

AD _____

GRANT NUMBER DAMD17-97-1-7271

TITLE: A Novel Fuzzy Topological Approach to the Detection of Mammographic Lesions and Quantification of Parenchymal Density

PRINCIPAL INVESTIGATOR: Jayaram K. Udupa, Ph.D.

CONTRACTING ORGANIZATION: University of Pennsylvania
Philadelphia, Pennsylvania 19104-3246

REPORT DATE: August 1998

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Distribution authorized to U.S. Government agencies only (proprietary information, Aug 98). Other requests for this document shall be referred to U.S. Army Medical Research and Materiel Command, 504 Scott Street, Fort Detrick, Maryland 21702-5012.

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

DTIC QUALITY INSPECTED 2

NOTICE

USING GOVERNMENT DRAWINGS, SPECIFICATIONS, OR OTHER DATA INCLUDED IN THIS DOCUMENT FOR ANY PURPOSE OTHER THAN GOVERNMENT PROCUREMENT DOES NOT IN ANY WAY OBLIGATE THE U.S. GOVERNMENT. THE FACT THAT THE GOVERNMENT FORMULATED OR SUPPLIED THE DRAWINGS, SPECIFICATIONS, OR OTHER DATA DOES NOT LICENSE THE HOLDER OR ANY OTHER PERSON OR CORPORATION; OR CONVEY ANY RIGHTS OR PERMISSION TO MANUFACTURE, USE, OR SELL ANY PATENTED INVENTION THAT MAY RELATE TO THEM.

LIMITED RIGHTS LEGEND

Contract Number: DAMD17-97-1-7271
Contractor: University of Pennsylvania
Location of Limited Rights Data (Pages): 1-23

Those portions of the technical data contained in this report marked as limited rights data shall not, without the written permission of the above contractor, be (a) released or disclosed outside the government, (b) used by the Government for manufacture or, in the case of computer software documentation, for preparing the same or similar computer software, or (c) used by a party other than the Government, except that the Government may release or disclose technical data to persons outside the Government, or permit the use of technical data by such persons, if (i) such release, disclosure, or use is necessary for emergency repair or overhaul or (ii) is a release or disclosure of technical data (other than detailed manufacturing or process data) to, or use of such data by, a foreign government that is in the interest of the Government and is required for evaluational or informational purposes, provided in either case that such release, disclosure or use is made subject to a prohibition that the person to whom the data is released or disclosed may not further use, release or disclose such data, and the contractor or subcontractor or subcontractor asserting the restriction is notified of such release, disclosure or use. This legend, together with the indications of the portions of this data which are subject to such limitations, shall be included on any reproduction hereof which includes any part of the portions subject to such limitations.

THIS TECHNICAL REPORT HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION.

Patricia Madson

2/12/99

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave blank)

2. REPORT DATE
August 1998

3. REPORT TYPE AND DATES COVERED
Annual (1 Aug 97 - 31 Jul 98)

4. TITLE AND SUBTITLE
A Novel Fuzzy Topological Approach to the Detection of Mammographic Lesions and Quantification of Parenchymal Density

5. FUNDING NUMBERS
DAMD17-97-1-7271

6. AUTHOR(S)
Udupa, Jayaram K., Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of Pennsylvania
Philadelphia, Pennsylvania 19104-3246

8. PERFORMING ORGANIZATION
REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING / MONITORING
AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

19990225 193

12a. DISTRIBUTION / AVAILABILITY STATEMENT

Distribution authorized to U.S. Government agencies only (proprietary information, Aug 98). Other requests for this document shall be referred to U.S. Army Medical Research and Materiel Command, 504 Scott Street, Fort Detrick, Maryland 21702-5012.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

This research focuses on mammographic image processing for the purpose of density quantification, lesion detection and classification. The approaches proposed are different from those taken in the literature in two respects: (1) They emphasize on identifying the dense regions and analyzing their parenchymal architecture. (2) They use a novel fuzzy connectedness method of object definition and image segmentation. During this report period, the following have been accomplished. (1) 120 mammograms from our hospital database have been digitized, converted to the 3DVIEWNIX format, and stored on a medium. More data are being gathered in this fashion. (2) A scale-based fuzzy affinity relation has been devised that is suitable for mammographic image segmentation within the fuzzy connectedness framework. (3) An automatic method has been developed for the segmentation and quantification of parenchymal density. It shows excellent correlation for the measures obtained for the same patient from CC and MLO images. (4) A method of characterizing abnormal parenchymal architecture has been developed. Its utility in detecting and classifying lesions is being investigated.

14. SUBJECT TERMS
Breast Cancer, Mammographic Image Processing, Lesion Detection, Density Quantification

15. NUMBER OF PAGES
24

16. PRICE CODE

17. SECURITY CLASSIFICATION
OF REPORT
Unclassified

18. SECURITY CLASSIFICATION
OF THIS PAGE
Unclassified

19. SECURITY CLASSIFICATION
OF ABSTRACT
Unclassified

20. LIMITATION OF ABSTRACT
Limited

FOREWORD

Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the U.S. Army.

____ Where copyrighted material is quoted, permission has been obtained to use such material.

____ Where material from documents designated for limited distribution is quoted, permission has been obtained to use the material.

____ Citations of commercial organizations and trade names in this report do not constitute an official Department of Army endorsement or approval of the products or services of these organizations.

____ In conducting research using animals, the investigator(s) adhered to the "Guide for the Care and Use of Laboratory Animals," prepared by the Committee on Care and use of Laboratory Animals of the Institute of Laboratory Resources, national Research Council (NIH Publication No. 86-23, Revised 1985).

____ For the protection of human subjects, the investigator(s) adhered to policies of applicable Federal Law 45 CFR 46.

____ In conducting research utilizing recombinant DNA technology, the investigator(s) adhered to current guidelines promulgated by the National Institutes of Health.

____ In the conduct of research utilizing recombinant DNA, the investigator(s) adhered to the NIH Guidelines for Research Involving Recombinant DNA Molecules.

____ In the conduct of research involving hazardous organisms, the investigator(s) adhered to the CDC-NIH Guide for Biosafety in Microbiological and Biomedical Laboratories.



PI - Signature

8/27/98
Date

TABLE OF CONTENTS

Front Cover	1
SF298	2
Foreword	3
Table of Contents	4
Introduction	5
Methods	7
Results	11
Conclusions	12
References	12
Appendices	15

Introduction

Objectives:

This research has the following main aims.

- 1 To develop and implement a fuzzy object definition method for the detection and delineation of parenchymal density, masses and microcalcifications in digitized mammograms.
- 2 To develop and implement a fuzzy object definition method for the classification of lesions and mammographic densities.
- 3 To conduct evaluation studies using histologically verified mammographic data to determine the efficacy of the proposed methods of lesion detection and quantitative classifications.

