PRRC 07-05

A Customizable Fuzzy Expert System for Regional and Local Play Analysis

Second Annual Report

for the period

January 1, 2006 through March 31, 2007

by
Robert S. Balch, Principal Investigator
Ronald F. Broadhead, Co-Principal Investigator

March 2007

Work Performed Under Contract No. DE-FC26-04NT15512

Prepared for
U.S. Department of Energy
Jim Barnes, Project Manager

Prepared by
Reservoir Evaluation and Advanced Computing Technology Group
New Mexico Petroleum Recovery Research Center
New Mexico Institute of Mining and Technology
Socorro, NM 87801

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

ABSTRACT

The objective of this project is to create a user definable and customizable fuzzy expert system tool to dramatically speed local and regional play analysis and to reduce subsequent drilling risk. This general tool will not require significant knowledge of computer programming, and will guide users through the process of building a successful expert system to evaluate plays from field to basin scale using public and/or private data and their own or public data and knowledge. To demonstrate the effectiveness of the tool, a secondary objective of analyzing the Pennsylvanian play of southeastern New Mexico is being performed. Public data has been organized for an analysis of this outstanding, bypassed-pay play, which will provide an example of the usage of the system while simultaneously providing a significant opportunity for identifying new reserves.

Four major tasks are to be performed in order to accomplish the project goals. Task 1 is the development of a customizable fuzzy expert system (CFS), with the ability to self-generate software and fully customize integral components of the expert system for non-computer programmers. Task 2 is the development of interfaces to simplify the goals of Task 1. The creation of wizards to aid in the work flow process will make expert system development both quicker and easier. In Task 3 the system will be validated using the Pennsylvanian carbonate play of SE New Mexico. Development of this bypassed pay play will aid in the creation of wizards to assist users in Task 2. An ongoing effort in Tech Transfer comprises Task 4, as the ultimate test of the successful program will require that it is widely available and utilized.

TABLE OF CONTENTS

DISCLAIMER ii

ABSTRACT iii

TABLE OF CONTENTS iv

List of Figures v

List of Tables xii

EXECUTIVE SUMMARY 1

REPORT DETAILS 2

1. Introduction 2
2. Expert System Development 3
3. Geologic Data Acquisition and Analysis 71
4. Experimental Methods 133
5. Technology Transfer 133
6. Conclusions 135

REFERENCES 137

List of Figures

TABLE OF CONTENTS

DISCLAIMER ii

ABSTRACT iii

TABLE OF CONTENTS iv

List of Figures v

List of Tables xii

EXECUTIVE SUMMARY 1

REPORT DETAILS 2

1. Introduction 2
2. Expert System Development 3
3. Geologic Data Acquisition and Analysis 71
4. Experimental Methods 133
5. Technology Transfer 133
6. Conclusions 135

REFERENCES 137

List of Figures

Fig. 1.1. Gantt chart of project tasks. 3

Fig. 2.1. CFS overall interface with all modules active. 5

Fig. 2.2 . Stages in the development of a CFS knowledgebase. 7

Fig. 2.3. Algorithm pseudocode: distance-based data cleaning. 9

Fig. 2.4. Overview of the Statistics Module. 14

Fig. 2.5. Calculation of P(x) 15

Fig. 2.6. K-S statistic for normal data. 17

Fig. 2.7. K-S statistic for not-normal data. 17

Fig. 2.8. Histogram for normal data. 18

Fig. 2.9. Illustration of k-s statistic for normal data. 18

Fig. 2.10. Histogram for not-normal data. 19

Fig. 2.11. Illustration of k-s statistic for not-normal data. 19

Fig. 2.12. Fuzzy curves for *normal* data. 20

Fig. 2.13. Fuzzy curves for *not-normal* data having two non-adjacent high frequency bins. 22

Fig. 2.14. Fuzzy curves for *not-normal* data having two adjacent high frequency bins. 23

Fig. 2.15. Fuzzy curves for *not-normal* data having one high frequency bin. 24

Fig. 2.16. Fuzzy curves for *not-normal* data with no high frequency bin. 24

Fig. 2.17. Fuzzy curves for *normal* data. 25

Fig. 2.18. Fuzzy curves for the not-normal data having adjacent high frequency bins. 26

Fig. 2.19. Fuzzy curves for not- normal data having two nonadjacent high frequency bins. 27

Fig. 2.20. Fuzzy curves for the Not-normal data having one high frequency bin. 28

Fig. 2.21. Screenshot of the variable select panel. 29

Fig. 2.22. Screenshot of the Fuzzy Set Operations Panel. 30

Fig. 2.23. Screenshot of the Fuzzy Curves Input Panel. 31

Fig. 2.24. Screenshot of the Fuzzy Sets Input Panel. 31

Fig. 2.25. Screenshot of the Add Fuzzy Sets Input Panel. 32

Fig. 2.26. Screenshot of the Delete Fuzzy Sets Input Panel. 32

Fig. 2.27. Data management subsystem interface. 35

Fig. 2.28. Data management subsystem architecture. 36

Fig. 2.29. Performance analysis sequential search. 38

Fig. 2.30. Performance analysis binary search. 39

Fig. 2.31. Index files with a gap of 25, 50, 100 and 800. 40

Fig. 2.32. Index files search time for different set of values. 42

Fig. 2.33. Data management subsystem, binary data file. 42

Fig. 2.34. Data Retrieval module, By Grid option. 44

Fig. 2.35. Data Retrieval module output. 45

Fig. 2.36. Data Retrieval module, By Latitude/Longitude option. 45

Fig. 2.37. Data Retrieval module, All Data result. 46

Fig. 2.38. Data Modification module, Add Variable option. 47

Fig. 2.39. Data Modification module, Add Variable screen. 48

Fig. 2.40. Data management subsystem, binary data file after Add. 48

Fig. 2.41. Data management subsystem, Update option. 50

Fig. 2.42. Data management subsystem, Update Variable screen. 50

Fig. 2.43. Data management subsystem, Data File Selection. 51

Fig. 2.44. Data management subsystem, Delete Variable screen. 52

Fig. 2.45. Data management subsystem, binary data file after Delete.....52

Fig. 2.46. Binary to text conversion – Confirmation Dialog. 53

Fig. 2.47. Binary to text conversion—Notification Dialog. 53

Fig. 2.48. Use of map data in knowledgebase generation for the CFS.....55

Fig. 2.49. Algorithm: Inverse Distance Weighted (IDW). 57

Fig. 2.50. Application screen with Project Management application running.....64

Fig. 2.51. Sample screen for loading an existing project.....64

Fig. 2.52. Browse data application.....65

Fig. 2.53. Select file screen.....65

Fig. 2.54. Browse data application.....66

Fig. 2.55. Application screen for managing fuzzy variables. 66

Fig. 2.56. Displayed fuzzy curves for a defined variable. 67

Fig. 2.57. Set definition screen for fuzzy variables. 67

Fig. 2.58. Interface management screen. 68

Fig. 2.59. In this application, text of questions can be entered. 68

Fig. 2.60. Sample output for entered questions. 69

Fig. 2.61. Inference parameter setup application introduction screen. 69

Fig. 2.62. Boolean logic can be used to define inference parameters.....70

Fig. 2.63. Defined weights of variables for a sample projects.....70

Fig. 3.1. Location of project areas in New Mexico. 73

Fig. 3.2. Location of project areas in southeastern New Mexico in relationship to boundaries of the Permo-Pennsylvanian carbonate reservoir oil play and major tectonic elements. Shown in gray are oil reservoirs that have produced more than 1 million bbls oil from Upper Pennsylvanian and Lower Permian carbonate reservoirs. Modified from Broadhead and others (2004).....76

Fig. 3.3. Stratigraphic chart of the Northwest Shelf of the Permian Basin emphasizing the productive Bough carbonate reservoirs. The Bough C member (highlighted in yellow) is the major productive Bough member within the intrashelf project area. Canyon strata contain the major productive reservoirs within the shelf-margin project area with secondary production provided by reservoirs in Cisco strata. 77

Fig. 3.4. Typical well log through the Bough members within the Bough intrashelf project area. From Broadhead (1999a).....78

Fig. 3.5. Reservoirs productive from Upper Pennsylvanian and Lower Permian carbonates in the Bough intrashelf project area.....81

Fig. 3.6. Location of the area selected for more detailed analysis of the Bough formation. The detailed work in this area was undertaken by graduate student Jake Sharp. 82

Fig. 3.7. Location of wells used for geologic and production analysis in the detailed project area. 83

Fig. 3.8. Contour map of cumulative oil production from wells productive from the Bough A, Bough B, Bough C, and Bough D zones in the detailed project area. 84

Fig. 3.9. Contour map of cumulative oil production from wells productive from the Bough C zone in the detailed project area. 85

Fig. 3.10. Locations of cross sections and wells utilized in constructing the cross sections in the detailed study area.....87

Fig. 3.11. Isopach map of the Bough A member, detailed project area. Contours in feet.....88

Fig. 3.12. Isopach map of the Bough B member, detailed project area. Contours in feet.....89

Fig. 3.13. Isopach map of the Bough C member, detailed project area. Contours in feet.....90

Fig. 3.14. Isopach map of the Bough A member, detailed project area with boundaries of reservoirs productive from Upper Pennsylvanian and Lower Permian strata superimposed. Contours in feet. 91

Fig. 3.15. Isopach map of the Bough B member, detailed project area with boundaries of reservoirs productive from Upper Pennsylvanian and Lower Permian strata superimposed. Contours in feet. 92

Fig. 3.16. Isopach map of the Bough C member, detailed project area with boundaries of reservoirs productive from Upper Pennsylvanian and Lower Permian strata superimposed. Contours in feet. 93

Fig. 3.17. Isopach map of Bough A member. Dots indicate wells used to construct the map. Red symbols indicate wells productive from the Bough A; black symbols indicate well that are not productive from the Bough A. Contours in feet.....94

Fig. 3.18. Isopach map of Bough B member. Dots indicate wells used to construct the map. Red symbols indicate wells productive from the Bough B; black symbols indicate well that are not productive from the Bough B. Contours in feet.....95

Fig. 3.19. Isopach map of Bough C member. Dots indicate wells used to construct the map. Red symbols indicate wells productive from the Bough C; black symbols indicate wells that are not productive from the Bough C. Contours in feet. 96

Fig. 3.20. Structure contours on top of Abo Formation (Lower Permian), Bough intrashelf project area.....98

Fig. 3.21. Structure contours on top of the Mississippian System, Bough intrashelf project area.....99

Fig. 3.22. Structure contours on top of Abo Formation (Lower Permian), Bough intrashelf project area with boundaries of reservoirs productive from Upper Pennsylvanian and Lower Permian strata superimposed. Contours in feet. Datum = sea level. 100

Fig. 3.23. Structure contours on top of the Mississippian System, Bough intrashelf project area with boundaries of reservoirs productive from Upper Pennsylvanian and Lower Permian strata superimposed. Contours in feet. Datum = sea level. 101

Fig. 3.24. Structure contour map of Bough A member. Dots indicate wells used to construct the map. Red symbols indicate wells productive from the Bough A; black symbols indicate wells that are not productive from the Bough A. Contours in feet. Datum = sea level.....102

Fig. 3.25. Structure contour map of Bough B member. Dots indicate wells used to construct the map. Red symbols indicate wells productive from the Bough B; black symbols indicate wells that are not productive from the Bough B. Contours in feet. Datum = sea level.....103

Fig. 3.26. Structure contour map of Bough C member. Dots indicate wells used to construct the map. Red symbols indicate wells productive from the Bough C; black symbols indicate wells that are not productive from the Bough C. Contours in feet. Datum = sea level.....104

Fig. 3.27. Reflection seismic line across a Permo-Pennsylvanian phylloid algal mound, showing thinning of Upper Pennsylvanian and Lower Permian strata over a paleostructure. From Cys (1986).....107

Fig. 3.28. Isopach map of the stratigraphic interval between the top of the Abo Formation (Lower Permian) and the top of the Mississippian System, Bough intrashelf project area.....108

Fig. 3.29. Location of Late Pennsylvanian paleostructural positive elements as determined from the Abo-Mississippian isopach map.....109

Fig. 3.30. Abo-Mississippian isopach map superimposed on a wireframe relief map of Mississippian structure, Bough intrashelf project area.....110

Fig. 3.31. Wireframe relief map of Mississippian structure, Bough intrashelf project area.....110

Fig. 3.32. Isopach map of the Abo-Mississippian interval, Bough intrashelf project area with boundaries of reservoirs productive from Upper Pennsylvanian and Lower Permian strata superimposed.....111

Fig. 3.33. Wells used in mapping and analyzing the detailed study area within the Bough intrashelf project area. Wells with cores are shown as larger symbols. 112

Fig. 3.34. Photomicrograph of biomicrite facies, transmitted light. Darker areas are micrite and lighter areas are calcite cement. Slide is vertically bisected by recrystallized phylloid algal blade with preserved wall structure. Field of view is 1.8 mm. From Sharp (2006).....113

Fig. 3.35. Photomicrograph of biosparite facies, transmitted light. Darker areas are micrite and lighter areas are sparry calcite cement. Recrystallized phylloid algal blade is in center of slide. Field of view is 0.9 mm. From Sharp (2006).....114

Fig. 3.36. Photomicrograph of dolomitized facies, transmitted light. Lighter areas are subhedral dolomite crystals and darker areas are crystal faces. The blue areas are intercrystalline pores filled with blue epoxy. Field of view is 0.9 mm. From Sharp (2006).....115

Fig. 3.37. Oil and gas reservoirs that are productive principally from Upper Pennsylvanian strata within the shelf-margin project area. See Table 3.3 for reservoir data. 117

Fig. 3.38. Production history curve of the Dagger Draw reservoirs, showing periods of discovery and initial development of a seemingly small reservoir, and final full redevelopment that revealed the true size of this very large oil reservoir. After Broadhead (1999a).....120

Fig. 3.39. Structure contour map on top of productive dolostone shelf-margin carbonate buildup facies, Dagger Draw and Dagger Draw South reservoirs indicating wells drilled during phases of (1) discovery and initial development, and (2) redevelopment. Structure contours from Reddy (1995).....121

Fig. 3.40. Depositional model of shelf-margin phylloid algal mounds and intramound facies at shelf margin, Dagger Draw reservoirs. From Cox et al. (1998). 122

Fig. 3.41. West-east structural cross section though Upper Pennsylvanian strata in Dagger Draw South reservoir showing morphology of dolomitized shelf-margin reservoir facies, impermeable backreef limestones, and basinal shales. From Speer (1993). 122

Fig. 3.42. Locations of wells analyzed for this project in the shelf-margin project area and locations of type logs for the shelf area, basin area and Dagger Draw North reservoir (Figs. 3.43–3.45). 124

Fig. 3.43. Type log through Upper Pennsylvanian shelf facies. See Fig. 3.42 for location. 125

Fig. 3.44. Type log through Upper Pennsylvanian basinal facies. See Fig. 3.42 for location. 126

Fig. 3.45. Type log through Upper Pennsylvanian shelf-margin facies, Dagger Draw North reservoir, indicating productive zone. See Fig. 3.42 for location.....127

Fig. 3.46. Structure contours on top of Mississippian strata, shelf-margin project area. Contours in feet. Datum = sea level.....128

Fig. 3.47. Structure contours on top of upper Wolfcamp shale, shelf-margin project area. Contours in feet. Datum = sea level.....129

Fig. 3.48. Isopach map of stratigraphic interval between the top of the upper Wolfcamp shale marker and the top of the Mississippian System, shelf-margin project area. 130

Fig. 3.49. Upper Pennsylvanian (Cisco and Canyon) reservoirs and Mississippian structure, shelf-margin project area. 131

Fig. 3.50. Upper Pennsylvanian (Cisco and Canyon) reservoirs and upper Wolfcamp shale structure, shelf-margin project area. 132

Fig. 3.51. Upper Pennsylvanian (Cisco and Canyon) reservoirs and isopach map of stratigraphic interval between the top of the upper Wolfcamp shale marker and the top of the Mississippian System, shelf-margin project area.....133

List of Tables

Table 2.1. Results of Distance-Based Outlier Algorithm 10

Table 2.2. Experimental Results for the Proposed Data Cleaning Algorithm 11

Table 2.3. Normal Data 16

Table 2.4. Not-Normal Data 16

Table 2.5. Frequency at Each Bin Having Two Non-Adjacent High Frequency Bins 22

Table 2.6. Frequency at Each Bin Having Two Adjacent High Frequency Bins 23

Table 2.7. Frequency at Each Bin Having One High Frequency Bin 23

Table 2.8. Frequency at Each Bin Having No High Frequency Bins 24

Table 2.9. Frequency of Each Bin When Size Is 64 26

Table 1.10. Frequency of Each Bin When Size Is 75 (Case II) 27

Table 2.11. Frequency of Each Bin When Size Is 75 (Case III) 28

Table 2.12. Data Management Subsystem, Binary File Format 43

Table 2.13. Data Management Subsystem, Binary File Format after Add 49

Table 2.14. Data Management Subsystem, Binary File Format after Delete 52

Table 3.1. Permo-Pennsylvanian Carbonate Reservoirs Productive within the Bough Intrashelf Project Area (See Fig. 3.5 for Locations of Reservoirs) 74

Table 3.2. Cumulative Oil, Gas and Water Production Data for the 241 Wells Analyzed in the Detailed Study Area 86

Table 3.3. Permo-Pennsylvanian Carbonate Reservoirs Productive within the Shelf Margin Project Area (See Fig. 3.37 for Locations of Reservoirs) 118

EXECUTIVE SUMMARY

The objective of this project is to create a user definable and customizable fuzzy expert system tool to dramatically speed local and regional play analysis and to reduce subsequent drilling risk. Two particular challenge for this general tool is that it must not require computer programming skill, and that it will instead use software wizards and highly adaptable software subsystems to guide users through the process of building a successful expert system to evaluate plays from field to basin scale using public and/or private data and their own or public expertise. To demonstrate the effectiveness of the tool, an analysis of the Pennsylvanian play of SE New Mexico, a play with significant potential for bypassed pay, is being used as a test case during the design and testing phases of the system.

