships from observed data (in this case rocks). The function to be optimized is:

$$E_1 = \sum_{a,b=1} P_{ab}\left(\|\psi_b - y_a\|^2 - \mu_1\right) + \sum_{b,i=1} R_{bi}\left(\|x_i - \psi_b\|^2 - \mu_2\right)$$
$$+\sum_{b,i=1} Q_{li}\left(\|c_i - cf_l\|^2 - \mu_3\right) + \mu_4 \sum_{a,l=1} F_{al} - \mu_5 \sum_{a,b,l,i=1} P_{ab} R_{bi} Q_{li} F_{al}$$

subject to the following constraints:

$$0 \le P_{ab} \le 1, \sum_{a=1}^{A} P_{ab} \le 1; 0 \le R_{bi} \le 1, \sum_{b=1}^{B} R_{bi} \le 1; 0 \le Q_{li} \le 1, \sum_{l=1}^{L} Q_{li} \le 1$$

where $F$: membership matrix of facies $l$ in deposit $a$ (the key many-to-many relationship which expresses the "intertwining" of mineral and spatial information); $P$: membership matrix of patch $b$ in deposit $a$; $R$: membership matrix of rock $i$ in patch $b$; $Q$: membership matrix of rock $i$ in facies $l$; $y_a$: spatial location of center of deposit $a$; $\psi_b$: location of center of patch $b$; $cf_l$: composition vector for facies $l$; $x_i$: observed spatial location of rock $i$; $c_i$: observed composition of rock $i$; $\mu_1, \ldots, \mu_5$: reward/penalty weights.

These constraints are enforced by adding the following entropy $(-TS)$ and Lagrangian-multiplier $(\lambda_b, \nu_i, \gamma_i)$ terms to the objective function:

$$E_2 = T\sum_{a=0}^{A}\sum_{b=1}^{B} P_{ab}\left[\log(P_{ab})-1\right] + T\sum_{b=0}^{B}\sum_{i=1}^{N} R_{bi}\left[\log(R_{bi})-1\right]$$

$$+T\sum_{l=0}^{L}\sum_{i=1}^{N} Q_{li}\left[\log(Q_{li})-1\right] + T\sum_{a,l}\left[F_{al}\log(F_{al}) + (1-F_{al})\log(1-F_{al})\right]$$

$$E_3 = \sum_{b=1}^{B}\lambda_b\left(1-\sum_{a=0}^{A}P_{ab}\right) + \sum_{i=1}^{N}\upsilon_i\left(1-\sum_{b=0}^{B}R_{bi}\right) + \sum_{i=1}^{N}\gamma_i\left(1-\sum_{l=0}^{L}Q_{li}\right)$$

The overall objective function $E$ is the sum of the partial objective functions:

$$E = E_1 + E_2 + E_3 .$$

An essential difference between our algorithm over, e.g., EM for mixtures of Gaussians, is the fourth-order *PRQF* term in $E_1$ which allows information from the mineral clustering (facies) and the two-level spatial clustering subproblems (patches and deposits) to interact and mutually constrain one another.

### 2.3 Clustering Algorithm and Optimization Process

To perform the constrained optimization we use deterministic annealing with clocked objective functions and Soft-assign, converging to a fixed point as shown in the following pseudo-algorithmic excerpt:

```
T = T_max;
energy = EvalEnergy(parameters,T);
while ( T > T_min )                          // deterministic annealing loop
{
  UpdateDepositMeans(parameters);            // y_a
  UpdatePatchMeans(parameters);              // psi_b
  UpdateFaciesMeans(parameters);             // cf_l
  UpdateDepositPatchMemberships(parameters,T); // P
  UpdatePatchRockMemberships(parameters,T);  // R
  UpdateFaciesRockMemberships(parameters,T); // Q
  UpdateDepositFaciesMemberships(parameters,T); // F
  T = T * T_rate;                            // anneal temperature T
  energy = EvalEnergy(parameters,T);
}
```

The necessary update equations for the clustering algorithm are derived in a straightforward calculation not presented here by setting the partial derivatives of the objective function $E$ with respect to each variable equal to 0:

$$\frac{\partial E}{\partial y_{ak}} = \frac{\partial E}{\partial c_{lk}} = \frac{\partial E}{\partial \psi_{bk}} = \frac{\partial E}{\partial P_{ab}} = \frac{\partial E}{\partial R_{bi}} = \frac{\partial E}{\partial Q_{li}} = \frac{\partial E}{\partial F_{al}} = 0 .$$

## 3 Results

To demonstrate the algorithm we show an example using three deposits, nine patches, and three six-dimensional facies.