The focus during the report period has been on the following three tasks:

- 1 Select data sets required for Technical Objective 3, digitize them, transmit via PACS to MIPG facilities and store them in auxiliary storage medium (optical disks) for later use. Recruit Post-Doctoral Fellow.
- 2 Modify current implementation to handle vector-valued features and multivariate affinity functions.
- 3 Implement lesion detection method, verify effectiveness with a few sample data sets and refine method if needed.

Purpose:

Existing studies on false negative mammograms have found patient age, tumor histology and interpretive variability to contribute to false negative diagnosis. However, breast density appears to be the primary cause of missed carcinomas. The radiographic appearance of female breast differs from woman to woman in relation to the amounts of

fat and fibroglandular (connective and epithelial) tissue present. Areas of fat are radiographically lucent while fibroglandular tissues are radiographically dense. There have been many studies looking at the relationship between mammographic density and risk of developing breast cancer. Although a few studies reported no association with increased risk, the majority of studies have found an association between parenchymal patterns and breast cancer risk. A recent meta-analysis [1] of all studies confirms that subjects with mammographic densities have an increased risk of breast cancer relative to those without densities. The risk increases with the density of the breast [2]. The Wolfe classification was proposed many years ago to identify groups of woman at high risk for breast cancer [3]. This scheme was widely used for many years, but has fallen into disuse because of several limitations. For example, inter-observer variability is a problem when the radiologists' subjective assessment is used to classify the amount of density present [4]. Secondly, the magnitude of the increased risk has varied widely in the published studies [1]. Thirdly, identification of this risk factor for a given woman has not altered screening recommendations [5,6]. Recently an computer-assisted user-interactive method to quantify mammographic density has been published [5], which concluded that quantitative classification of densities allows for more specific gradients of risk than do Wolfe's classifications.

An objective, repeatable quantitative measure of risk derived from mammographic densities will be useful in recommending alternative screening process. An architecture dependent quantitative analysis of the mammographic densities will make the screening process more effective. Image processing efforts toward this goal seem to be very sparse in the literature, and automatic and efficient methods for generating this measure do not seem to exist.

The focus of this research is on utilizing and extending recently developed fuzzy connectedness method [17] to fulfill the main objectives. This methodology has been successfully applied in several applications including multiple sclerosis lesion detection [7-11], MR angiography [12] and craniofacial soft tissue imaging [13]. The approach of integrating density-derived information (total density and architecture) into the lesion detection method will hopefully further improve this accuracy.

Scope:

Computer-assisted analysis of mammographic density would provide an objective, quantitative measure of cancer risk factor. This measure will be useful in total risk analysis in several ways. First, such risk analysis could influence the choice of alternative screening paradigms such as intervals between mammograms or use of other modalities such as MRI. Second, this measure could be useful in selecting a group of women for whom the risk-benefit ratio of a potentially toxic preventive measure, such as tamoxifen, would be favorable [14,15]. Third, this measure could be used to signal the need for more careful interpretation of a subset of mammograms. For example, double-reading might be indicated for mammograms above a certain level of density. Fourth, a variety of computer-assisted techniques continue to be developed for mammographic lesion detection. No single algorithm is optimal for all mammograms. Objective density quantification could be helpful for selecting the most appropriate computerized method, or as we propose here, for tailoring a selected algorithm. Clearly, an automatic, accurate, and objective method for density quantification will allow the study of the effect of this variable on the implementation of mammograms.

Methods

Task 1:

A postdoctoral fellow has been recruited for this project. This research utilizes only existing image data. No patients are imaged expressly for the purpose of this project. We have so far digitized (at a resolution of 100 microns) existing mammograms which consists of 39 normal cases, 43 masses and 42 calcifications from our patient database in the hospital. These images have been transmitted to our (MIPG) facility, converted to our local format (the IM0 format of 3DVIEWNIX [18], the software platform used in the project), and stored on a medium. We continue to get more digital data.

Task 2:

We have spent a considerable amount of time investigating the affinity relation that are appropriate for segmenting digitized mammograms. A novel scale-based fuzzy affinity and connectedness method has been developed, implemented in 3DVIEWNIX [18] and tested. A manuscript has been submitted [16] on this topic. A brief summary of its approach is given below.

Fuzzy affinity between two nearby pixels is a reflexive and symmetric relation whose strength lies between 0 and 1 and indicates how the pixels locally “hang together” in the image. The notion of fuzzy affinity is expressed as a nondecreasing function of fuzzy adjacency (dependent only on the distance between pixels), their homogeneity and their agreement to some global intensity-based object property or feature. In determining the homogeneity and feature-based components of affinity between two pixels p_1 and p_2 , a neighborhood around both p_1 and p_2 are considered. The size of this neighborhood, called scale, is not fixed but depends on the size of the largest homogeneous region locally. The scale is first computed automatically for all pixel in the image.

Fuzzy connectedness between any two (not necessarily nearby) pixels p_1 and p_2 is a fuzzy relation whose strength lies between 0 and 1. It is determined by considering all possible paths between p_1 and p_2 . A path is simply a sequence of nearby pixels. A strength of connectedness is assigned to each path which is simply the smallest affinity of successive pixels along the path. The strength of connectedness between p_1 and p_2 is the largest of the strengths of all possible paths between p_1 and p_2 .

In one way, we can imagine that two pixels are strongly connected to each other if there is a path between them through locally homogeneous regions. We call it *homogeneity based connectedness*. In another way, the connectedness between two pixels could be imagined as their likeliness to fit the object features and unlikeliness to fit the background features. We call it *feature based connectedness*. One may note that feature based connectedness does not have the notion of path homogeneity and thus fails to overcome the effects of slowly varying components over the image. On the other hand, homogeneity based connectedness suffers from the fact that in many applications objects merge to background so smoothly that always there is a path from object to background

through locally homogeneous regions; this is more pronounced when blurring or partial voluming is high. Moreover, especially in a thin branch of an object, often, there are small regions with high inhomogeneity that stop the homogeneity based paths for the rest of the branch. Therefore, the two different notions of connectedness are combined. Additionally, the notion of scale allows the correct estimation of homogeneity and object features that is insensitive to noise.

A detailed mathematical formulation of scale-based fuzzy affinity, connectedness, the associated algorithms, and their validation on both 2D and 3D clinical images and phantoms is presented in [16]. A careful statistical evaluation indicates that the scale-based method is much superior to the original method described in [17]

Task 3:

We have devised and implemented several methods for characterizing abnormal parenchymal architecture using the scaled-based connectedness. These are currently being tested and evaluated for their efficacy in identifying and characterizing lesions.