In the first project year a large set of rules that form default settings based on play type and dimensions was gathered using the Delphi method. These rules were used to populate databases with rules, variables, and fuzzy sets that can be defined or redefined on demand. Preliminary software that allowed customization using a graphical interface to add, subtract or modify rules based on the user's were also developed. A prototype of the scalable and adaptable inference engine was completed, and the initial data for an outstanding, bypassed-pay play, the Pennsylvanian play in New Mexico, were collected.

In the second project year Work continued on making fuzzy sets and rules scalable and easily configured, as well as self-checking for conflicts and consistency. Software development resulted in a prototype of the CFS system which is currently in in house testing (alpha testing) of software components, evaluation of the efficiency of the integration of software components, and in production of preliminary expert systems. The software is ready to be tested and modified by a small group of future users (beta testing) in the final project year. Completion of data collection for the Pennsylvanian play, including organization of maps and data for testing of the CFS software, at the end of year 2 will will allow generation of an example of the usage of the system while simultaneously providing a significant opportunity for identifying new reserves.

This second report contains a summary of progress to date, problems encountered, and plans for the next year and future progress. The emphasis of this report is on work completed during the second project year, January 2006 through March 2007.

REPORT DETAILS

1. Introduction

Four major tasks must be accomplished to complete the project goals: Progress on these four tasks will be discussed in the remainder of this report. Task 1.0 is the development of a customizable fuzzy expert system (CFS) with the ability to self-generate software and fully customize integral components of the expert system for non-computer programmers. Task 2.0 is the development of graphical user interfaces and wizards to simplify the goals of Task 1.0 by using wizards that allow users to quickly and easily build and customize the system. Task 3.0 represents data and rule collection from the Pennsylvanian Carbonate play of southeastern New Mexico and subsequent validation of the utility of the system. A continuous effort in tech transfer comprises Task 4.0, as the ultimate test of the successful program will require that it is widely available and utilized.

The Gantt chart shown in Fig. 1.1 gives the approximate time distribution of the tasks and some of the major subtasks discussed in the remainder of the report. Tasks 1.3 (Generalized Fuzzy Engine), 2.1 (User Interfaces and Wizards), 2.2 (Expert System Design Wizards), 3.1 (Collect Data), 3.2 (Application of CFS) and 4 (Web Pages) were active during the second project year. Tasks 1.3, 2.1 and 3.1 were completed during the second project year.

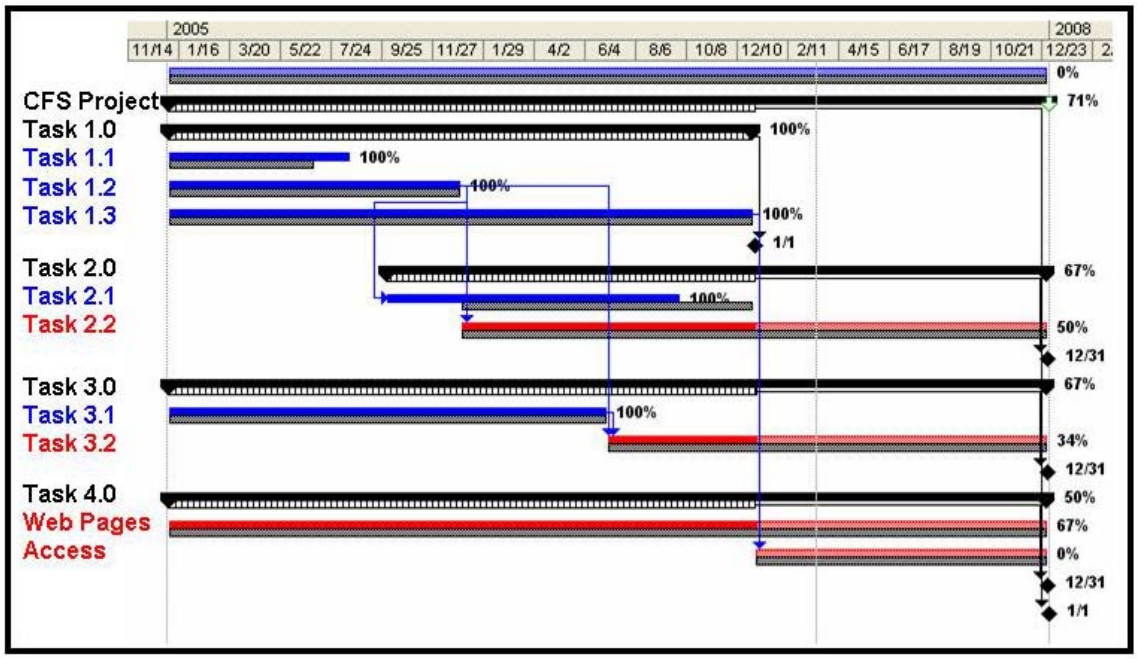


Fig. 1.1. Gantt chart of project tasks.

2. Expert System Development

The primary goal of Task 1 is to create a CFS that can quickly and easily adapt to exploration problems in a wide range of settings, with little–to-significant user input and modifications as required. Task 1 is divided into three main subtasks: Task 1.1 creating a generalized knowledgebase, Task 1.2 creating databases and self-generating software, and Task 1.3 building the generalized fuzzy inference engine. Tasks 1.1 and 1.2 were completed during the first project year and reported on in the first annual report. Task 1.3 saw initial development during the first project year and was completed during the second year.

Task 2.0 was the focus of the majority of software development work during the second project year. Individual software components were implemented using wizard-based design. The entire workflow for designing and implementing a customized fuzzy expert system requires implementation in an easy-to-use fashion. Design specifications necessary to accomplish this were determined to include:

- One overall user interface. Multiple interfaces and starting and stopping programs lead to unneeded complexity.
- Task tracking. The software needs to keep track of where in the workflow the user is working. If the software is stopped, then restarted, the workflow begins at the point.
- Multiple layer undo. Users need to be able to revert to older variable definitions seamlessly. This adds database complexity.
- The workflow needs to allow users to easily move between modules. Including:
    - Skipping modules (with automated information adjustment).
    - Returning to previously finished tasks to review and/or change items.
- Variables need to be able to change type, and keep track of several generations of changes.
- The entire process needs to be as short as an hour for defaults systems (using a pre-set pay type). But powerful enough to allow complete customization at field to regional scales.

Essentially the user interface allows all of the previously built components to work together in a seamless, wizard driven environment with guided workflows. A screen shot of the interface with all modules active is shown in Fig.2.1.

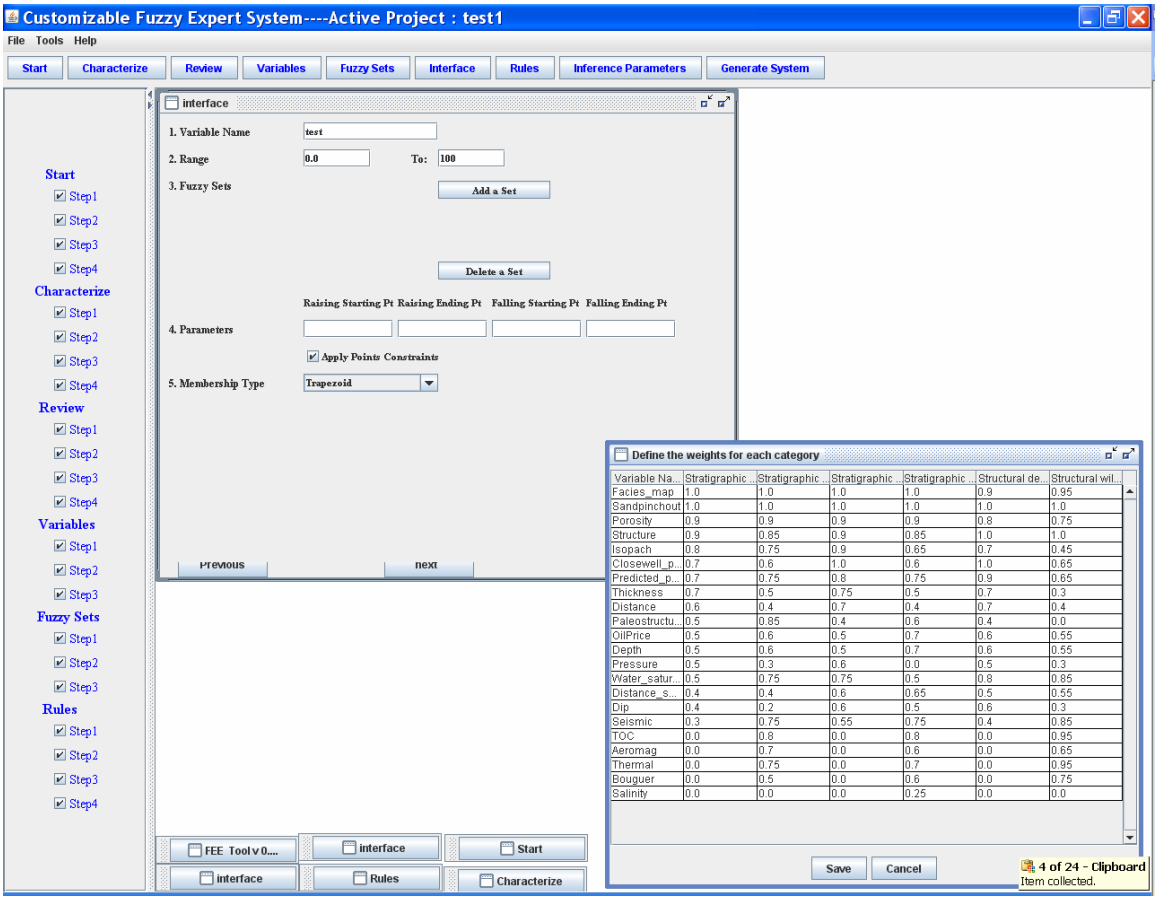


Fig. 2.1. CFS overall interface with all modules active.

2.1 Generalized Fuzzy Inference Engine—Task 1.3

The generalized inference engine required the ability to access the rule and fuzzy set data stored in databases. At the core of the system, the inference engine organizes the rules, checks for conflicts and outputs a resulting interpretation, in this case, an estimate of how successful a prospect will be. A fuzzy inference engine is not dependent on large amounts of specific data and, with properly formed questions, can perform inference with scarce data. The challenge for developing this inference engine was allowing a large number of variables and ranges to be defined by the user. The software needed to adapt to each study based on an analysis of rules, variables, and fuzzy ranges described and stored in external database files that can and will be changed on demand by users. This subtask required the first two project years and its development during the second year was linked

with the database classes developed for subtask 1.2 in order to allow seamless integration of data and rules.

Building on the success of the first year prototype tests, a functional version of the software has been completed and has been integrated with user interfaces. It is anticipated that only minor changes to inference software will be required, as the Pennsylvanian data is processed using the CFS software. The prototype has been used to replicate the results of the Delaware FEE Tool by giving it access to the data already generated for that play. Specifics on design and testing of the inference engine were reported in the first annual report.

2.2 User Interfaces and Expert System Design Wizards

Interfaces to Populate and Manage Knowledgebase Information

A knowledgebase is defined as a "machine-readable resource for the dissemination of information, generally online or with the capacity to be put online," (May, 2001) and as "a collection of facts, relations, procedures etc., which constitute the knowledge about a particular domain." (Hart, 1986). The pre-set knowledge base of the CFS comprises expert knowledge from interviews, to which is added both user-defined and data-derived rules. The knowledgebase of an expert system consists of the rule sets whose input is numerical, or will be converted to a numerical format for computer analysis. This numerical data can also be used to generate fuzzy membership functions (or curves) that can be used by an inference engine to evaluate the rules using fuzzy logic rather than discrete or "crisp" inference techniques.

A key for successful implementation of the CFS is the ability to introduce and assimilate user data in a variety of formats and convert them for use by the inference engine. Therefore, a necessary component is user-friendly wizard-driven software that reviews the quantitative data provided by the user and develops fuzzy curves for that data and thus aids in development of data derived rules from the user provided raw data.

Data-derived rules have to be developed when the user provides only the raw data as input. Therefore, a tool that reviews the quantitative data provided by the user and develops initial fuzzy curves for that data was designed. The architecture consists of three key modules – Data Cleaning, Normality Testing and Fuzzy Set Definitions. The fuzzy curves can be viewed using a separately developed module where they can be viewed and modified by the user before being used in their customized expert system. The architecture flowchart is shown in Fig. 2.2.

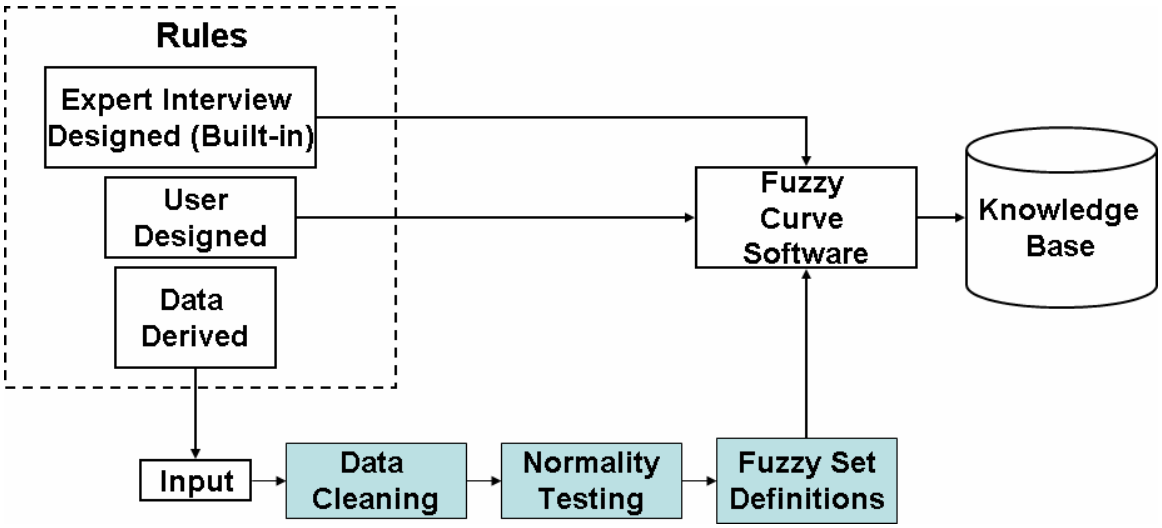


Fig. 2.2 . Stages in the development of a CFS knowledgebase.