### 3.1 Generation of Synthetic Data Sets

The rocks together with their respective facies, the patch centers, the facies centers, and the deposit centers are generated from 1-D Gaussian distributions for each dimension (two dimension for x- and y-spatial coordinates, six dimensions for facies). The variances we used are as follows: deposit-center-variances = 10.0, patch-center-variances = 3.0, facies-center-variances = 10.0, rock-location-variances = 0.3, and rock-facies-variances = 1.0.

### 3.2 Simulation Results

Figures 2 and 3 show the simulation results for an example synthetic data set generated using the above variances. In Figure 2a the results of the RPFD-algorithm group the data into the same deposit and patch clusters as the source data labels. Figure 3a shows the clustering results in the mineral domain, where the facies are correctly labeled. The EM algorithm clusters only in the joint spatial and mineral feature space, thus leading to an incorrect clustering result as depicted in Figures 2b and 3b for the spatial and mineral domains, respectively.

## 4 Discussion

We have introduced a clustering method using clocked objective functions and Soft-assign techniques to optimize an appropriately formulated objective function, that allows heterarchical clustering in mutually constraining but disjunct feature spaces. In our study the disjunct features are spatial and mineral features with which the relationships within a geological rock-patch-facies-deposit model are inferred. We demonstrated the algorithm using synthetic data generated according to the rock-patch-facies-deposit data model. We further showed that standard clustering algorithms such as EM fail to cluster correctly in the joint feature space.

Since the optimal choice of the reward/penalty weights $\mu_1,\ldots,\mu_5$ must be determined we used a simulated annealing based algorithm to obtain an optimized set of weights. As criteria for the quality of the weight set we employed two measures: (1) we calculated the normalized sum of the smallest Euclidian differences between the original and the nearest calculated deposit means, patch means, and facies means; (2) we computed a confusion matrix for each type of label (deposit, patch, and facies) and determined the best assignment of original class labels to estimated classes obtained with our algorithm. The score for each label type is given by the percentage of correct rocks.

Since we report on preliminary simulation results, future work would look at the scalability of our method, e.g., increased number of deposits, patches, facies, and rocks and would also examine its performance under more complicated circumstances, such as one deposit being completely embedded in another. In the absence of ground truth information (e.g., knowing the generating means and variances of the involved distributions) cross-validation could be used to determine the optimal reward/penalty weights [Smyth, 1996].