In addition, a major part of Task 6 (stated in the proposal) has been completed. This task has to do with developing a method to automatically quantify parenchymal density and to validate it. We felt it was more natural to accomplish this task, once the fuzzy connectedness method is in place. Besides, we also felt that the segmentation and quantification of parenchymal density would be useful in lesion detection and characterization. A brief summary of this work is given below. A manuscript on this work is now completed [20] and being readied for journal submission.

Segmentation of breast from background: At the very beginning, using 3DVIEWNIX [18] supported LIVE-WIRE [19] tool, regions corresponding to pectoral muscles are interactively excluded when those are projected in the image. In the entire process, this is the only step requiring operator intervention. Fuzzy connectivity is used as the underlying technique in segmenting breast from background. To apply the fuzzy connectivity model, we need to estimate different parameters. Studying all the 120 images from all the 60 patients, we found that the intensity histogram always contains a highly prominent peak at the lower intensities; and that peak is mostly contributed by background. The first

prominent peak in the intensity histogram is detected and used to (roughly) calculate mean and standard deviation of background intensity. To apply the fuzzy connectivity algorithm, we need to select a set of reference (seed) pixels. For this purpose, we assume that the rightmost column in the image always lies in background and include all these pixels as the reference set. Fuzzy connectedness processing starting from those pixels gives us a fuzzy connectivity image for background. We discard connectivity strengths in the upper half and keep the lower half as the breast region. Figures (a) and (b) show an original digitized mammogram and the segmented breast.

Fuzzy connectivity image for glandular tissue: Fuzzy connectivity method is used to enhance glandular dense regions and to suppress fat tissues; the resulting fuzzy connectivity image, in turn, is used for automatic segmentation of glandular region from fat. The major task in applying the fuzzy connectivity model is to estimate the parameters of the algorithm and to select the set of reference pixels. After ignoring the upper 0.01 percentile intensities, the mean and standard deviation parameters (for the homogeneity and feature-based affinity) are estimated from those part of the breast region falling in the upper 25% of the intensity range. Finally, the pixels in the breast region falling in the upper 15% of the intensity range are selected as reference pixels. The fuzzy connected image for the glandular tissue is computed (Figure). In the next paragraph we describe an automatic threshold selection method for quantifying density from this image.

Automatic threshold selection: For any threshold, the image is divided into two regions. Local homogeneity based affinity between every pair of spatially adjacent pixels is used to define their likeliness of belonging to the same object or of not belonging to the same object. The optimum threshold is determined from the associated statistics called threshold energy. The threshold with the minimum threshold energy is selected as the optimum threshold. The segmented binary image for our example is shown in Figure (d).

Finally, we generate several descriptors from the segmented and the original image to quantify the glandular tissues as described below. Note that all steps are completely automatic except the exclusion of pectoral muscles if they are included in the mammogram.

Results

The method has been tested on over 60 studies (each study produces two digitized mammograms) from routine exams from two projections (CC and MLO). The images were scanned on a Lumisys scanner at a resolution of 100 microns. The population included normal as well as cancer cases (masses and calcifications). Except the exclusion of pectoral muscles, the entire method has worked automatically on all images wherein all parameters required by the algorithms are selected automatically. The algorithms produced visually acceptable segmentations in all 120 images. From the segmented regions and the image intensities in them, we compute a set of density related parameters including total glandularity(TG), TG/total fat(TF), TG/average fat(AvF), TG/area of breast(AB), area of glandularity(AG), AG/area of fat(AF), AG/AB. TG and TF are computed by integrating radiographic intensity over respective segmented regions while AG, AB and AF are computed by counting the number of pixels in respective regions; finally, AvF is computed by dividing TF by AF. The correlation between the parameters from the two projections (CC and MLO) is studied. The correlations for TG, TG/TF, TG/AvF, TG/AB, AG, AG/AF and AG/AB are 0.967, 0.902, 0.951, 0.944, 0.959, 0.915 and 0.941, respectively; the plots are shown in Figures 2.(a)–(g). In all figures, the horizontal axis represents the estimated parameter from the CC projection while the vertical axis represents the estimate from the MLO projection.

Conclusions

- 1 The new scale-based fuzzy connectedness method is more robust and effective than the original method [17]. It is very effective for mammographic image segmentation.
- 2 The method seems to be appropriate for characterizing architectural abnormalities and for lesion detection. Its evaluation for these purpose is currently underway.

- 3 Glandularity is considered to be one of the strongest factors for breast cancer. Automatic breast glandularity quantification from digitized mammograms is practical using the proposed method. The method removes the subjectivity inherent in interactive threshold selection techniques currently used. The ability of the computed glandularity parameters in evaluating risk is currently being investigated.

References

- 1 E. Warner et. al., "The risk of breast-cancer associated with mammographic parenchymal patterns: A meta-analysis of the published literature to examine the effect of method of classification", *Cancer Detection and Prevention*, **16**(1):67-72, 1992.
- 2 A. M. Oza and N. F. Boyd, "Mammographic parenchymal patterns: A marker of breast cancer risk", *Epidemiologic Reviews*, **15**(1):196-208, 1993.
- 3 J. N. Wolfe, "Risk for breast cancer development determined by mammographic parenchymal pattern", *Cancer*, **37**(5):2486-92, 1976.
- 4 M. Moscowitz, P. Gartside and C. McLaughlin, "Mammographic patterns as markers for high-risk benign breast disease and incident cancers", *Radiology*, **134**(2):293-5, 1980.
- 5 N. F. Boyd et. al., "Quantitative classification of mammographic densities and breast cancer risk: Results from the Canadian National Breast Screening Study", *Journal of the National Cancer Institute* **87**(9):670-675, 1995.
- 6 J. N. Wolfe, A. F. Saftlas and M. Salane, "Mammographic parenchymal patterns and quantitative evaluation of mammographic densities: A case control study", *American Journal of Roentgenology*, **148**(6):1087-92, 1987.
- 7 J. K. Udupa, L. Wei, Y. Miki, and R. I. Grossman, "A system form comprehensive analysis of multiple sclerosis lesion load based on MR imagery", *SPIE Proceeding*, **3031**:610-618, 1997.
- 8 J. K. Udupa, L. Wei, S. Samaraesekera, Y. Miki, M. A. van Buchem, and R. I. Grossman, "Multiple sclerosis lesion quantification using fuzzy connectedness principles", *IEEE Medical Imaging*, **16**(5):598-609, 1997.
- 9 Y. Mike, R. I Grossman, S. Samarasekera, J. K. Udupa, M. A. van Buchem, B. S. Cooney, S. N. Pollack, D. L. Kolson, M. Polansky, L. J Mannon, "Clinical correlation of computer assisted enhancing lesion quantification in multiple sclerosis", *American Journal of Neuroradiology*, **18**:705-710, 1997.