Data Cleaning Module

The user's data is first examined in the Data Cleaning module where the data is cleaned in two stages—Data Collection and Data Analysis (Xiong, 2006). In the data analysis stage, the data cleaning algorithms remove irrelevant or weakly irrelevant objects for the purpose of improving the results of data analysis. Detection and removal of errors is not the key focus in this stage of data cleaning. The data analysis can be done with different methods: distance-based, density-based and clustering-based. Xiong *et al.* (Xiong, 2006)

described the data cleaning algorithms in the form of outlier detection algorithms. Distance-based algorithms identify the outliers based on the distance measure. Density-based algorithms identify the outliers in data sets with varying densities. Clustering-based algorithms identify the small clusters that are far away from other major clusters as sets of outliers. Using these definitions, the selected data-cleaning algorithm falls under the category of distance-based data cleaning algorithms. The cleaned data is then passed to the Normality Testing module to test the normality of the data.

Data Collection Stage: The data-cleaning algorithm takes user data and calculates two ranges, the error range and the warning range. A value outside the error range is typically a data entry error, and a value outside the warning range may be a possible data entry error or an outlier. In order to determine these ranges, the user defines a variable type. The types are proportional (decimal or %), nonnegative real, real, integer, whole, and natural.

The warning range is created by using distance-based outlier detection. The mean of the data ($\bar{x}$) and the standard deviation (s) are calculated. The interval $[\bar{x} - 2s, \bar{x} + 2s]$ is then computed, which implies that even if the data is not normally distributed, it will contain at least 75% of all values regardless of the distribution of the data set. The intersection of the interval above and the error range based on the variable type gives the warning range.

Example: Suppose the user wants to enter data for connate water saturation. The data is usually represented as a percentage, so the variable type selected is % proportion. The error range is thus [0,100].

To compute the warning range, suppose the mean of the user's data is 11% with a standard deviation of 6%. The interval built in the second step would be [11-2(6), 11+2(6)] or [-1, 23]. The intersection of the set [-1, 23] and [0,100] is [0, 23], so this is the warning range.

A data value of -2% or 105% would fall outside the error range; a pop-up tells the user that the value is outside the domain of the data. A data value of 50% would fall outside the warning range, so a pop-up will recommend that the user verify the value. A data value of 16% would be inside both ranges, and would not raise any flags.

Data Analysis Stage: The selected algorithm uses a distance-based outlier detection algorithm based on Xiong *et al.* (Xiong, 2006). They define a distance-based method as follows:

"An object in a data set D is a distance-based outlier if at least a fraction α of the objects in D are at a distance greater than r."

Figure 2.3 describes the pseudocode of the proposed distance-based data cleaning algorithm. r and α are accepted from the user as input. N is the total number of data points. x is an one-dimensional array that contains all the N data points. N_r is the number of data points that lie within a distance r from the mean. N_α is the number of data points that represent a fraction α of the N. The **For** loop traverses the whole array and determines the number of data points that lie within a distance r from its mean. The data points are sorted with respect to the distance from the mean. The points that are farthest away from the mean are considered as noise and hence removed. To do this, the algorithm considers the greater of the two, N_r or N_α. The first N_r or N_α (whichever is greater) data items are considered as clean and the rest are considered as noise.

```
Nr := 0;
For i:= 1 to N do
  Distance[i] :=Mod(x[i] – mean)/sd;
  If (Distance[i]< r) then
     Nr := Nr + 1;
  End If
End For
Sort (X,ascending,distance from mean);
Nα := N * α;
If (Nr < Nα)then
  Clean Data Set := X[1.. Nα];
Else
  Clean Data Set := X[1..Nr];
End If
```

Fig. 2.3: Algorithm pseudocode: distance-based data cleaning.

Summary and Justification: The selected algorithm has a complexity of O(n) while the algorithm proposed by Xiong *et al.* has a complexity of $O(n^2)$. This is because the algorithm by Xiong *et al.* calculates data items with the least number of neighbors within a specified radius and requires two **For** loops to achieve the same. The selected algorithm in this research is faster than the algorithm by Xiong *et al*. The drawback of the proposed algorithm is that it does not clean data efficiently when the data set is irregular or skewed.

Table 2.1 illustrates the selected algorithm with an example consisting of 20 data items. The first column represents the index of the array x. The second column represents all the unsorted data elements and their corresponding distance from the mean. The third column represents the data elements that are sorted in ascending order of distance from the mean. Let the values of the global parameters be as follows: $r = 2$, $\alpha = 0.95$. After executing the algorithm, it was found that $N_r = 18$, and $N_\alpha = 19$. Since $N_\alpha > N_r$, a clean set consists of 19 items from the sorted data set.

Table 2.1. Results of Distance-Based Outlier Algorithm

	Unsorted		Sorted according to distance	
Index (i)	X[i]	Distance from mean	X[i]	Distance from mean
1	15.75	0.26	15.36	0.03
2	16.16	0.50	15.26	0.03
3	14.97	0.21	15.15	0.10
4	19.25	2.34	15.51	0.12
5	13.35	1.17	15.58	0.16
6	15.51	0.12	15.61	0.18
7	14.15	0.70	14.97	0.21
8	15.26	0.03	15.75	0.26
9	15.36	0.03	16.15	0.49
10	13.41	1.13	16.16	0.50
11	15.15	0.10	16.45	0.68
12	14.12	0.71	14.15	0.70
13	16.63	0.78	14.12	0.71
14	19.47	2.47	16.63	0.78
15	15.58	0.16	13.86	0.87
16	16.15	0.49	13.41	1.13
17	12.66	1.58	13.35	1.17
18	13.86	0.87	12.66	1.58
19	15.61	0.18	19.25	2.34
20	16.45	0.68	19.47	2.47

Nr

Nalpha

Mean=15.32, SD=1.68, r=2, N_r=18, α =0.95, N_α =19

The drawback of distance-based algorithms is that they cannot consider points in different densities.

The experiments are performed using two data sets for *normal* and *not normal* data with the parameters of r and α changed arbitrarily. The data sets are generated using excel having size 100. The *normal* data set has a mean of 4.9 and standard deviation of 0.99.

Table 2.2 shows the percentage of data used for normality testing for *normal* and *not-normal* distributions after filtering the data based on changing the values of r and α. For *normal* data, 68% of the data falls between $\bar{x} - s$ and $\bar{x} + s$ and 96% of the data falls between $\bar{x} - 2s$ and $\bar{x} + 2s$.

The *normal* data with r = 1.00 and α = 0.50 filters 32% of the data which makes the distribution of the data as *not-normal*. Thus, parameter selection plays a significant role in determining the output of the program.

Table 2.2. Experimental Results for the Proposed Data-Cleaning Algorithm

r	α	Normal Data	Not Normal Data
1.00	0.50	0.68	0.64
1.50	0.50	0.85	0.89
2.00	0.50	0.96	0.99
1.00	0.75	0.75	0.75
1.50	0.75	0.85	0.89
2.00	0.75	0.96	0.99
1.00	0.90	0.90	0.90
1.50	0.90	0.90	0.90
2.00	0.90	0.96	0.99

Normality Testing Module

Kolmogorov-Smirnov (K-S), Lilliefors, and Shapiro-Wilks are commonly described normality tests available in the literature. Kolmogorov-Smirnov is based on maximum difference between the sample cumulative distribution and the hypothesized cumulative distribution. Lilliefors is an extension of K-S but the mean and SD of the hypothesized distribution are estimated from sample data. Shapiro-Wilks has good power properties but is difficult to implement. One of the most common used goodness-of-fit tests is the chi-square test. The advantage of this test is that it can be used with any type of input data (sample, density or cumulative) and any type of distribution function (discrete or continuous). A weakness of the chi-square test is that there are no clear guidelines for selecting intervals, which increases difficulty of implementation for lay users. The test is considered subjective because it depends on the user-supplied intervals specifications (Jankauskas, 1995). The K-S test does not depend on the number of intervals, which makes it more powerful than the chi-square test, though it has weaknesses in detecting tail discrepancies.

The Normality Test module uses the K-S algorithm to test the normality of the data. It is based on the maximum difference between sample cumulative distribution and the hypothesized distribution. The fuzzy set definitions are then derived based on the fuzzy set definition algorithms for both normal and not-normal data. The Fuzzy Set Definitions module provides the starting point with a number of basic rule sets that apply to a variety of formation types. Users can then add to, delete or modify these rule sets. The fuzzy curve software allows users to view and modify the fuzzy curves before being used by their customized expert system.

Normality Testing: The cleaned data that resulted from the Data Cleaning module is given to the Normality Testing module. This is used to test if the data provided by the user is normally distributed or not. This section explains the K-S algorithm which is used

for normality testing and the illustrations and histograms for the *normal* and *not normal* data.

Statistical Algorithm—Normal and Not-Normal Data: Figure 2.4 illustrates the Statistics module used in this project. There are different subtasks in the statistics module. The user data is read and then analyzed for errors. Then, the data is passed on to the Normality Test submodule. Here, the data is tested for normality. After this test, the data is labeled either as *normal* or *not-normal*. The histograms/box plots are drawn as graphical output. The labeled data is then passed on to the Fuzzy Set Definition module, which generates the definitions for the fuzzy sets. The definitions can be viewed using fuzzy curve software.

The Normality Test module uses the K-S algorithm to test the normality of the data. It is based on the maximum difference between the sample cumulative distribution and the hypothesized cumulative distribution. The algorithm is listed below:

1. The step cumulative distribution function for the data: $S_N(x)$ gives the fraction of the data points to the left of a given value of x. It is constant between consecutive values of x_i and jumps by 1/n at each x_i.

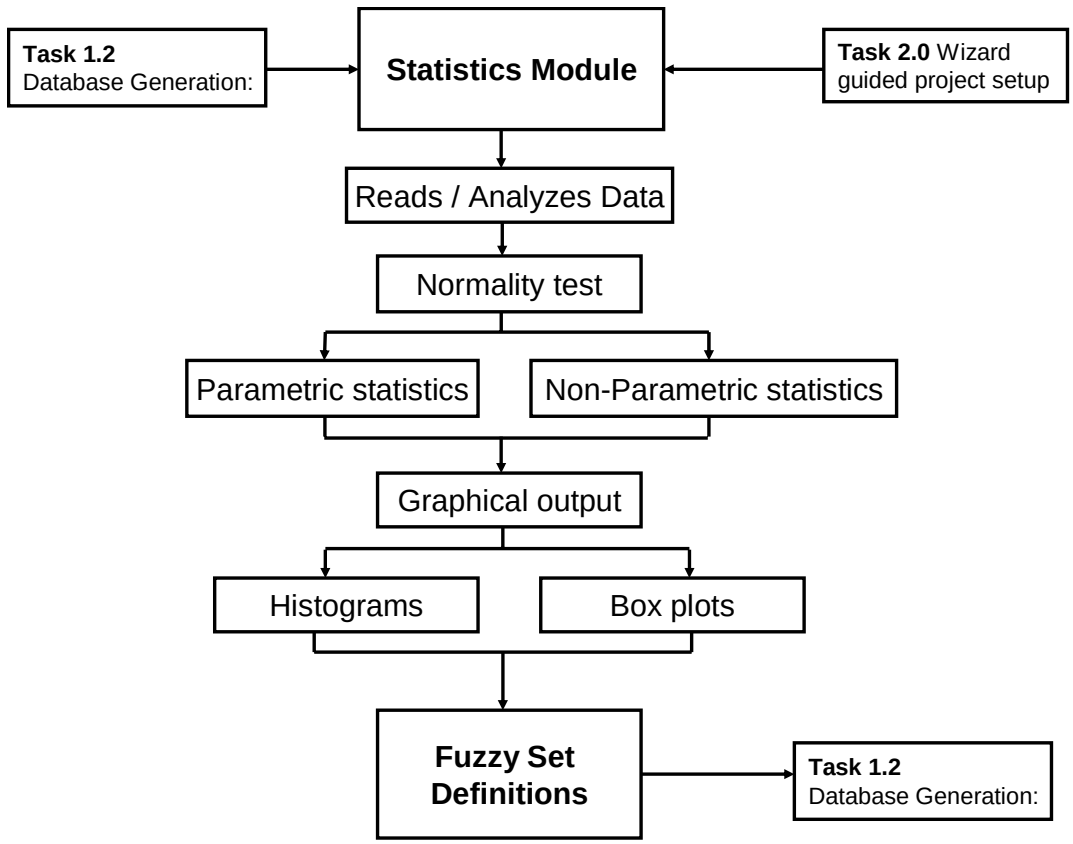


Fig. 2.4. Overview of the Statistics Module.

2. Cumulative normal distribution function is calculated using the following equation:

$$P(x) = \frac{1}{2}\left(1 + erf\left(\frac{x - \bar{x}}{s\sqrt{2}}\right)\right)$$

3. With the function P(x) and $S_N(x)$, K-S statistic, D is computed as below:

$$D = \max_{-\infty < x < \infty} \left| S_N(x) - P(x) \right|$$

4. Then the significance of D is calculated by computing Q_{KS}:

$$Q_{KS}(\sqrt{n}D) = 2\sum_{j=1}^{\infty} (-1)^{j-1} e^{-2j^2(\sqrt{n}D)^2}$$

This is an infinite series that should converge to a value between 0 and 1. If this value of Q is small, the data is not normally distributed (Q < 0.10) is good for these purposes.

In step 2 of the algorithm, the formula given does not contain an explicit closed-form solution for calculation of the integral. Abramowiz and Stegun (Abramowicz, 1964) contains an approximation of P(x). Figure 2.5 lists that part of the java code below from 'NormalityTest.java':

```
public double CNDF(double z)
{
        if (z> 6.0) return 1.0;
        if (z<-6.0) return 0;
        double b1 = 0.31938153;
        double b2 = -0.356563782;
        double b3 = 1.781477937;
        double b4 = -1.821255978;
        double b5 = 1.330274429;
        double p = 0.2316419;
        double c2 = 0.3989423;
        double a = Math.abs(z);
        double t = 1.0/(1.0 + a*p);
        double b = c2 * Math.exp((-z)*(z/2.0));
        double n = ((((b5 * t + b4)*t+b3)*t+b2)*t+b1)*t;
        n = 1.0 - b*n;
        if (z<0.0) n = 1.0 - n;
        return n;
}
```

Fig. 2.5. Calculation of P(x).

Implementation of the code for normality testing requires validation. Four datasets (two *normal* distributions and two *not-normal* distributions) were used for the validation process. Tables 2.3 and 2.4 show the results of applying K-S test to *normal* data and *not-normal* data, respectively.

Table 2.3. Normal Data

S(x)	P(x)
0.05	0.012890285
0.10	0.051603599
0.15	0.083870383
0.20	0.173737036
0.25	0.239007063
0.30	0.281125892
0.35	0.322937607
0.40	0.367116629
0.45	0.440579556
0.50	0.496210107
0.55	0.528102197
0.60	0.621909537
0.65	0.637018176
0.70	0.648212782
0.75	0.677421284
0.80	0.702117353
0.85	0.705573883
0.90	0.925136958
0.95	0.969980081
1.00	9.93E-01

Table 2.4. Not-Normal Data

S(x)	P(x)
0.05	0.063380528
0.10	0.065905357
0.15	0.072558233
0.20	0.075359044
0.25	0.085812715
0.30	0.102577716
0.35	0.109911365
0.40	0.117613876
0.45	0.131997214
0.50	0.134147013
0.55	0.182203376
0.60	0.252386009
0.65	0.255590402
0.70	0.364295739
0.75	0.40632846
0.80	0.477272551
0.85	0.556854913
0.90	0.944419902
0.95	0.999999961
1.00	1

Figures 2.6 and 2.7 show the K-S statistic for the *normal* and *not-normal* data, respectively. The y-axis in these graphs follows a log-normal scale. In the case of the *normal* data (Fig. 2.6), the maximum distance between P(x) and S(x) is very less. In the case of the *not-normal* data (Fig. 2.7), there is significant difference between the cumulative distributions.