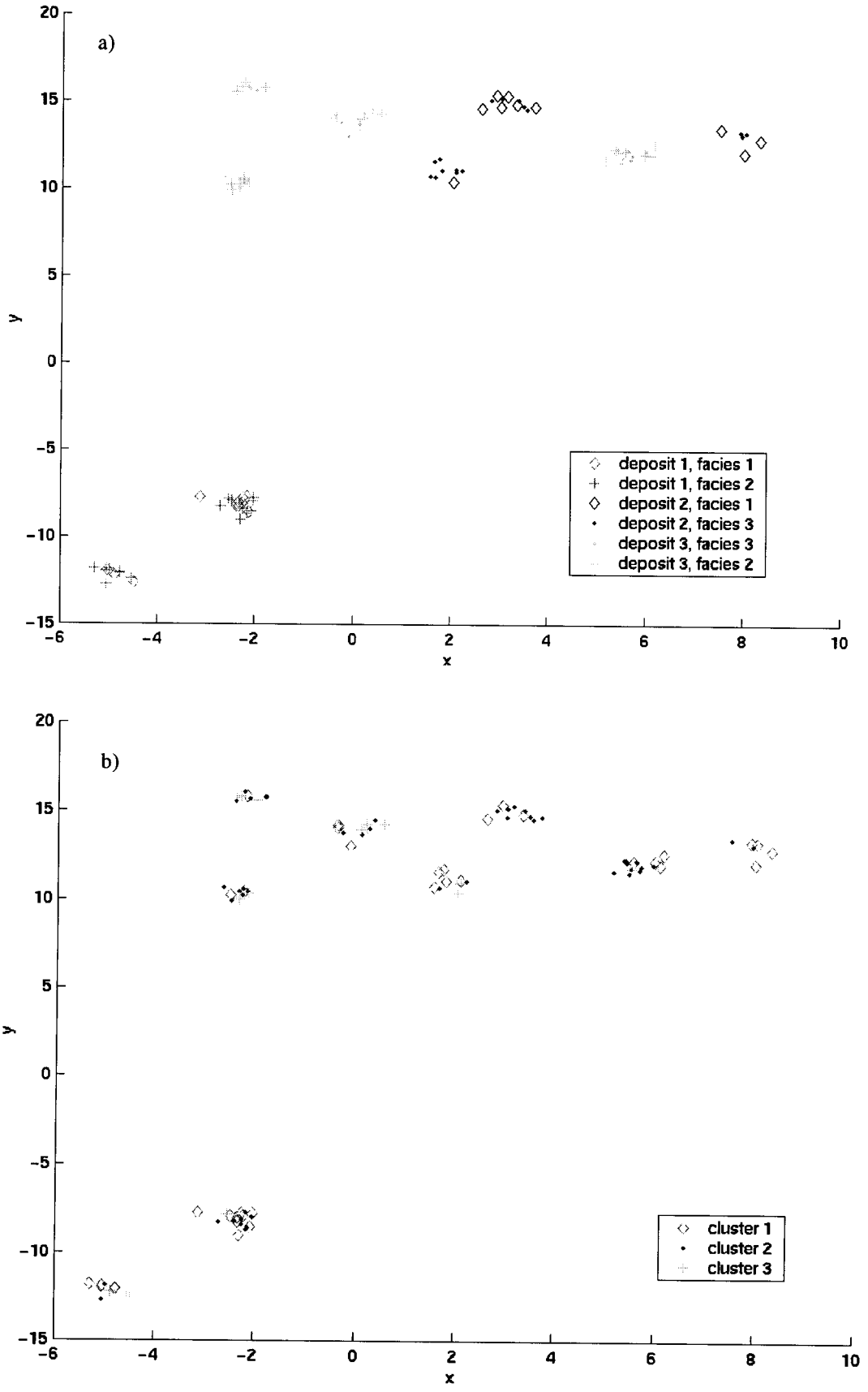


*Figure 2.* a) calculated RPFD-clusters – spatial domain – corresponding with source data labels; b) calculated EM-clusters – spatial domain – incorrectly clustering source data labels.