- 10 Y. Miki, R. I Grossman, J. K. Udupa, L. Wei, D. L. Kolson, L. J Mannon, "Isolated U-fiber involvement in MS: Preliminary observations", *Neurology*, **50**:1301-1306, 1998.
- 11 A. Kumar, W. Bilker, J. K. Udupa, G. Gottlieb, "Late onset minor and major early evidence for common neuroanatomical substrates detected by using MRI", proceedings of the *National Academy of Science*, **95**:7654-7658, 1998.
- 12 J. K. Udupa, D. Odhner, J. Tian, G. Holland, and L. Axel, "Automatic clutter-free volume rendering for MR angiography using fuzzy connectedness", *SPIE Proceeding*, **3034**:111-119, 1997.
- 13 J. K. Udupa, J. Tian, D. C. Hemmy, P. Tessier, "A Pentium-based craniofacial 3D imaging and analysis system", *Journal of Craniofacial Surgery*, **8**(5):333-339, 1997.
- 14 S. M. Lippman, T. L. Bassford and F. L. Meyskens, Jr., "A quantitatively scored cancer-risk assessment tool: Its development and use", *Journal of cancer Education*, **7**(1): 15-36, 1992.
- 15 R. T. Chlebowski et. al., "Breast cancer chemoprevention. Tamoxifen: Current issues and future prospective", *Cancer* **72**(3):1032-7 Supplement, 1993.
- 16 P. K. Saha, J. K. Udupa, and D. Odhner, "Fuzzy connected object definition; A scale based formulation", submitted to *Graphical Models Image Processing*.
- 17 J. K. Udupa, and S. Samarasekera, "Fuzzy Connectedness and Object Definition: Theory, Algorithms, and Applications in Image Segmentation", *Graphical Models and Image Processing*, **58**(3):246-261, 1996.
- 18 J. K Udupa, D. Odhner, S. Samarasekera, R. J. Goncalves, K. Iyer, K. Venugopal, S. Furuie, "3DVIEWNIX: A open, transportable, multidimensional, multimodality, multiparametric imaging system, proceedings of *SPIE*, **2164**:58-73, 1994.
- 19 A. X. Falcao, J. K. Udupa, S. Samarasekera, S. Sharma, "User-steered image segmentation paradigms: Live wire and live lane", *Graphical Models Image Processing* **60**:233-260, 1998.
- 20 P. K. Saha, J. K. Udupa, E. F. Conant, D. P. Chakraborty, "Automatic segmentation of parenchymal density in digitized mammograms", *under preparation*.

Appendix

Figure Captions:

Figure 1: (a) A digitized mammographic image. (b) Segmented breast region. (c) Connectivity image for glandular tissue. (d) Segmented dense regions.

Figure 2: Graphical representations for different parameters computed from two different projections (CC and MLO). In all graphs, the x-value represents the estimated parameter from CC projection while the y-value represents the estimated parameter from MLO projection. (a) Total glandularity(TG). (b) TG/total fat(TF). (c) TG/average fat(AvF) (d) TG/area of breast(AB). (e) Area of glandularity(AG). (f) AG/area of fat(AF). (g) AG/AB.