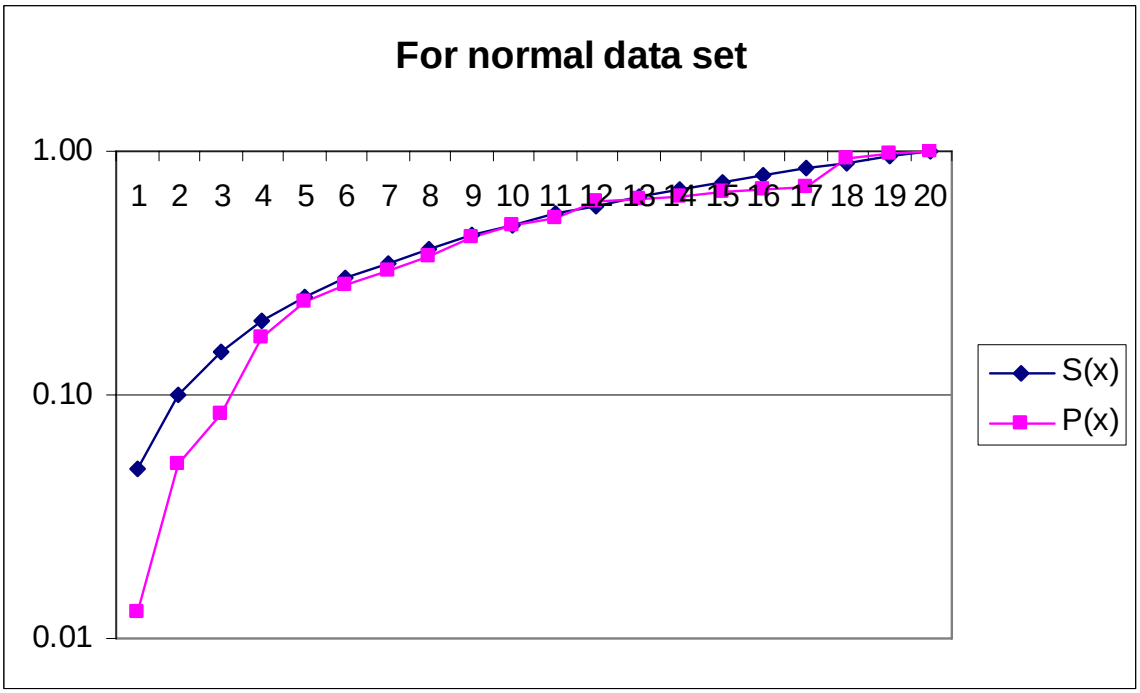


Fig. 2.6. K-S statistic for *normal* data.

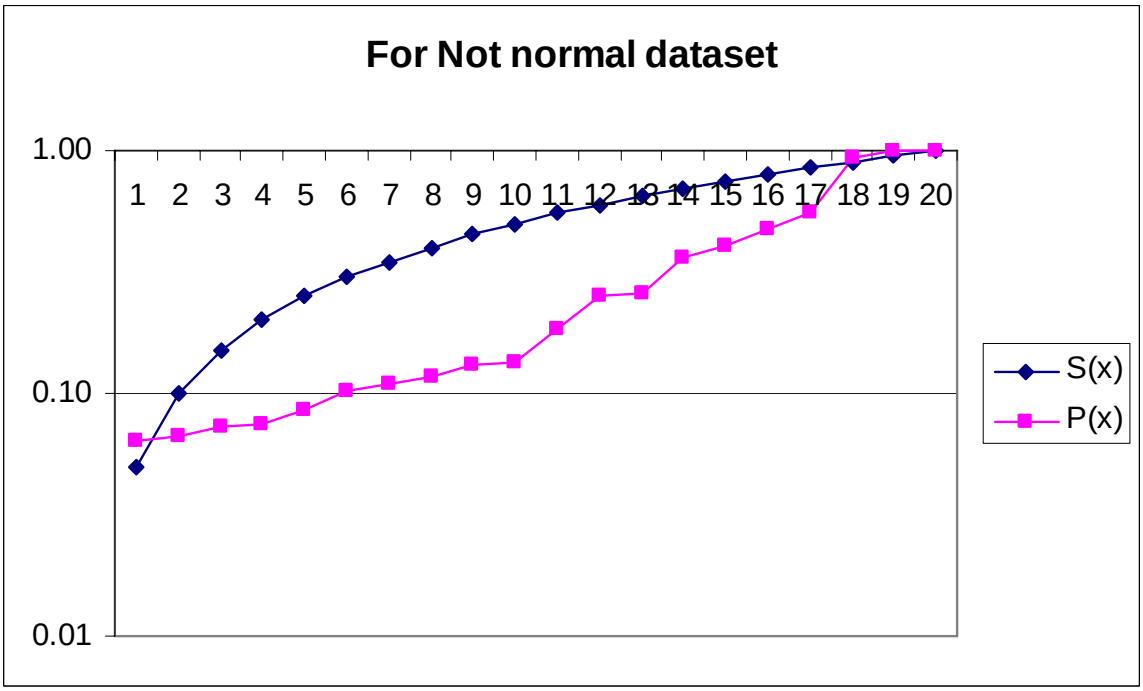


Fig. 2.7. K-S statistic for *not-normal* data.

Figure 2.8 shows the histogram for a larger test set of *normal* data having mean 4.9 and standard deviation of 0.99. It has the shape of a bell curve.

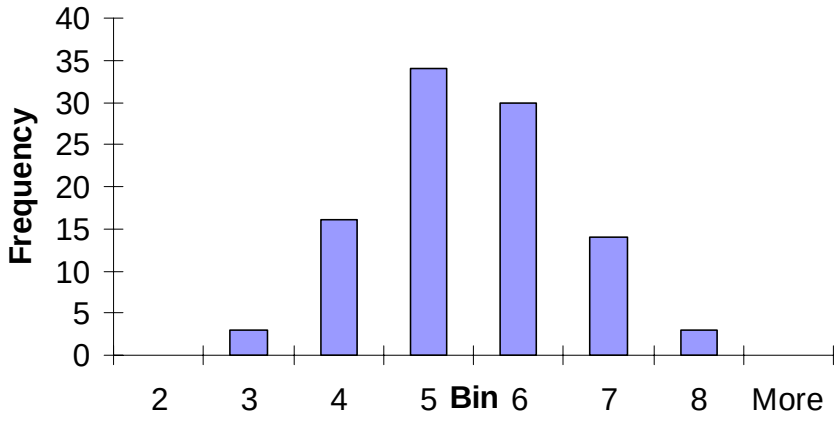


Fig. 2.8. Histogram for *normal* data.

Figure 2.9 shows the k-s statistic for *normal* data. It is a lognormal scale graph having the values of S(x) and P(x). The x-axis has the number of data points, which is 100. It is observed that the difference between S(x) and P(x) is the minimum for the *normal* data.

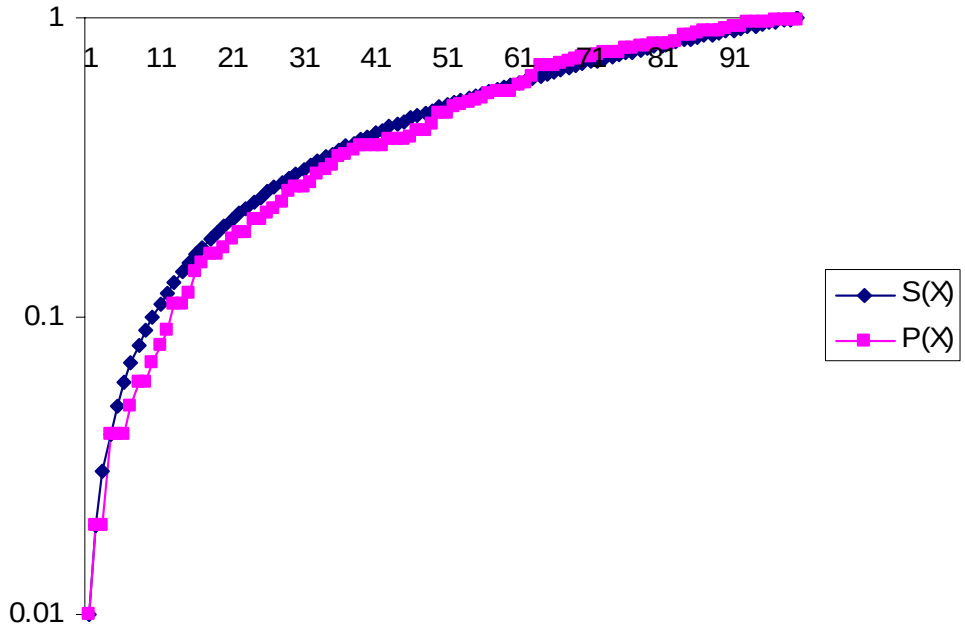


Fig. 2.9. Illustration of k-s statistic for *normal* data.

Figure 2.10 shows the histogram for the *not- normal* data having mean 36.46 and standard deviation of 30.65. Figure 2.11 shows the k-s statistic for *not-normal* data. It is a lognormal scale graph having the values of S(x) and P(x). The maximum difference between P(x) and S(x) is large in case of *not-normal* data.

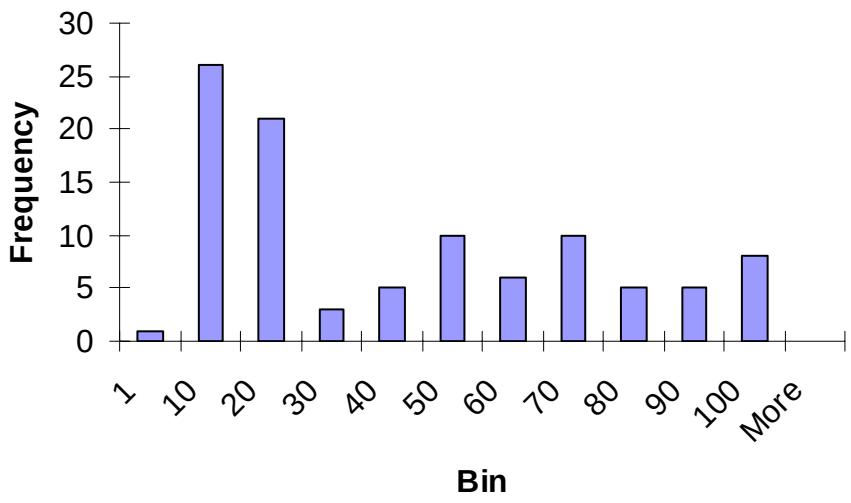


Fig. 2.10. Histogram for *not-normal* data.

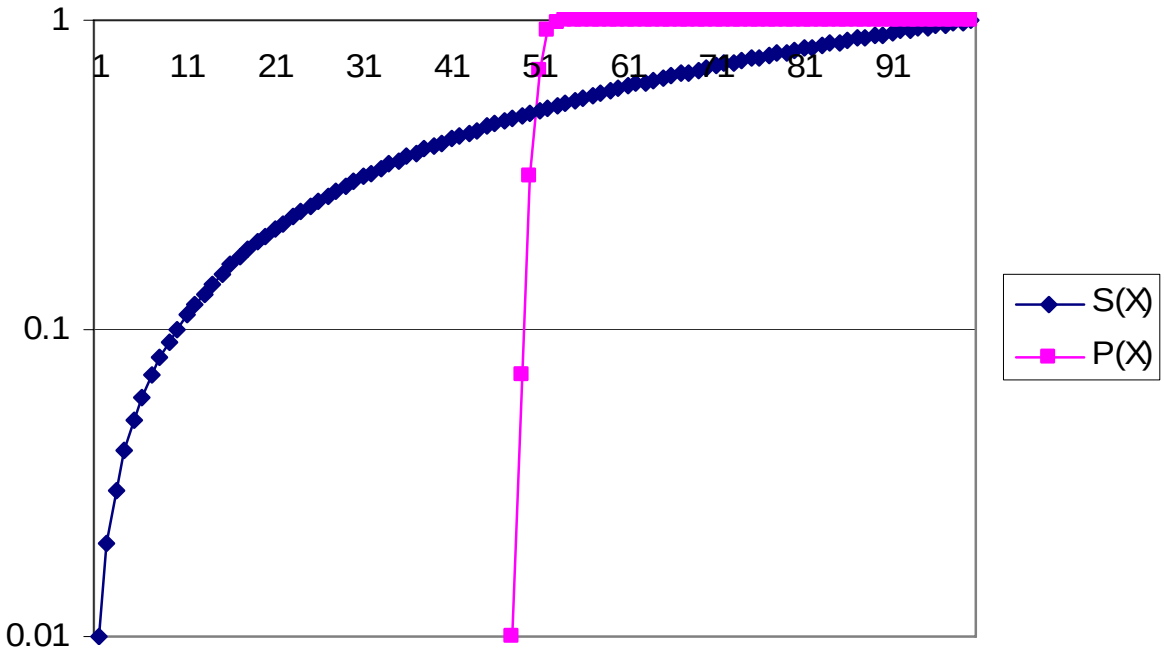


Fig. 1.11. Illustration of k-s statistic for *not-normal* data.

These examples illustrate the need for the fuzzy curve generation software to understand the data distributions, and then address *normal* and *not- normal* data in the curve selection process. Values of the K-S statistic provide a ready measurement of which approach to take when generating fuzzy sets for user-provided data streams

Fuzzy Curve Generation Module:

Fuzzy Set Definition Algorithm—*Normal* Data: The empirical rule states that normally distributed data has 68% of its data points between $\bar{x} \pm s$, 95 % between $\bar{x} \pm 2s$ and 99% of the data between $\bar{x} \pm 3s$. If data is from a population that is normally distributed, fuzzy curves are drawn as in Fig. 2.12 with end point values from b0 to b9.

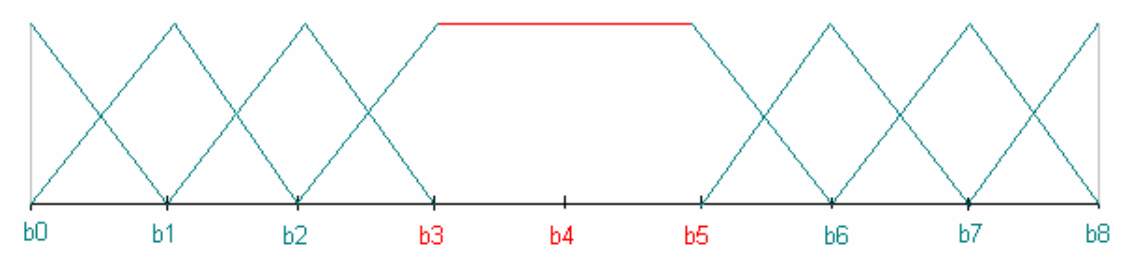


Fig. 2.12. Fuzzy curves for normal data.

Note:

$$b0 = \bar{x} - 3s, b1 = \bar{x} - 2s, b2 = \bar{x} - s, b3 = \bar{x} - s/2, b4 = \bar{x}, b5 = \bar{x} + s/2, b6 = \bar{x} + s, b7 = \bar{x} + 2s, b8 = \bar{x} + 3s$$

It is observed that 68% of all data points will fall between b2 and b6, 95% of all data points will fall between b1 and b7 and 99% of all data points will fall between b0 and b8.

Fuzzy Set Definition Algorithm—*Not-Normal* Data: In this case, a cumulative frequency distribution chart is developed and the frequencies determine the shape of the distribution. The variable n is the sample size, m is the number of bins or classes, b_j is the boundary of the bin, f_j is the frequency for each bin, and r is the range x_n- x_1. The following steps illustrate the algorithm:

1. First the boundaries of the bins $\{b_j\}$ are calculated as follows:

$$
\begin{aligned}
&m = 9 \\
&b_0 = x_1 \\
&r = x_n - x_1 \\
&w = r / m \\
&for\ \ j = 1\ to\ m \\
&\quad b_j = b_{j-1} + w \\
&next
\end{aligned}
$$

2. All the frequencies are initialized to 0 before calculating the frequency for each bin:

$$
\begin{aligned}
&For\ i = 1\ to\ n \\
&\quad For\ j = 1\ to\ m \\
&\qquad if\ \ b_{j-1} \le x_i < b_j \\
&\qquad then\ \ f_j = f_j + 1 \\
&\quad next
\end{aligned}
$$

3. Large frequencies are determined by comparing the frequency at each bin with the sum of mean frequency and standard deviation. The mean and standard deviation of the $\{f_i\}$ are calculated using the equation

$$\bar{f} = \sum_{i=1}^{m} f_i / m \qquad s_f = \sqrt{\frac{1}{m}\sum_{i=1}^{m}\left(f_i - \bar{f}\right)^2}$$

4. The large frequencies obtained from the inequality $f_i \ge \bar{f} + s_f$ determine the shape of the fuzzy curve. In the cases where the above inequality is true, the fuzzy curve is drawn with a trapezoid. In the cases where it is false, the fuzzy curve is drawn as a triangle.