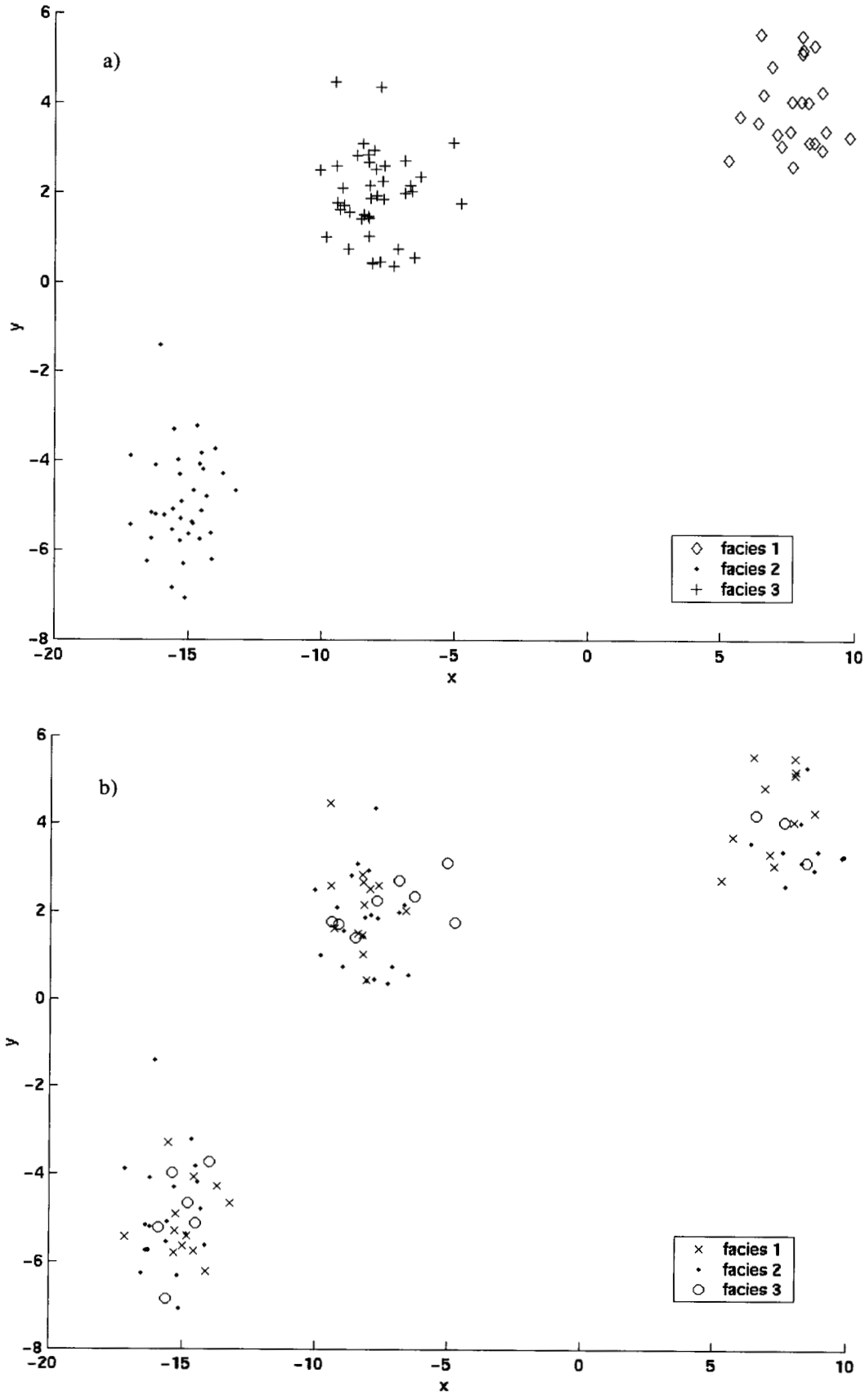


*Figure 3.* a) calculated RPFD-clusters – mineral domain – corresponding with source data labels; b) calculated EM-clusters – mineral domain – incorrectly clustering source data labels. The first two dimensions of the six-dimensional facies feature vectors are plotted.

At a deeper analytical level, that is, for scientific interpretation of the observed (and now classified) facies, mathematical models of physical/geological processes can be used to invert the distribution of materials to create the original (pre-process) distribution and quantify the strength of the process itself. Illustrative examples may be the reconstruction of flood events and the mapping of the ejecta around a simple impact crater. The rock-patch-facies-deposit model allows the different concentrations and ejecta sizes to be put into classes, and the resulting distributions, both mineralogical and physical (e.g., distribution of clast sizes and degree of shock), can be used within a model of crater formation and ejecta emplacement to determine original stratigraphy and mineralogy.

**Acknowledgments**

This work was supported in part by the Intelligent Data Understanding and Automated Reasoning components of the NASA Intelligent Systems Program, and also by the Thinking Systems Area of the Cross Enterprise Technology Development Program (CETDP). This work was performed by the Jet Propulsion Laboratory, California Institute of Technology under contract with NASA.

## References

[1] R. O. Duda, P. E. Hart, D. G. Stork, "Pattern Classification and Scene Analysis", John Wiley & Sons, 2nd edition, 2000.

[2] C. M. Bishop, "Neural Networks for Pattern Recognition", Clarendon Press, Oxford, 1995.

[3] C. K. I. Williams, "An MCMC Approach to Hierarchical Mixture Modelling", Advances in Neural Information Processing Systems 12, S. A. Solla, T. K. Leen, K.-R. Mueller, eds., MIT Press, 2000.

[4] W. Fink, R. Castaño, A. G. Davies, and E. Mjolsness, "Clustering Algorithm for Mutually Constraining Heterogeneous Features", Technical Report JPL-ICTR-01-5, 2001.

[5] S. Gold, A. Rangarajan, and E. Mjolsness, "Learning with Preknowledge: Clustering with Point and Graph Matching Distance Measures", Nerual Computation 8, pp. 787-804, 1996.

[6] A. Rangarajan, S. Gold, and E. Mjolsness, "A Novel Optimizing Network Architecture with Applications", Neural Computation 8, pp. 1041-1060, 1996.

[7] A. Rangarajan, A. Yuille, S. Gold, and E. Mjolsness, "A Convergence Proof for the Softassign Quadratic Assignment Algorithm", *Advances in Neural Information Processing Systems 9.* M. Mozer, M. Jordan, and T. Petsche, eds. MIT Press, 1997.

[8] A. Rangarajan, A. Yuille, and E. Mjolsness. Neural Computation, Convergence Properties of the Softassign Quadratic Assignment Algorithm", Neural Computation 11(6), pp. 1455-1474, 1999.

[9] E. Mjolsness, A. G. Davies, R. Castaño, J. Lou, and W. Fink, "Autonomous Rover-Based Scientific Investigation Using Invertable Mathematical Models", American Geophysical Union (AGU) Meeting, Fall 2000, San Francisco, California, 2000.

[10] P. Smyth, "Clustering using Monte Carlo Cross-Validation", Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining, AAAI Press, 1996.