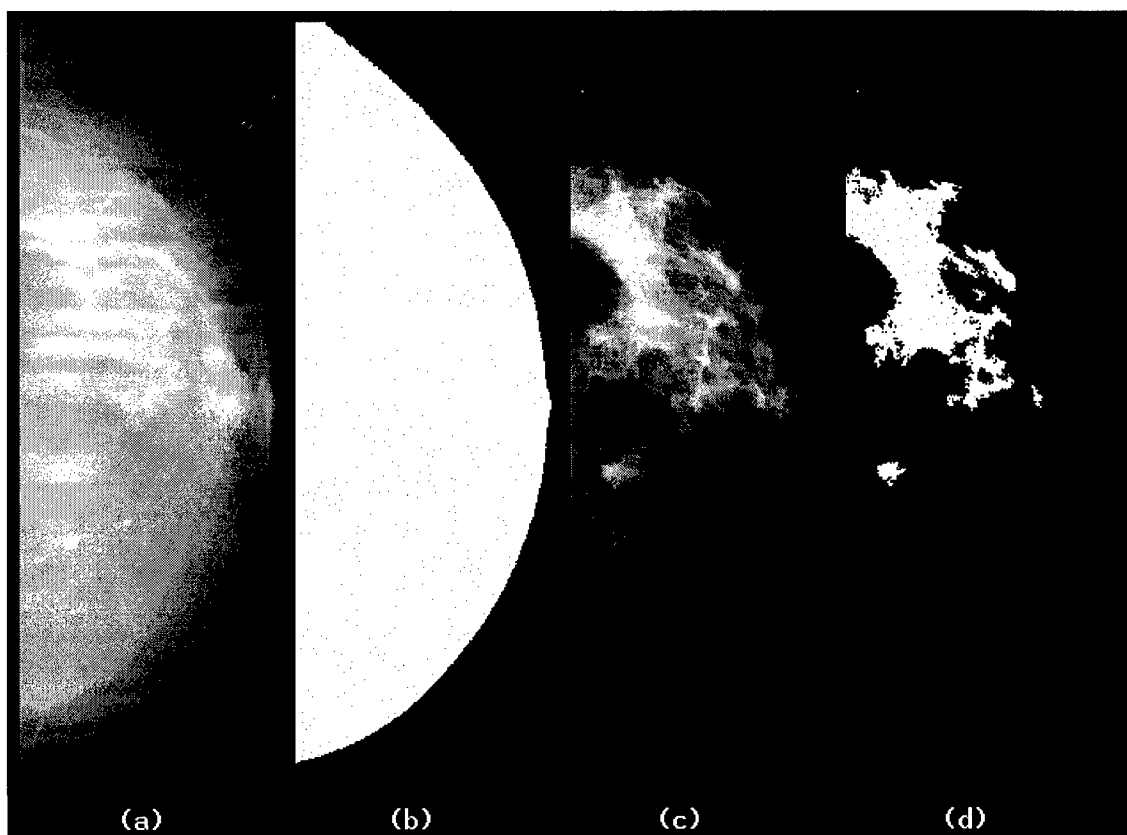


Figure 1

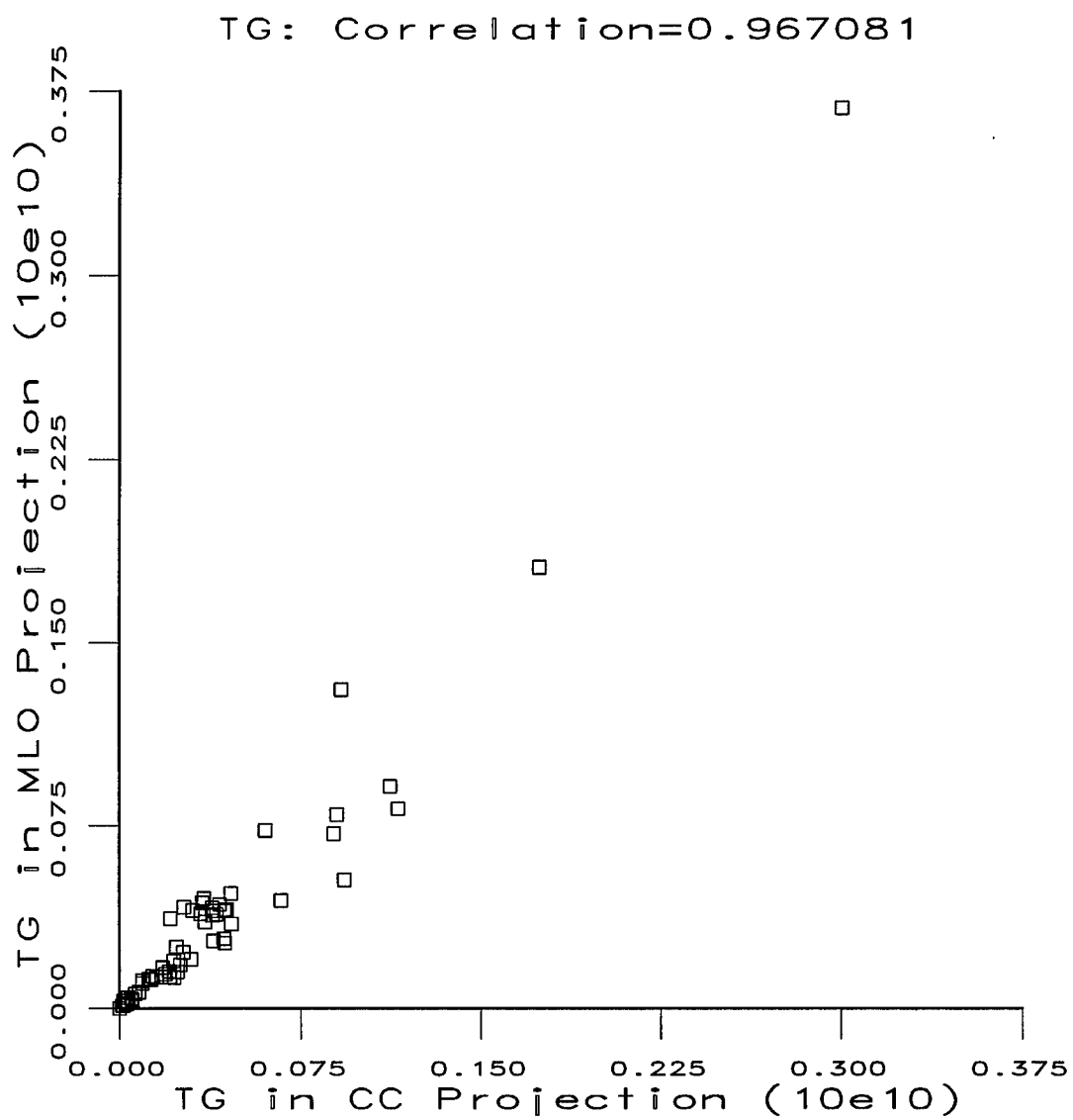


Figure 2.(a)

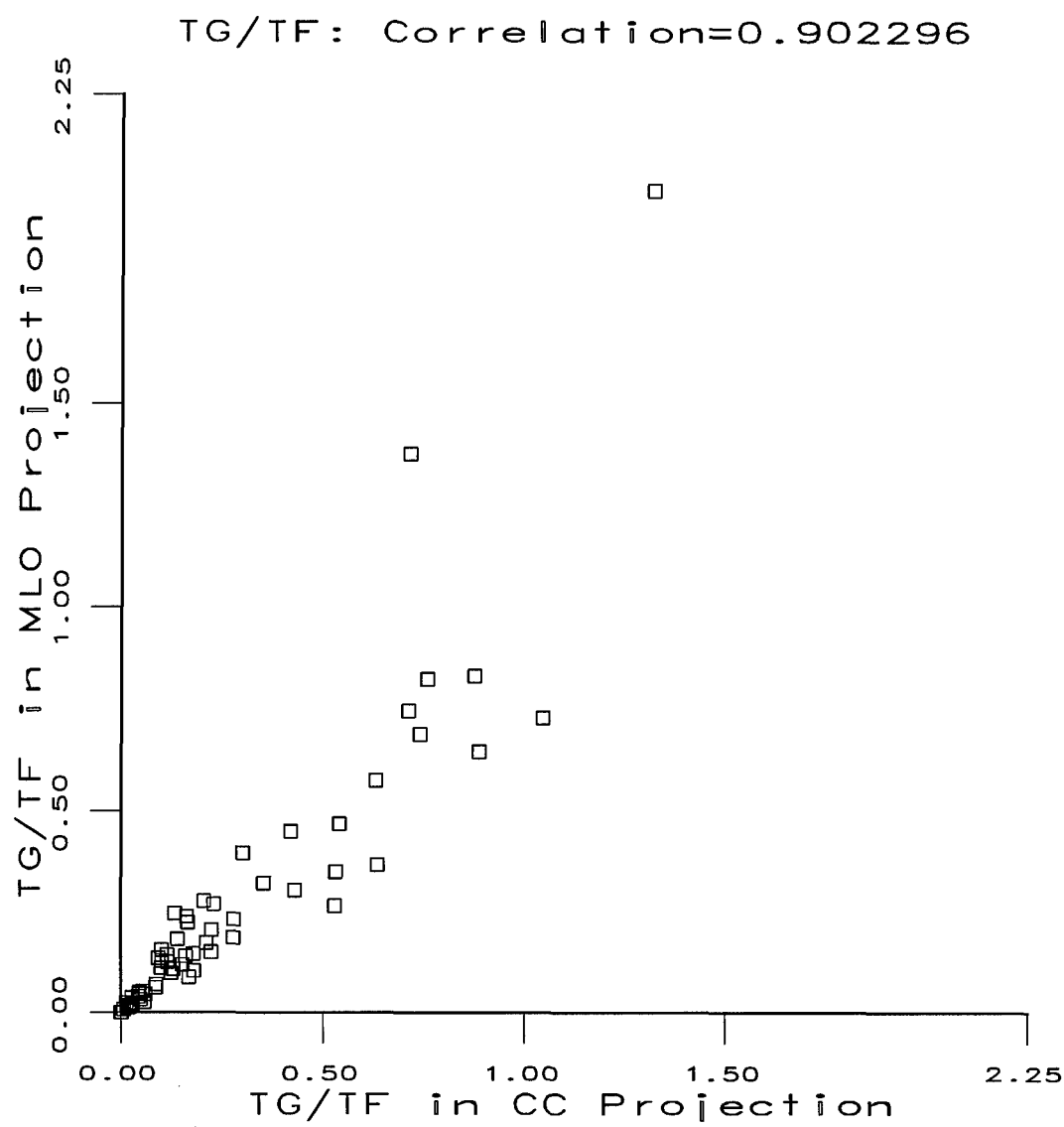


Figure 2.(b)