The different cases of n*ot- normal* data based on the high frequency bin are explained with the help of an example. The highlighted frequencies are above the sum of the mean and the standard deviation.

Case 1: The data has two non-adjacent bins where the frequency is "high." Table 2.5 shows the frequency at each bin having lower and upper bound. There are two high frequency bins, 21 and 17, which are not adjacent.

Table 2.5. Frequency at Each Bin Having Two Non-Adjacent High Frequency Bins

lower	upper	f	
b0	b1	4	9.3
b1	b2	3	6.1
b2	b3	7	15.4
b3	b4	21	
b4	b5	9	
b5	b6	7	
b6	b7	17	
b7	b8	11	
b8	b9	5	

mean $\bar{f}$
standard dev. s_f
mean + s.d.

Figure 2.13 shows the fuzzy curves for the *not-normal* data where there are two high frequency bins that are not adjacent.

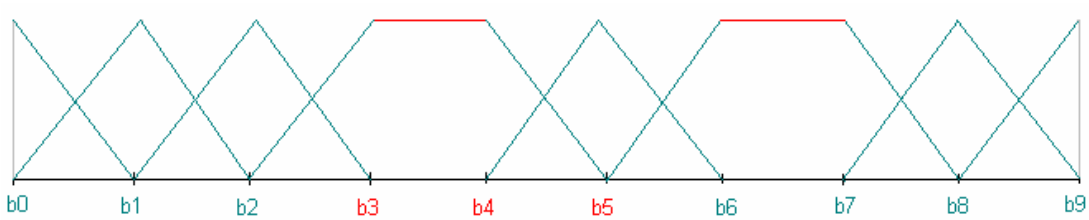


Fig. 2.13 Fuzzy curves for *not-normal* data having two non-adjacent high frequency bins.

Case 2: The data has two adjacent bins where the frequency is "high." Table 2.6 shows the frequency at each bin with its lower bound and upper bound. There are two high frequency bins, 21 and 19, which are adjacent to each other.

Table 2.6. Frequency at Each Bin Having Two Adjacent High Frequency Bins

lower	upper	f		
b0	b1	4	9	mean $\bar{f}$
b1	b2	3	6.5	standard dev. s_f
b2	b3	7	15.5	mean + s.d.
b3	b4	21		
b4	b5	19		
b5	b6	7		
b6	b7	9		
b7	b8	6		
b8	b9	5		

Figure 2.14 shows the fuzzy curves for the *not-normal* data having two adjacent high frequency bins. The adjacent high frequency bins merge into a single curve.

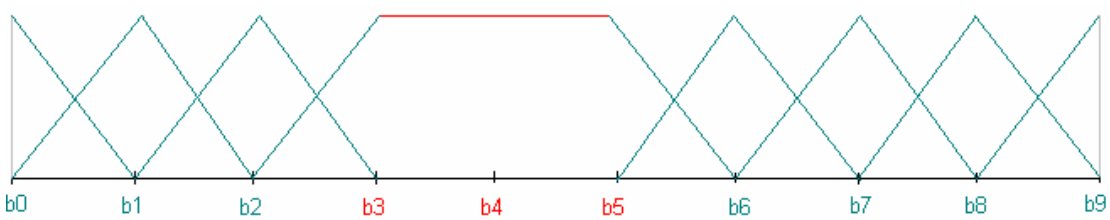


Fig. 2.14. Fuzzy curves for *not-normal* data having two adjacent high frequency bins.

Case 3: The data has one high frequency bin. Table 2.7 shows the lower bound, upper bound and frequency at each bin. There is one high frequency bin (Jankauskas, 1995). Figure 2.15 shows the fuzzy curves for the *not-normal* data having one high frequency bin.

Table 2.7. Frequency at Each Bin Having One High Frequency Bin

lower	upper	f		
b0	b1	4	5.7	mean $\bar{f}$
b1	b2	3	1.8	standard dev. s_f
b2	b3	7	7.5	mean + s.d.
b3	b4	6		
b4	b5	5		
b5	b6	7		
b6	b7	5		
b7	b8	9		
b8	b9	5		

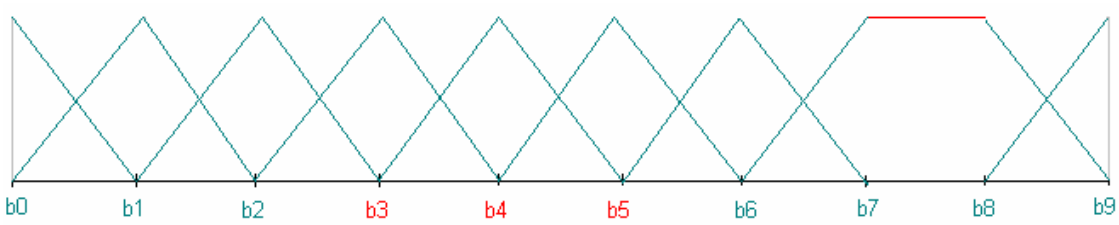


Figure 2.15 Fuzzy curves for *not-normal* data having one high frequency bin.

Case 4: The data has no high frequency bins and data is uniformly distributed. Table 2.8 shows the frequency at each bin with its lower and upper bound. There are no high frequency bins. Figure 2.16 shows the fuzzy curves of the *not-normal* data having no high frequency bins.

Table 2.8. Frequency at Each Bin Having No High Frequency Bins

lower	upper	f		
b0	b1	5	4.7	mean $\bar{f}$
b1	b2	4	0.5	standard dev. s_f
b2	b3	5	5.2	mean + s.d.
b3	b4	5		
b4	b5	4		
b5	b6	5		
b6	b7	5		
b7	b8	5		
b8	b9	4		

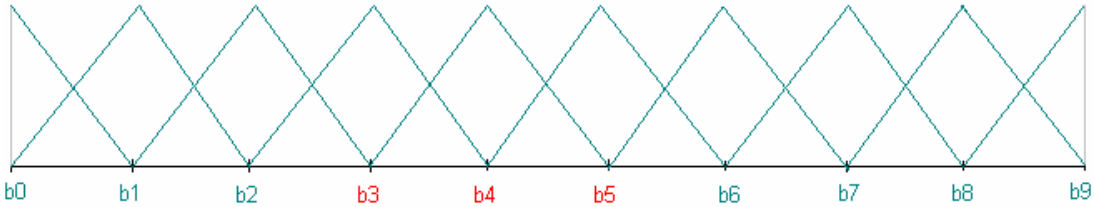


Fig. 2.16. Fuzzy curves for *not-normal* data with no high frequency bin.

It is observed that when the bins have high frequency, the curves are trapezoidal, with a horizontal top line connecting the lower and upper boundaries of the bins. If the bins do

not have a high frequency, the curves are triangular, with an apex at each bin endpoint. When there are two or more high frequency bins that are adjacent, the curves merge into one curve.

Fuzzy Curve Generation Software: Implementations of the fuzzy curve generation have been completed. The Fuzzy Curve Generation module generates fuzzy curves for *normal* data and *not-normal* data with minimal or intensive user interaction. Figure 2.17 shows a screen capture of software generated fuzzy curves for a *normal* data set.

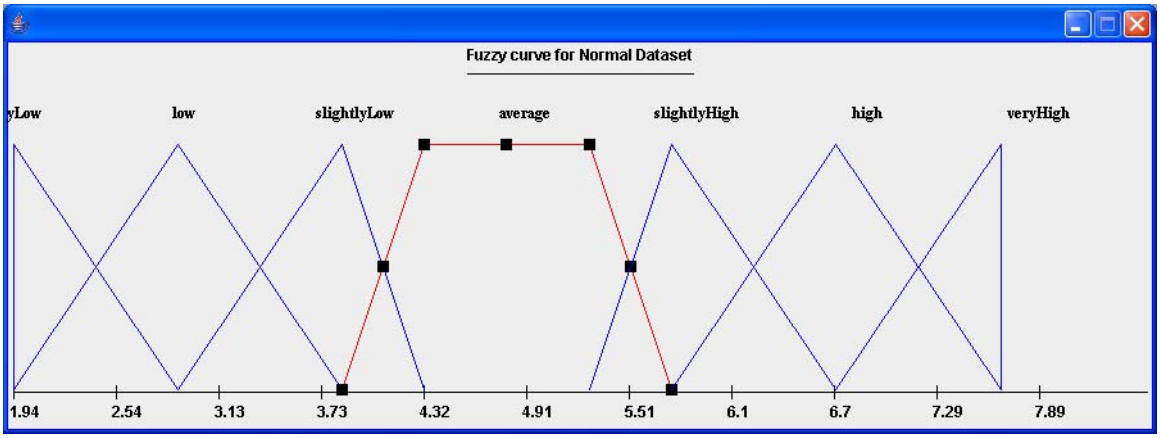


Fig. 2.17. Fuzzy curves for the *normal* data.

The *not-normal* data set has different cases based on the high frequency which are obtained after the filtering of data using data cleaning algorithm. Again, several cases or types of curve distributions are presented for the purpose of illustrating the concepts.

Case I: The size of the *not-normal* data after data analysis is 64, the sum of mean and standard deviation is 12.48. Table 2.9 has the lower bound and upper bound for each bin and its frequency. There are two high frequency bins whose value is greater than the sum of mean and standard deviation. In this case, two adjacent high frequency bins results in a single large curve covering the range of data for the two high frequency bins. Figure 2.18 shows the result of merging the two adjacent high frequency bins.

Table 2.9. Frequency of Each Bin When Size Is 64

Lower Bound	Upper Bound	Frequency
6.00	12.78	19
12.78	19.56	14
19.56	26.33	2
26.33	33.11	4
33.11	39.89	6
39.89	46.67	2
46.67	53.44	7
53.44	60.22	3
60.22	67.00	6

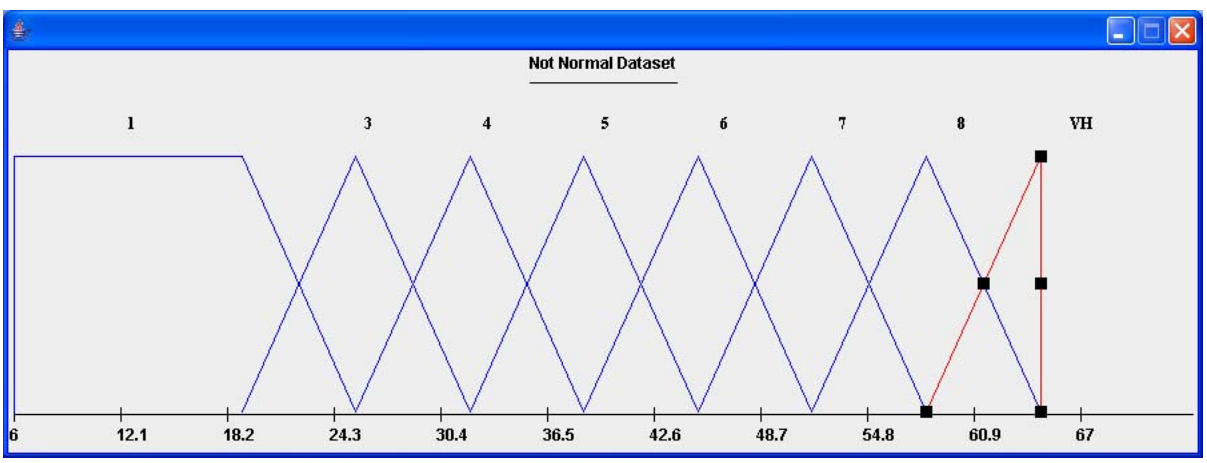


Fig. 2.18. Fuzzy curves for the *not-normal* data having adjacent high frequency bins.

Case II: Size of the *not-normal* data after data analysis is 75; the sum of the mean and standard deviation is 8.64. Table 2.10 has the lower bound and upper bound for each bin and its frequency. There are two nonadjacent high frequency bins whose value is greater than the sum of mean and standard deviation. Figure 2.19 shows these two high frequency bins that are not adjacent. Again, a bin without high frequency has curves which are triangular, with an apex at each bin endpoint.

Table 2.10. Frequency of Each Bin When Size Is 75 (Case II)

Lower Bound	Upper Bound	Frequency
13.00	21.22	9
21.22	29.44	8
29.44	37.67	8
37.67	45.89	8
45.89	54.11	9
54.11	62.33	8
62.33	70.56	8
70.56	78.78	8
78.78	87.00	8

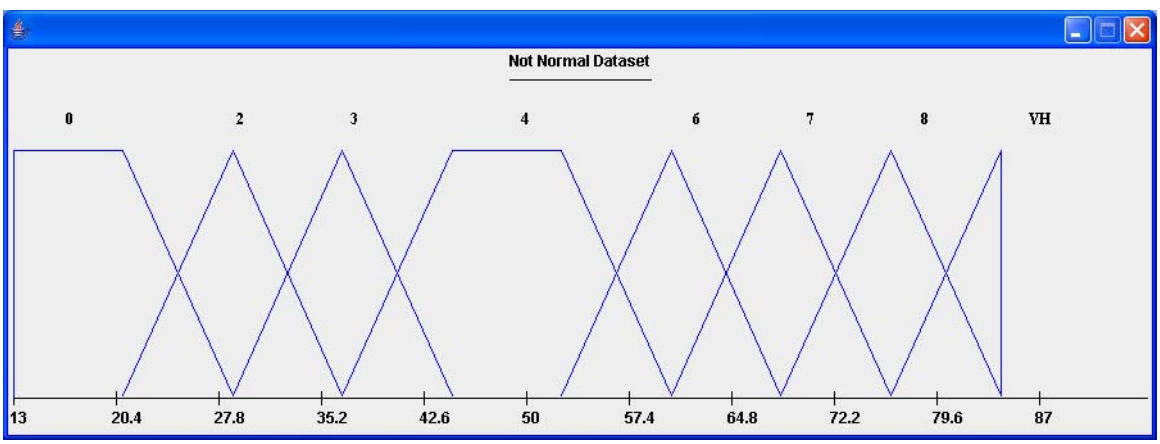


Fig. 2-19. Fuzzy curves for *not-normal* data having two nonadjacent high frequency bins.

Case III: The size of the *not-normal* data after data analysis is 100; the sum of the mean and standard deviation is 11. Table 2.11 has the lower bound and upper bound for each bin and its frequency. The frequency at each bin is equal to the sum of mean and standard deviation. Figure 2.20 shows the fuzzy curves for the *not-normal* data where there is a single frequency bin. It is observed that when there are two or more high frequency bins that are adjacent, the curves merge into one curve. Note this is a very unlikely scenario.

Table 2.11. Frequency of Each Bin When Size Is 75 (Case III)

Lower Bound	Upper Bound	Frequency
1.00	12.00	11
12.00	23.00	11
23.00	34.00	11
34.00	45.00	11
45.00	56.00	11
56.00	67.00	11
67.00	78.00	11
78.00	89.00	11
89.00	100.00	11

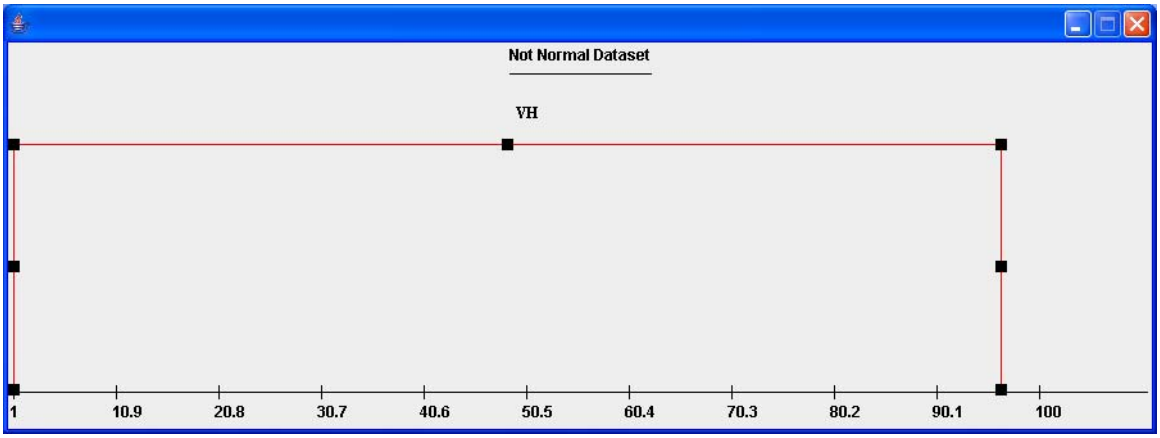


Fig. 2.20. Fuzzy curves for the *not-normal* data having one high frequency bin.