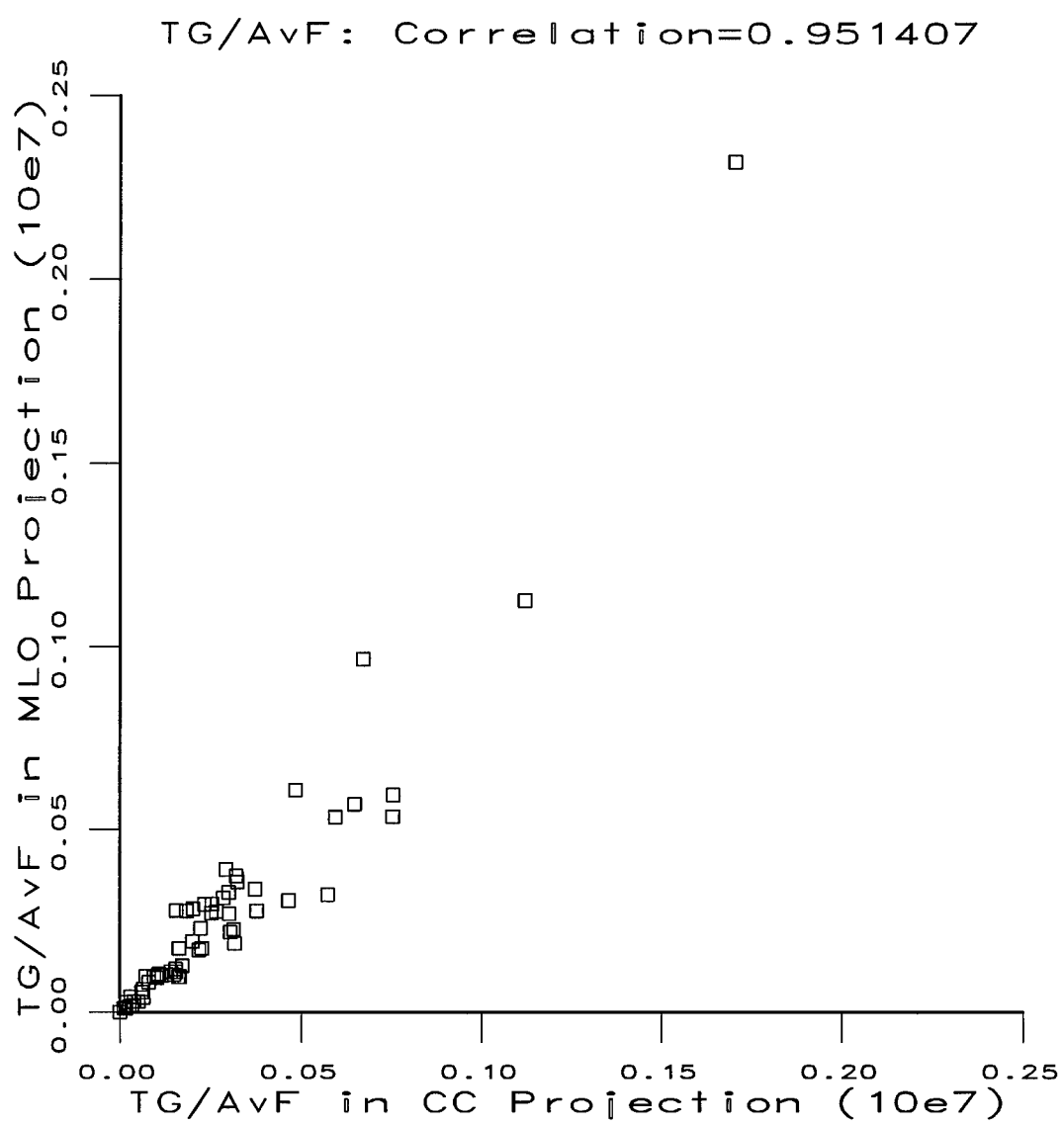


Figure 2.(c)

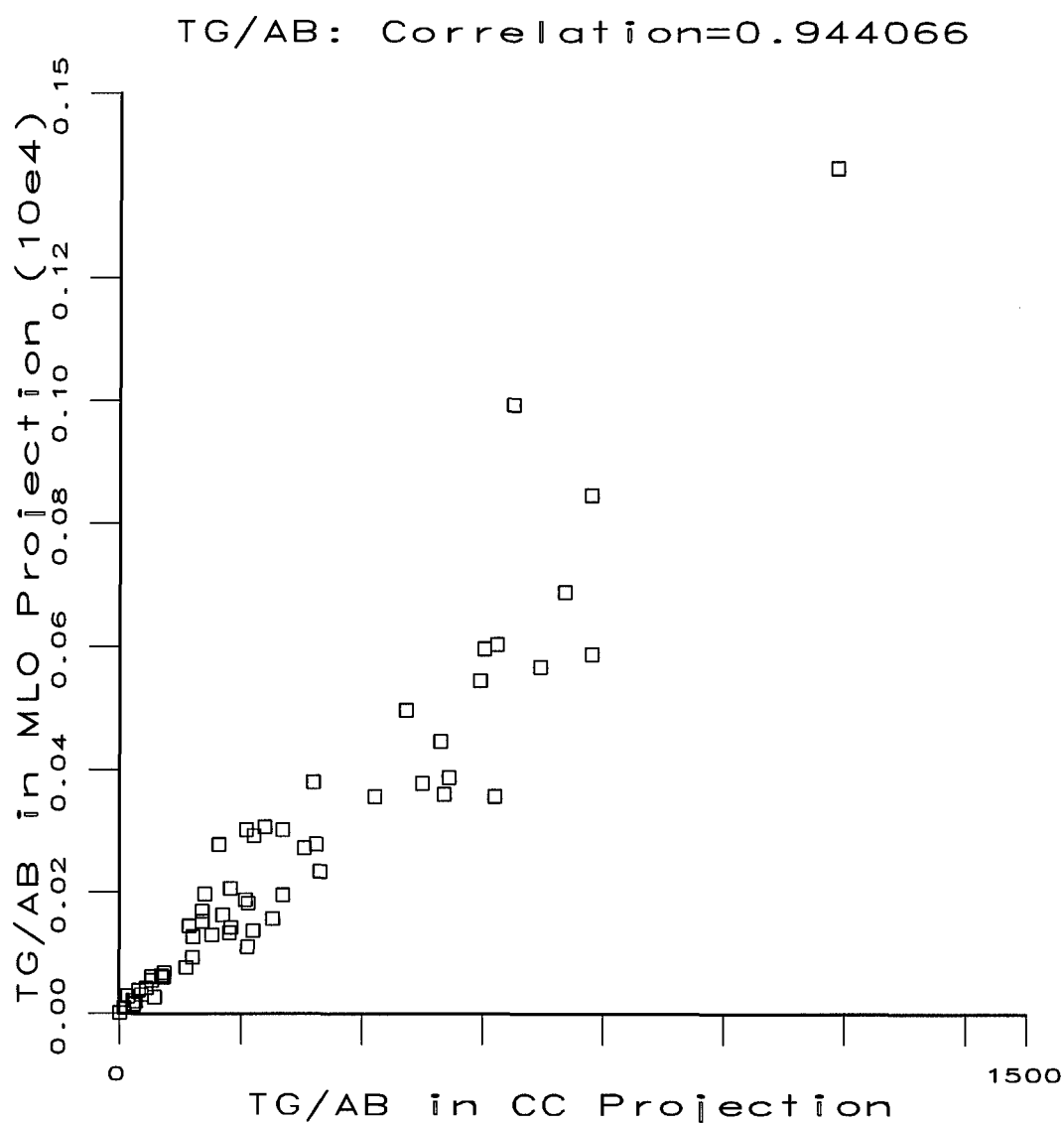


Figure 2.(d)

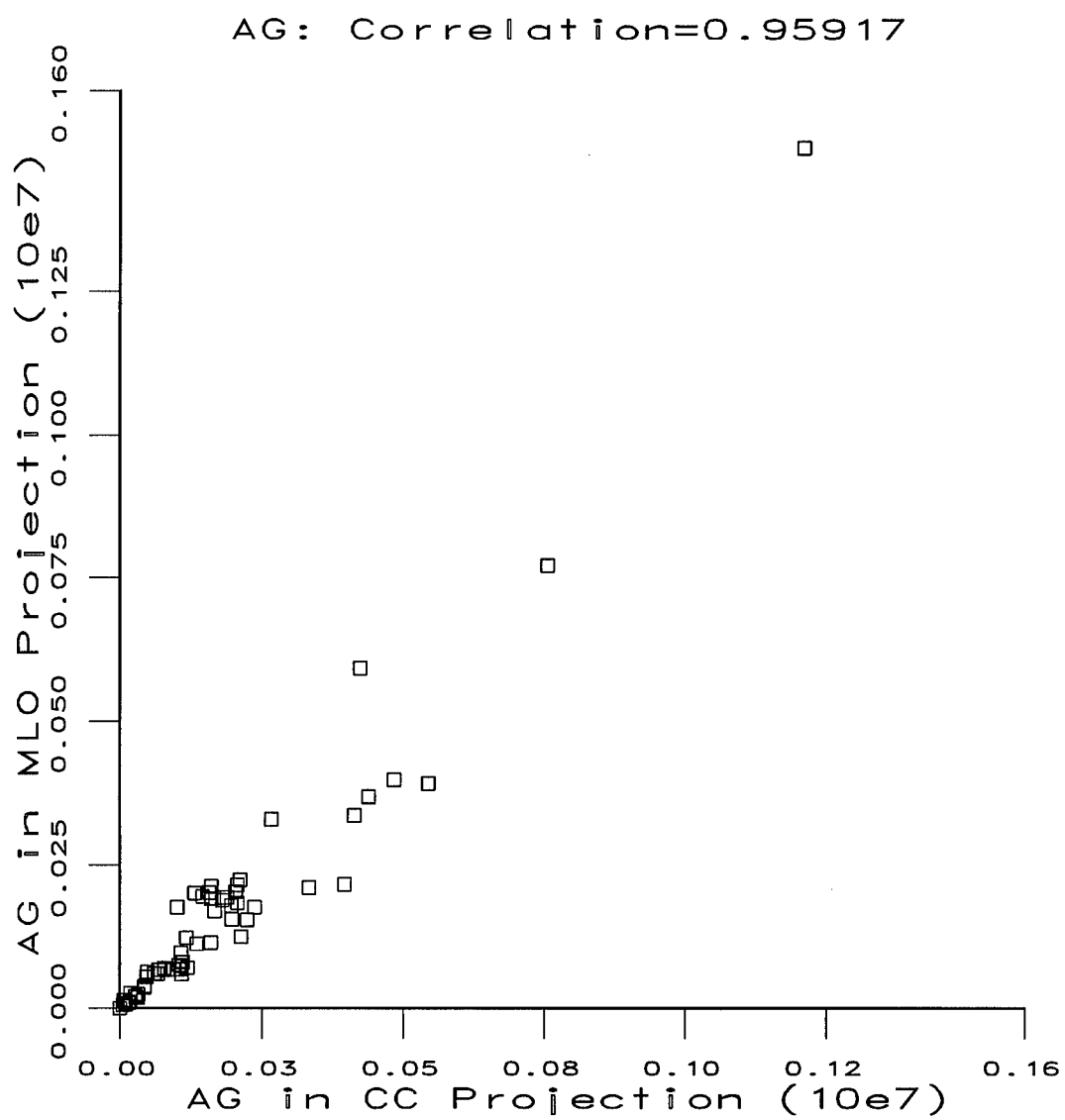


Figure 2.(e)