User Interface for Fuzzy Curve Software

The sequential flow of the Fuzzy Set Panel has four panels attached to it which allow the user to view curves and modify, add or delete fuzzy sets. The Fuzzy Set Panel is the main panel developed to allow the user to navigate between each of the panels. Figure 2.21 shows the *Variable Select Panel*, which allows the user to select a fuzzy variable to work with. The fuzzy curves can be viewed or modified using the panel.

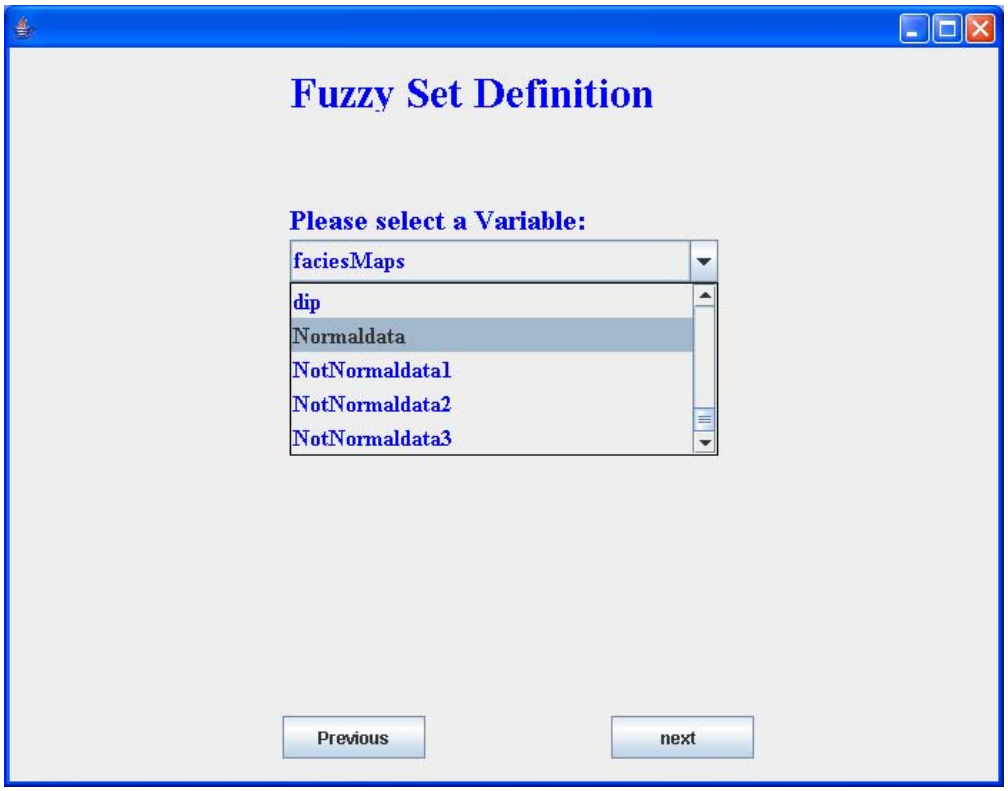


Fig. 2.21. Screenshot of the variable select panel.

Each step in the Fuzzy Set module was implemented as a sequential flow, which allows the user to select the options in the Fuzzy Set Operations Select Panel (Operations Panel). Figure 2.22 shows the operations panel in which the following actions take place when the “next” button is clicked:

a. When the View Fuzzy Curves option is selected in the Operations Panel, the user can view the fuzzy curves and the control goes back to the Operations Panel.
b. When the View Fuzzy Sets option is selected the user can view the Fuzzy Sets, view the Fuzzy Curves and then control goes back to the Operations Panel.

c. When the Add Fuzzy Sets option is selected the user can view the Add Fuzzy Sets panel, view the fuzzy curves and then control goes back to the Operations Panel.

d. When the Delete Fuzzy Sets option is selected the user can view the Delete Fuzzy Sets panel where the user can delete the fuzzy set, view the updated fuzzy curves and then control goes back to the Operations Panel.

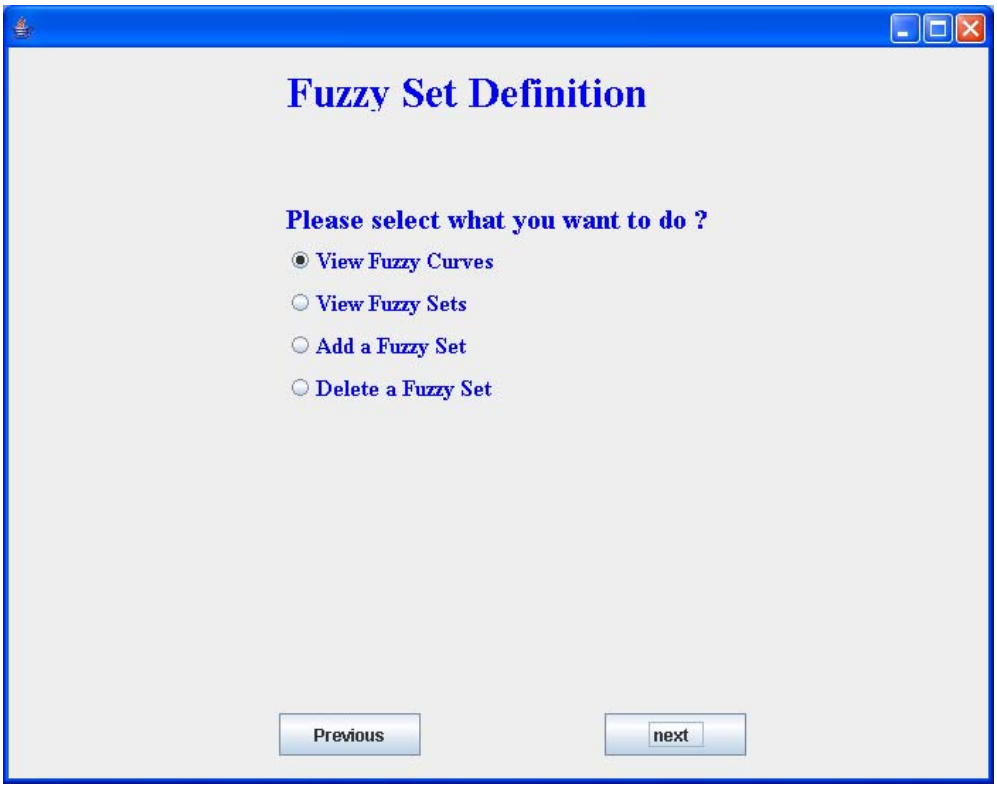


Fig. 2-22. Screenshot of the Fuzzy Set Operations Panel.

Figure 2.23 shows the Fuzzy Curves Input Panel, which shows the fuzzy curves for the selected variable. Users can view the fuzzy curves, parameters and change the membership type to Gaussian by selecting a fuzzy set. Figure 2.24 shows the Fuzzy Sets Input Panel, which contains the fuzzy sets and parameters of the selected variable. The user can modify the fuzzy sets by clicking the “Modify a set” button after changing the parameters. Figure 2.25 shows the Add Fuzzy Set Input Panel where users can add a Fuzzy Set Name with parameters and membership type to the selected variable. Figure 2.26 shows the Delete Fuzzy Set Panel where users can delete a fuzzy set of the selected variable.

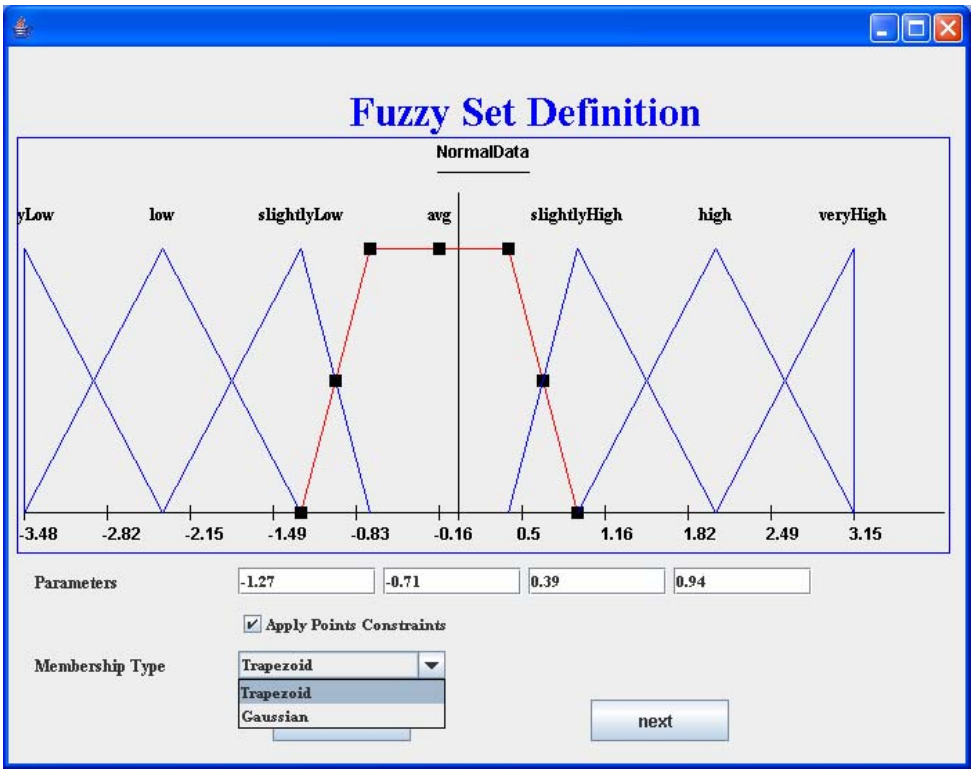


Fig. 2.23 Screenshot of the Fuzzy Curves Input Panel.

Fig. 2.24. Screenshot of the Fuzzy Sets Input Panel.

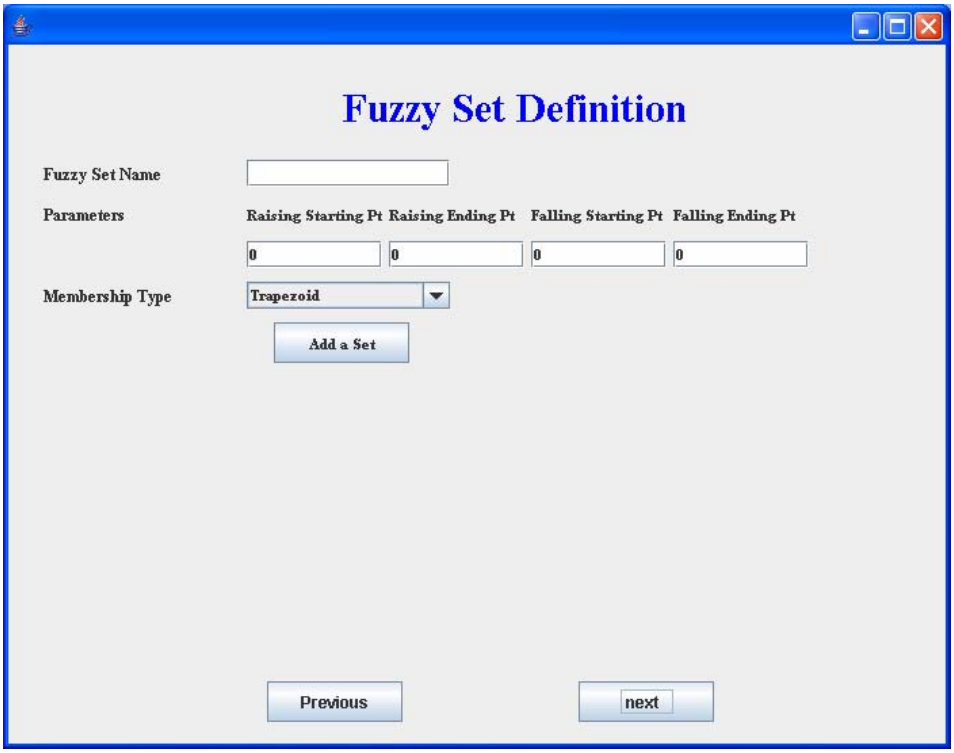


Fig. 2.25. Screenshot of the Add Fuzzy Sets Input Panel.

Fig. 2.26. Screenshot of the Delete Fuzzy Sets Input Panel.

Summary - Interfaces to Populate and Manage Knowledge Base Information

A user-friendly tool was developed for the CFS, which allows the user to build the knowledgebase. To integrate this process into the existing CFS software, Java was used as a programming language. The construction of the knowledgebase was identified as an important process of knowledge engineering. The architecture consists of three modules: data cleaning, normality testing, and fuzzy curve generation.

The two stages in which data cleaning is performed were identified as the data collection stage and the data analysis stage. A distance-based algorithm for data cleaning was proposed and analyzed. Subsequently, it was implemented, validated, and integrated into the main project. A K-S test was implemented to test the normality of the data. To give reasoning to a knowledgebase, fuzzy curves are important. The Fuzzy Set Definition algorithms are implemented for both *normal* and *not-normal* data. The fuzzy set definitions for *not-normal* data has four subcases. The fuzzy curves are viewed and modified using the fuzzy curve software. The interpolation algorithms, kriging algorithm and inverse distance weighted algorithms were also implemented.

This software, when implemented with the CFS package of software and wizards will aid in knowledgebase construction using efficient knowledge engineering techniques.

2.3 Data Management Subsystem

This project task involved creating a data management subsystem for the Customizable Fuzzy System (CFS). The data management subsystem for the CFS provides the user with the ability to perform a number of critical operations such as: view, modify, add and search, etc. on a binary data file. The data management subsystem operates on a binary data file (created by another module), which is in a tabular format using user-defined variables/properties, and default data. The integrity of the table format in the binary data file is strictly maintained with the use of white space in every user-defined variable/property of every record. The data management subsystem provides the functionality to add/delete/update a specific column in a binary data file. It also provides the function-

ality to add/delete/update a specific record in a binary data file. While the core software was largely completed during the first year, it was necessary to create user interfaces and links to wizards for other tasks. For this reporting period the work falls under Task 2.1 - Expert System Design Wizards.

Binary data formats were selected in response to potential user concerns that they would be forced to use particular relational database software. Noncommercial databases such as MySQL still require licensing fees if used for commercial purposes. In addition, non-commercial software presents a potential security risk that system administrators might find unacceptable. A final concern was updating software versions as a number of different and potentially incompatible releases for public domain software exist which could limit the functional portability of project-developed software.

Binary datafiles are compact and supported by all computer architectures which might use CFS software (computers capable of running aJava Virtual Machine). One issue that needed to be resolved was a method to swiftly search large binary tables. The work presented here represents our solution to that problem. Figure 2.27 shows the main interface screen of the resulting data management subsystem.

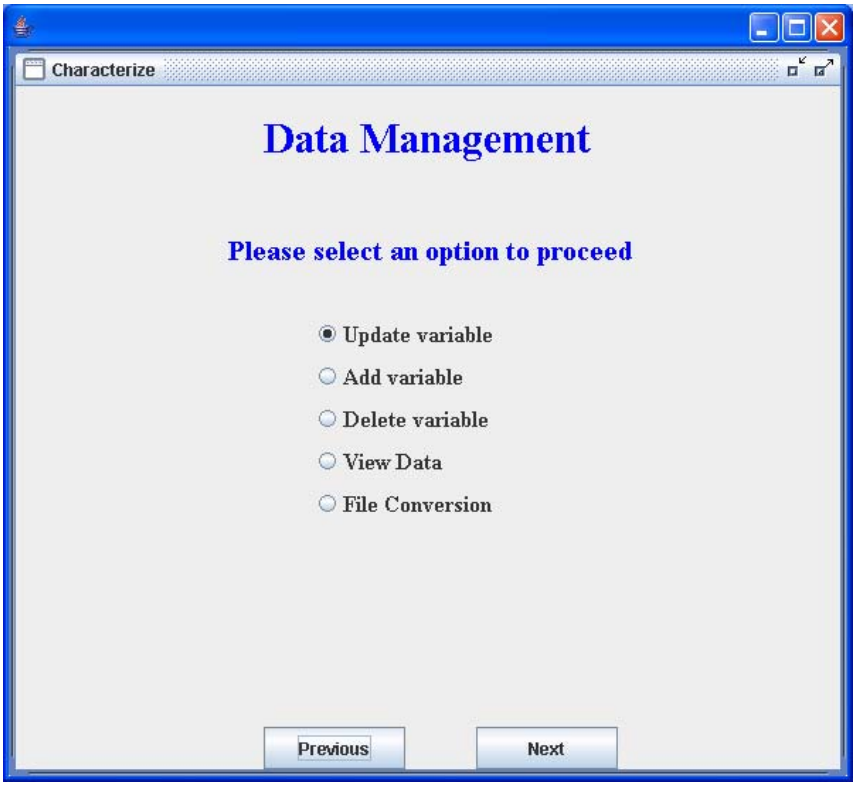


Fig. 2.27. Data management subsystem interface.