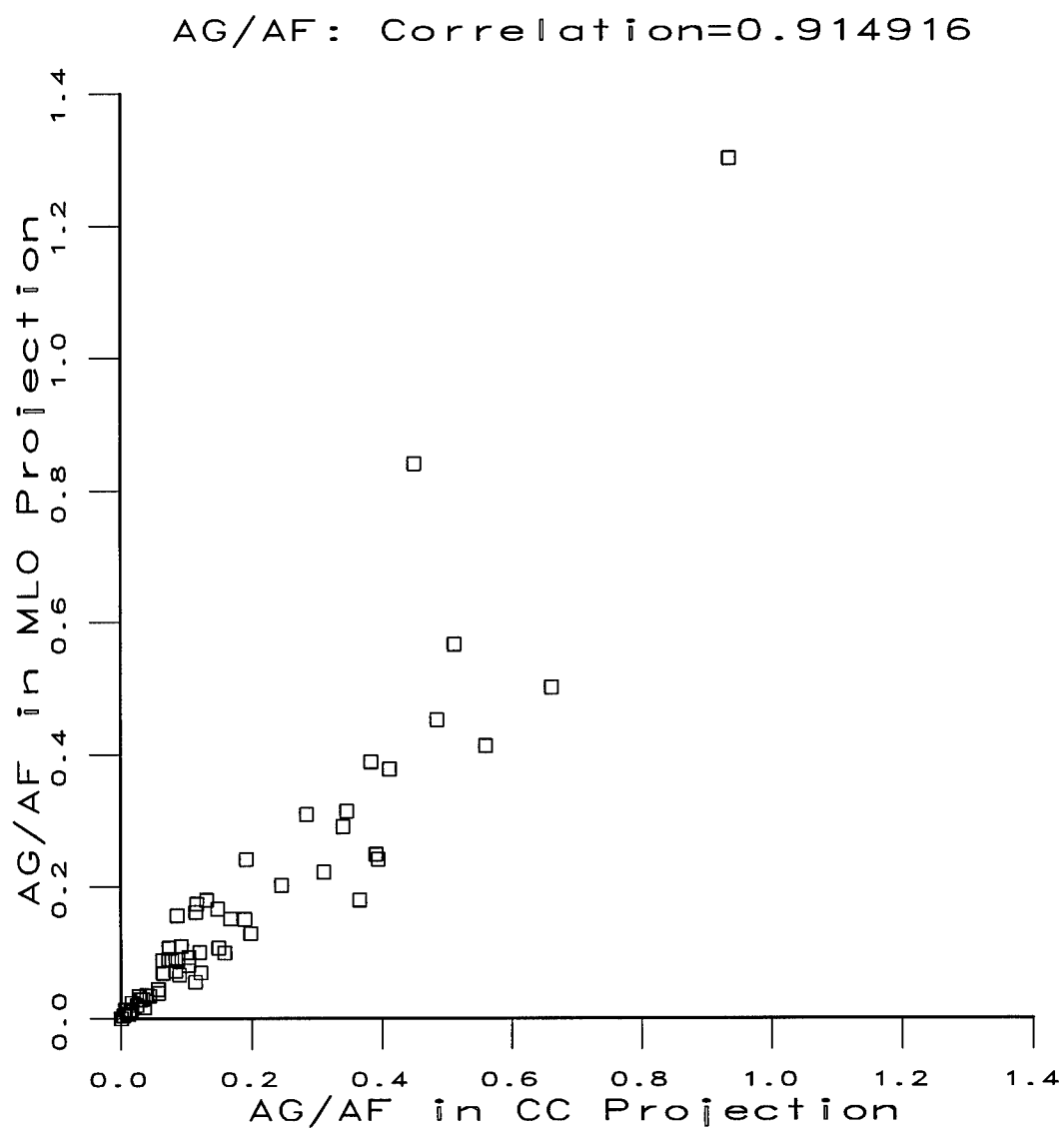


Figure 2.(f)

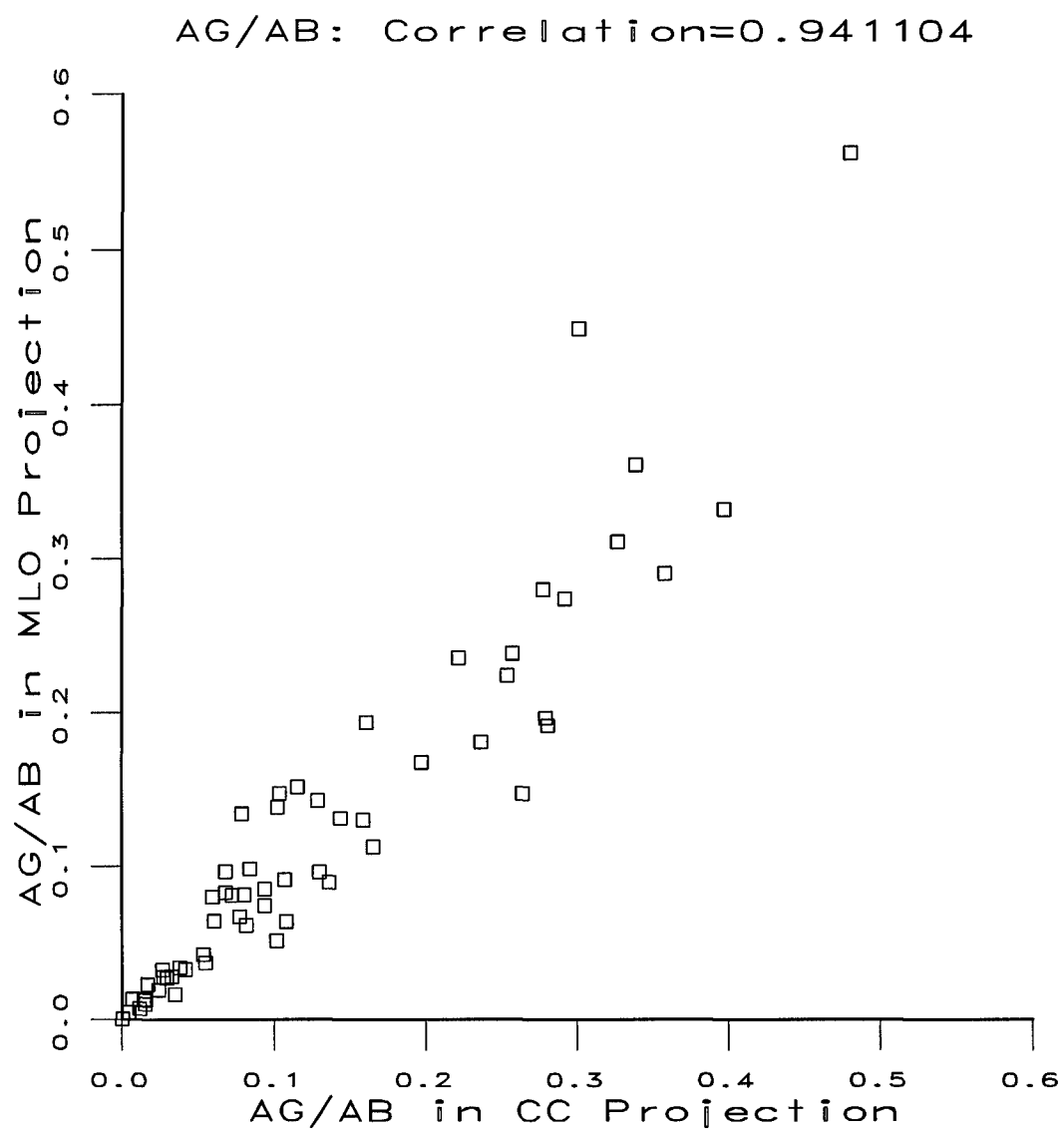


Figure 2.(g)