The data management subsystem utilizes a modular design. Figure 2.28 shows the architecture of the data management subsystem. The modules of the data management subsystem include the following:

- Data Retrieval Module
- Data Modification Module
- File Conversion Module
- Data Module
- Interface for the Customizable Fuzzy System (CFS)

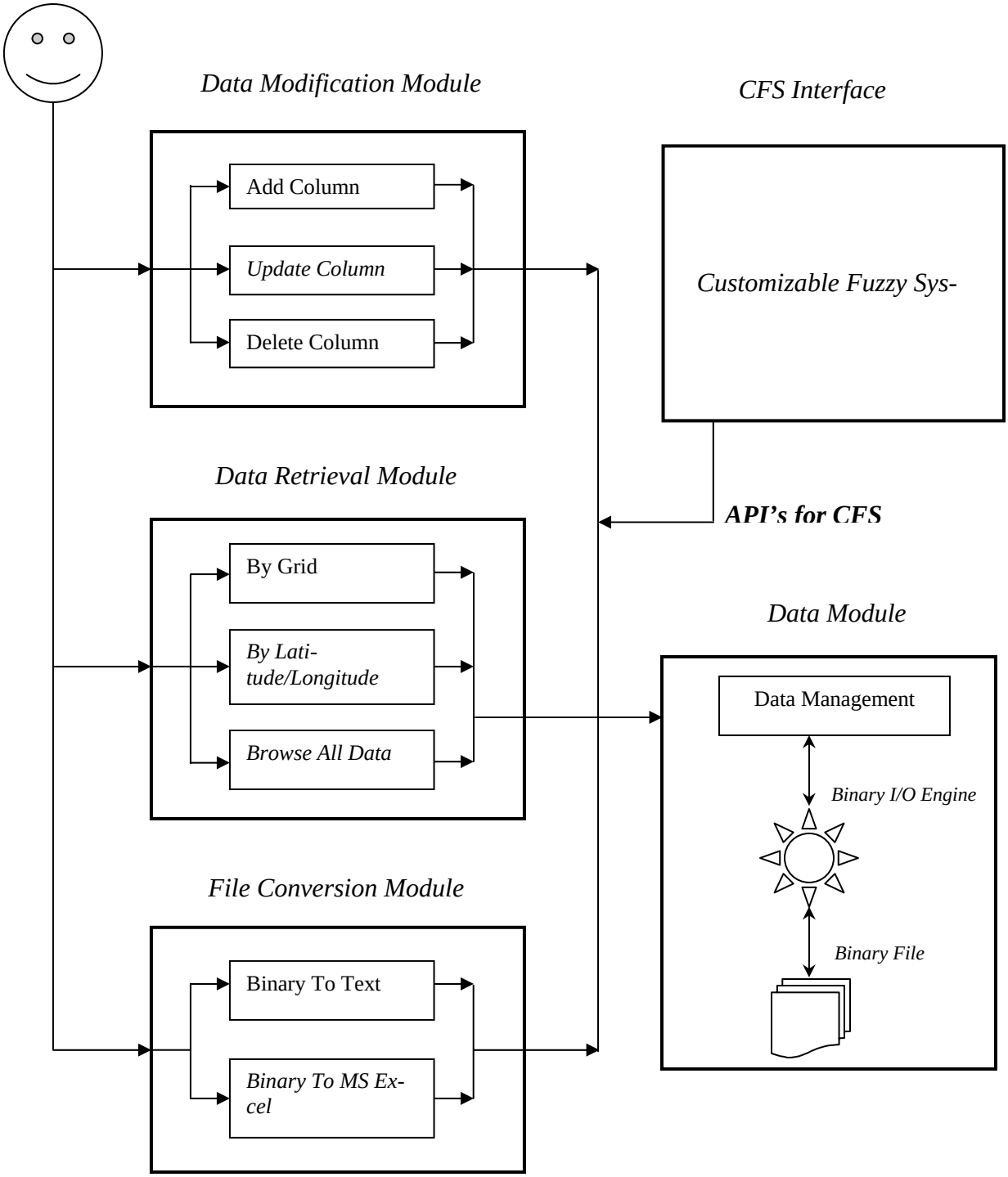


Fig. 2.28. Data management subsystem architecture.

Performance Analysis of the Binary Search Algorithm

Since it was expected that the data file could be very large, exceeding 50,000 records in many cases, searching for a specific record could be very time-consuming, particularly as the use of proprietary database software was precluded for the project. For this purpose a performance analysis, on an existing data set exceeding more than 60,000 records, was carried out using two experiments, sequential search and the binary search algorithm. The goal of this analysis was first to determine how effective the binary search algorithm could be in searching a particular record as compared to a sequential search in the entire binary data file. Second, if the binary search algorithm could reduce the search time significantly, then it should be adopted as the primary search algorithm for the data management subsystem. In both these experiments, the location (latitude and longitude) attributes were used to determine the exact record position of any particular data in the binary data file. In order to compare the results for both methods, a timer was set, which calculated the search time. After analyzing the results of both experiments it was apparent that the binary search algorithm reduced the search time well enough to allow the use of binary data tables in the project.

Sequential Search Experiment and Results: In the first experiment, the entire binary data file was searched sequentially for a particular location. A timer was started just before the search was to be carried out on a particular record in the binary data file. As the record was found the timer stopped and the time for the sequential search was recorded. Average times for three sequential searches for four different locations (latitude/longitude pair) are shown in Fig 2.29.

The results, as expected, were that as the record occurrence in the binary data file continued to increase, the search time also kept on increasing. The record pair 32.01417, -103.625 occurred early in the binary data file, whereas the record pair 32.80995, -103.8722 occurred later in the binary data file. For records sufficiently down the file the search time could be as long as 10 seconds, which was deemed unacceptable.

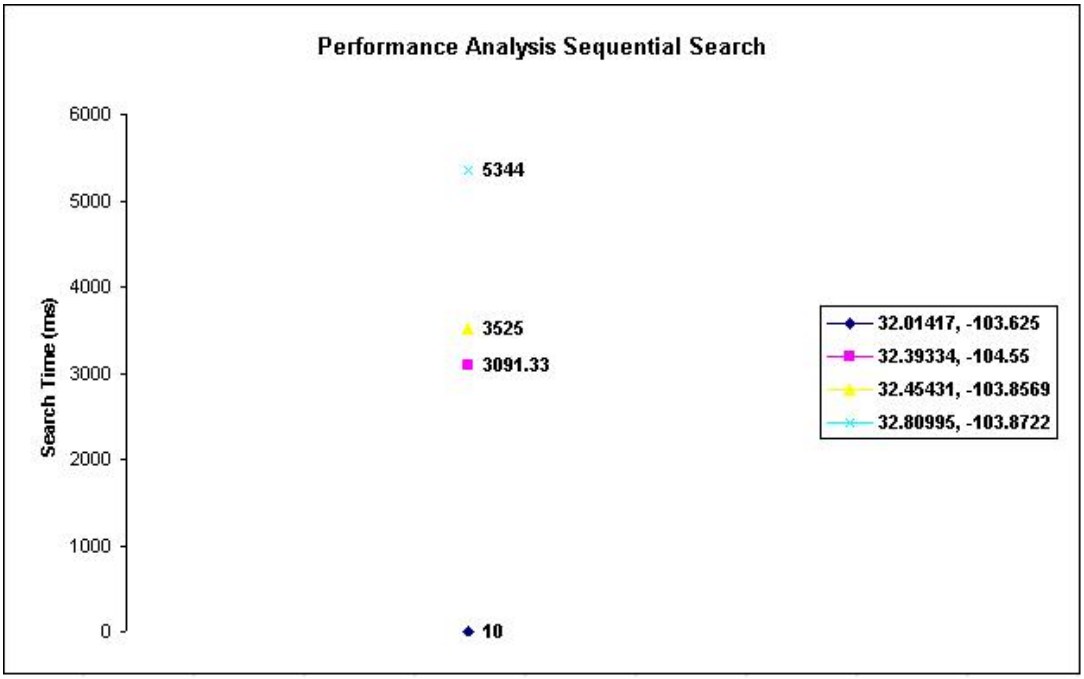


Fig 2.29. Performance analysis sequential search.

Binary Search Algorithm Experiment and Results: In the second experiment, a binary search algorithm was used to search for a particular record. A binary search algorithm is a technique for finding a particular value in a linear array by ruling out half the data at each step (bi-section). It was observed that the binary search algorithm considerably reduced the search time for a particular location. Initially, an index was created on the column "Latitude" of the binary data file. The index was created with two immediate values separated by a gap of 50 values. That is if the latitude values were sequential, starting from 1, then the index values would be 1, 51, 101, 151... and so on. After testing, the index technique, with a gap of 50 between two subsequent values, was found to be the most efficient record finding technique for the data set.

The binary search algorithm works by first determining the latitude range for the requested location within the binary search tree that is created out of the "Latitude"-based index file. Once the latitude range is determined the next step is to find an exact match for the latitude and longitude pair. Doing a sequential search, starting from the first record in the earlier determined latitude range, completes this operation.

In order to test the performance gain compared with the sequential search, a similar experiment was carried out. A timer was started just before the range of the value being searched was determined in the binary search tree. Once this range was determined, a sequential search was carried out for the searched value starting from the first record in the range in the binary data file. As the record was found the timer stopped and the time for the search was recorded. Average times for three searches, using binary search tree for four different locations (latitude/longitude pair), are shown in Fig. 2.30. These locations are the same as those used in the sequential search algorithm test.

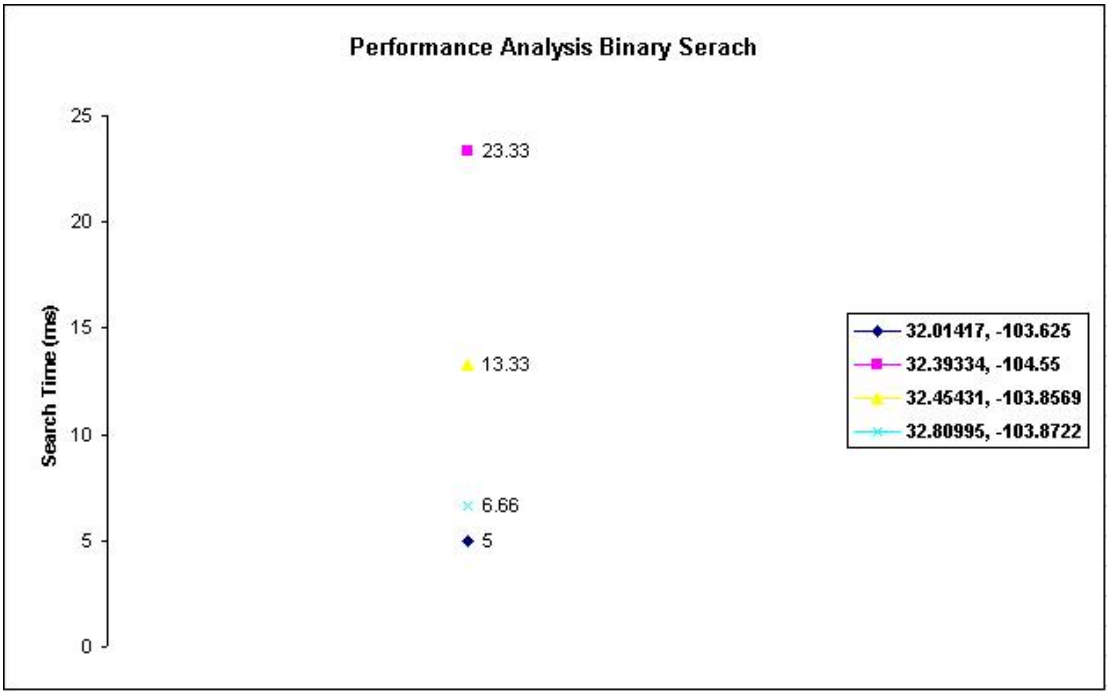
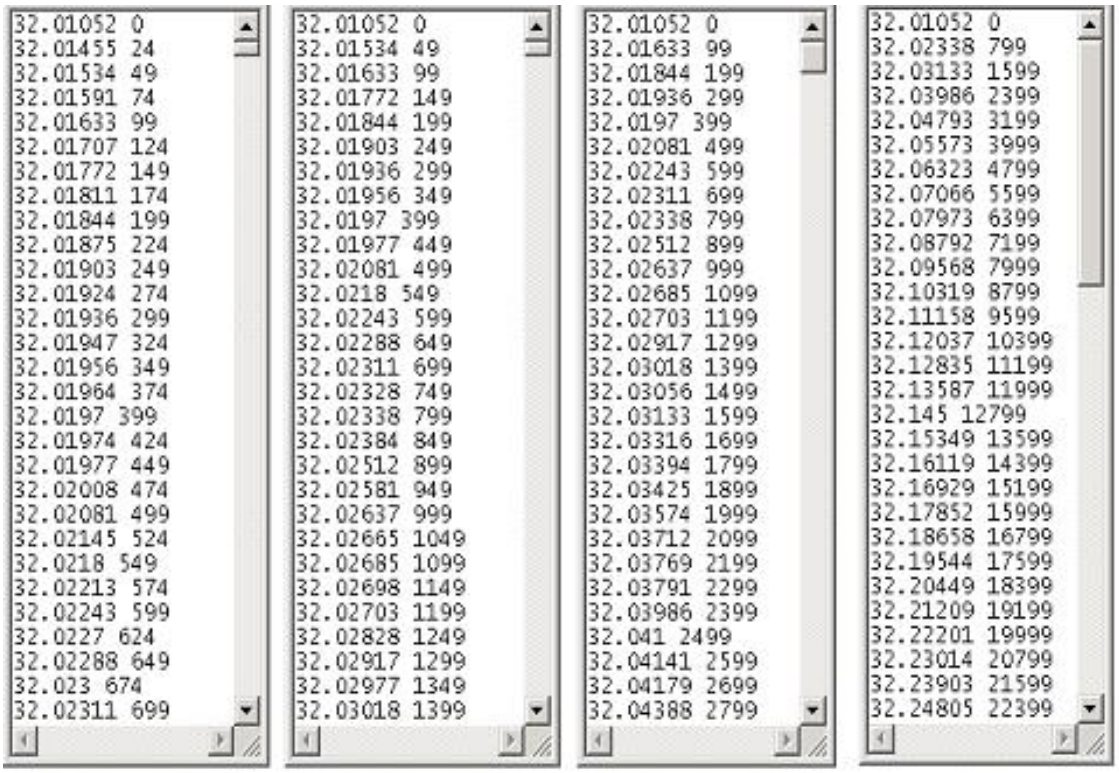


Fig 2.30. Performance analysis binary search.

The above results show that the binary search not only outperforms the sequential search for the problem but that the results were fast enough to justify the binary search algorithm as the primary search algorithm for the data management subsystem. For each CSF binary data file, an index file must be created and maintained as data is added or removed.

Index File Creation and Results: Seven different index files were generated before finally selecting one as the optimized one for the data management subsystem. Binary files of similar sizes should experience roughly similar search times and require roughly similar gaps. This will allow automation of the indexing process for CFS users. The seven different index files were created from the "Latitude" variable/column with a gap of 1, 25, 50, 100, 200, 400 and 800. That is a gap of 25 means that the 1st, 26th, 51st, 76th and so on records are stored in the index file with a gap of 25 between two subsequent values. Figure 2.31 is a screenshot of the index files with block sizes 25, 50, 100 and 800 respectively.



32.01052 0	32.01052 0	32.01052 0	32.01052 0
32.01455 24	32.01534 49	32.01633 99	32.02338 799
32.01534 49	32.01633 99	32.01844 199	32.03133 1599
32.01591 74	32.01772 149	32.01936 299	32.03986 2399
32.01633 99	32.01844 199	32.0197 399	32.04793 3199
32.01707 124	32.01903 249	32.02081 499	32.05573 3999
32.01772 149	32.01936 299	32.02243 599	32.06323 4799
32.01811 174	32.01956 349	32.02311 699	32.07066 5599
32.01844 199	32.0197 399	32.02338 799	32.07973 6399
32.01875 224	32.01977 449	32.02512 899	32.08792 7199
32.01903 249	32.02081 499	32.02637 999	32.09568 7999
32.01924 274	32.0218 549	32.02685 1099	32.10319 8799
32.01936 299	32.02243 599	32.02703 1199	32.11158 9599
32.01947 324	32.02288 649	32.02917 1299	32.12037 10399
32.01956 349	32.02311 699	32.03018 1399	32.12835 11199
32.01964 374	32.02328 749	32.03056 1499	32.13587 11999
32.0197 399	32.02338 799	32.03133 1599	32.145 12799
32.01974 424	32.02384 849	32.03316 1699	32.15349 13599
32.01977 449	32.02512 899	32.03394 1799	32.16119 14399
32.02008 474	32.02581 949	32.03425 1899	32.16929 15199
32.02081 499	32.02637 999	32.03574 1999	32.17852 15999
32.02145 524	32.02665 1049	32.03712 2099	32.18658 16799
32.0218 549	32.02685 1099	32.03769 2199	32.19544 17599
32.02213 574	32.02698 1149	32.03791 2299	32.20449 18399
32.02243 599	32.02703 1199	32.03986 2399	32.21209 19199
32.0227 624	32.02828 1249	32.041 2499	32.22201 19999
32.02288 649	32.02917 1299	32.04141 2599	32.23014 20799
32.023 674	32.02977 1349	32.04179 2699	32.23903 21599
32.02311 699	32.03018 1399	32.04388 2799	32.24805 22399

Fig 2.31. Index files with a gap of 25, 50, 100 and 800.

Since the data set on which the data management subsystem has to operate is potentially large, a sequential index file might not work efficiently. The way this process works is to first create a binary search tree for the index values from one of the index files. Starting from the index file, which has a gap of 1, the first step is to determine the range of the

latitude value for the data set. A timer is started just before the range of the latitude value, from the latitude/longitude pair, is determined in the binary search tree. Once this range is determined, a sequential search is carried out for the searched value starting from the first record in the range in the binary data file. As the record is found the timer stops and the time for the search is recorded.

Several experiments were conducted on different latitude/longitude pair values from the binary data file. Three tests were conducted on a particular latitude/longitude pair and the time it took to search a record in the binary data file using all of the seven index files was recorded. Thus, for a particular value, 21 tests were conducted. Then the average of the three tests was calculated for a particular index file to determine the time it takes to search a record in the binary data file. Below is a graph showing the time taken to search latitude/longitude pair values (32.01417, -103.625), (32.39334, -104.55) and (32.80995, -103.8722) for all the seven sparse index files.

It was observed that for the index file having block sizes 50 the search time was the minimum. That led to the selection of the index file having block size 50 as the primary index file for the data management subsystem, as shown in Fig 2.32.

The data management subsystem also provides the user the ability to convert the binary data file to text (.txt) format. The file conversion makes it possible to export data for use outside the data management subsystem. On initiation, the data management subsystem is provided with a binary data file generated by a separate wizard in another module. For better data integrity, the oil reservoir data is stored in a binary format. A snapshot of the data file on which the data management subsystem operates is shown in Fig 2.33. Each column in the data file represents a variable in the data management subsystem. The data management subsystem operates on these columns by performing different operations on them.

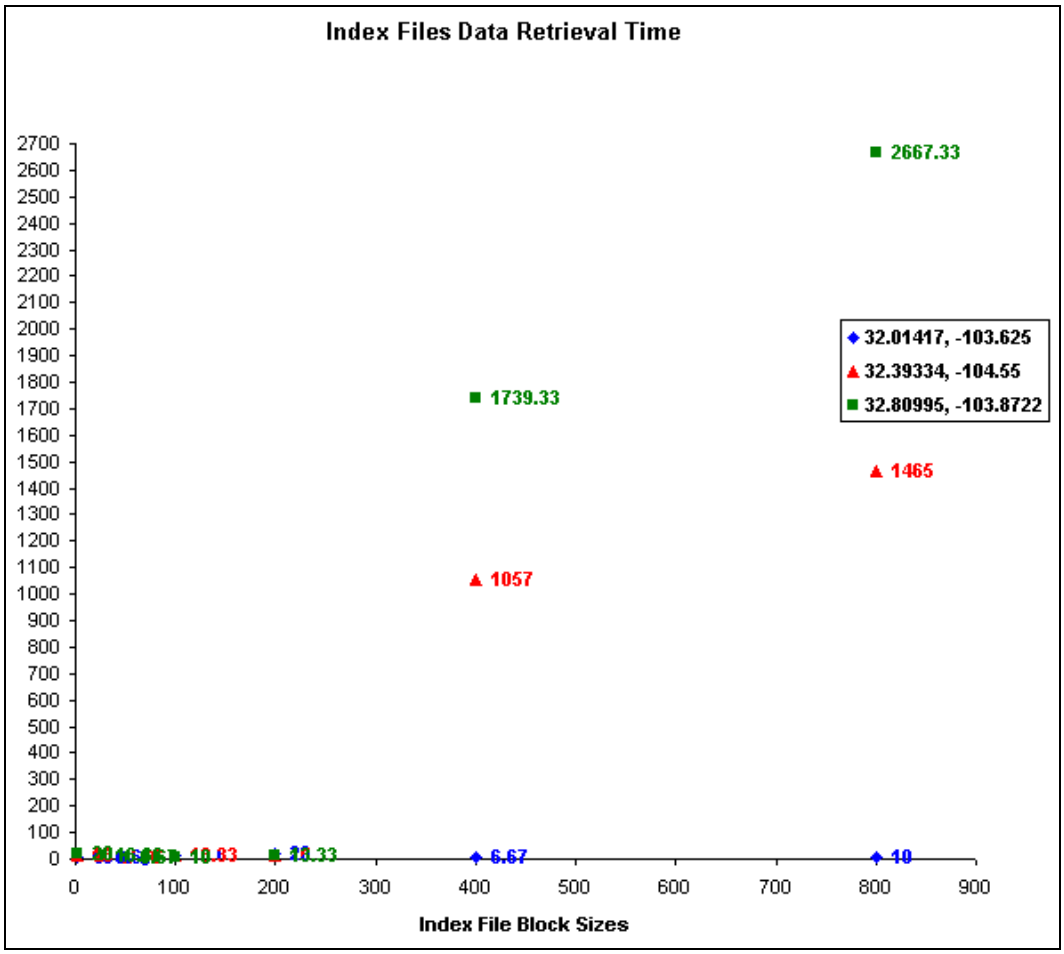


Fig. 2.32. Index files search time for different set of values.

165479.78	-3239.9	-5152.0	.....	4.533239	500
149463.41	-3810.9	-5152.0	.....	13.115721	500
141314.0	-4172.5	-5152.0	.....	1.1265149	500
134458.7	-4714.0	-5152.0	.....	0	500
126994.7	-5340.3	-5152.0	.....	0	500
114322.97	-5878.5	-5152.0	.....	0	500
108441.03	-5933.9	-5152.0	.....	0	500
103500.12	-5989.2	-4638.1	.....	0	1500
93001.471	-6028.2	-4638.1	.....	0	1500

Fig. 2.33. Data management subsystem, binary data file.

The initial format of the columns of the binary file, in each row/record, is shown in Table 2.12. The first five columns in Fig 2.34 are specific to the location of a reservoir. These five columns are not a part of the snapshot in Fig 2.33.

Table 2.12. Data Management Subsystem, Binary File Format

Column-Name	**Data Type**	**Size in Bytes**
	Int	4
X-Coordinate	Int	4
Y-Coordinate	Int	4
Latitude	Double	8
Longitude	Double	8
MinDistance	Double	8
DepthAtProspect	Double	8
DepthAtNearestWell	Grid#	8
. .	. .	. .
STDDeviation	Double	8
HighPredProd	Double	8

Each row in the binary data file is a complete record of all associated values with a particular location. After the user modifies, adds, deletes, or updates any column of the data file, through the data management subsystem, the size of a record in each row changes accordingly. The overall effect is to alter the size of the file; a representative file used for development purposes contains more than 60,000 records and occupies approximately 12MB of disk space.

Data Retrieval Module

The Data Retrieval module of the data management subsystem lets the user retrieve data from the data file on the basis of supplied parameters. There are different ways through which data can be retrieved. The view data screen (Fig 2.34) gives the user three choices to retrieve and display the data. The three options are: **By Grid, By Latitude / Longitude** and **Browse All Data.** When the user selects any of these options, the requisite controls are enabled and the others are disabled.

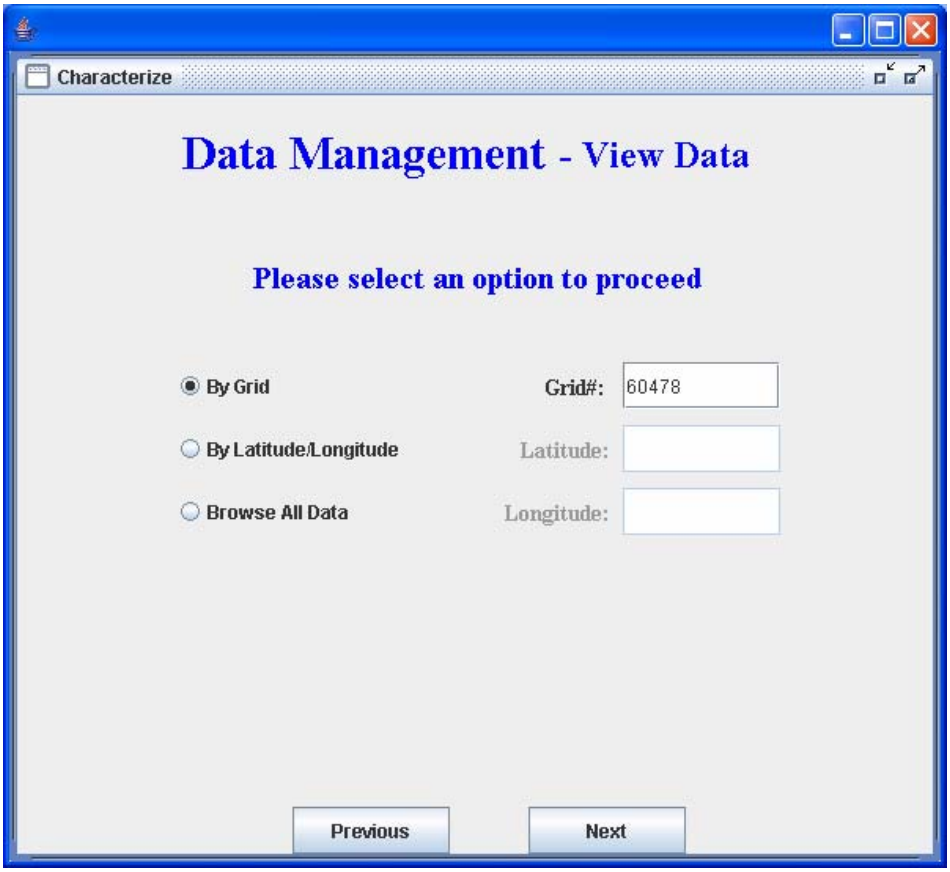


Fig 2.34. Data Retrieval module, By Grid option.

In Fig 2.34 the radio button to view the data on the basis of Grid# is selected. Now the user can enter the Grid# (as shown) before pressing the Next button. If data corresponding to the Grid# is available, the entire row/record will be displayed as in Fig 2.35, otherwise an appropriate message regarding the data unavailability will be displayed. The user can also choose to examine the data using the latitude and longitude values of the reservoir. In Fig 2.36 the user has selected to view the data using the latitude/longitude pair 32.01052/-103.1395.

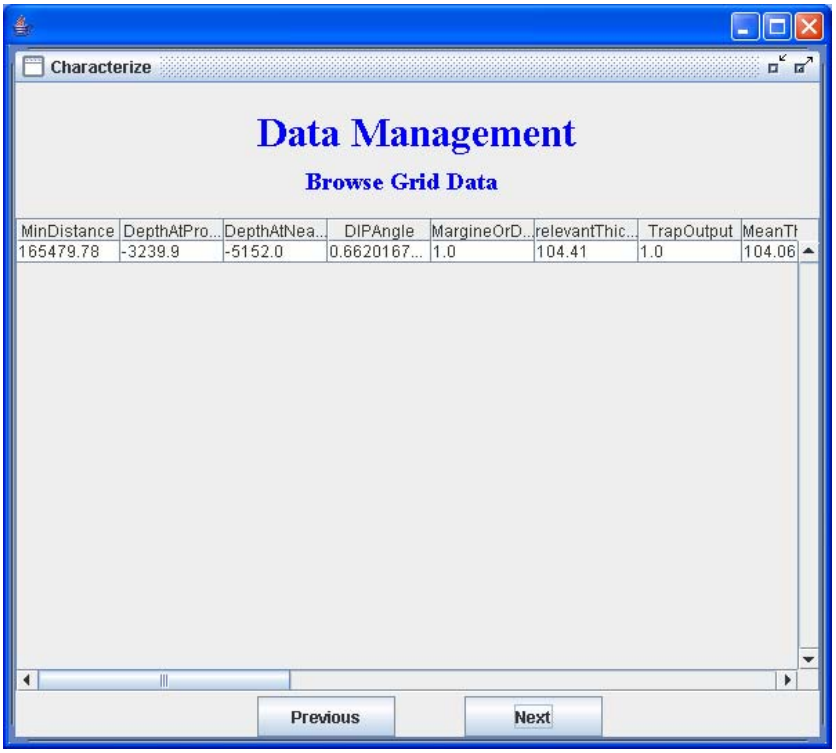


Fig. 2.35. Data Retrieval module output.

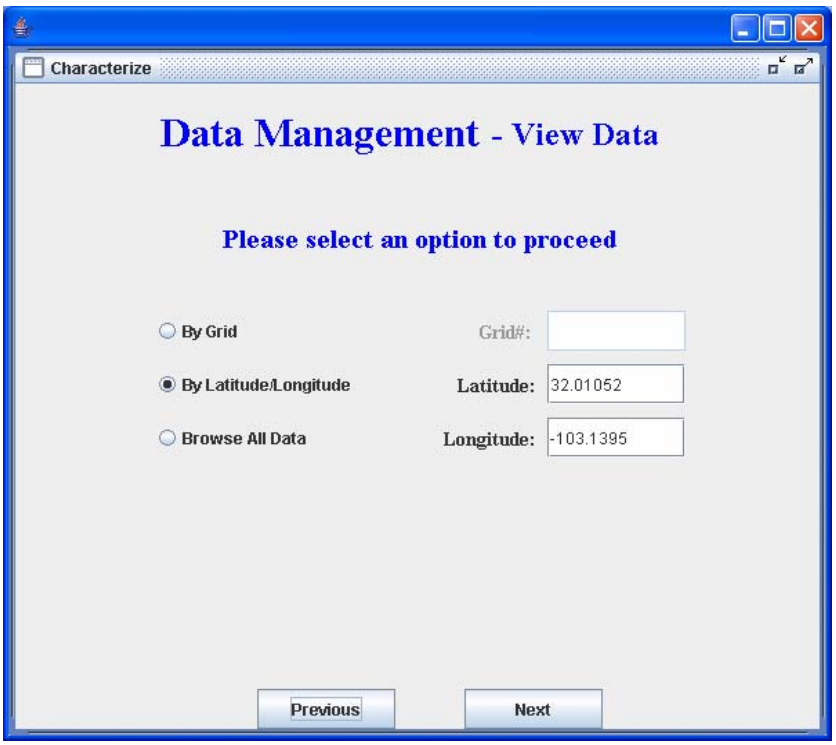
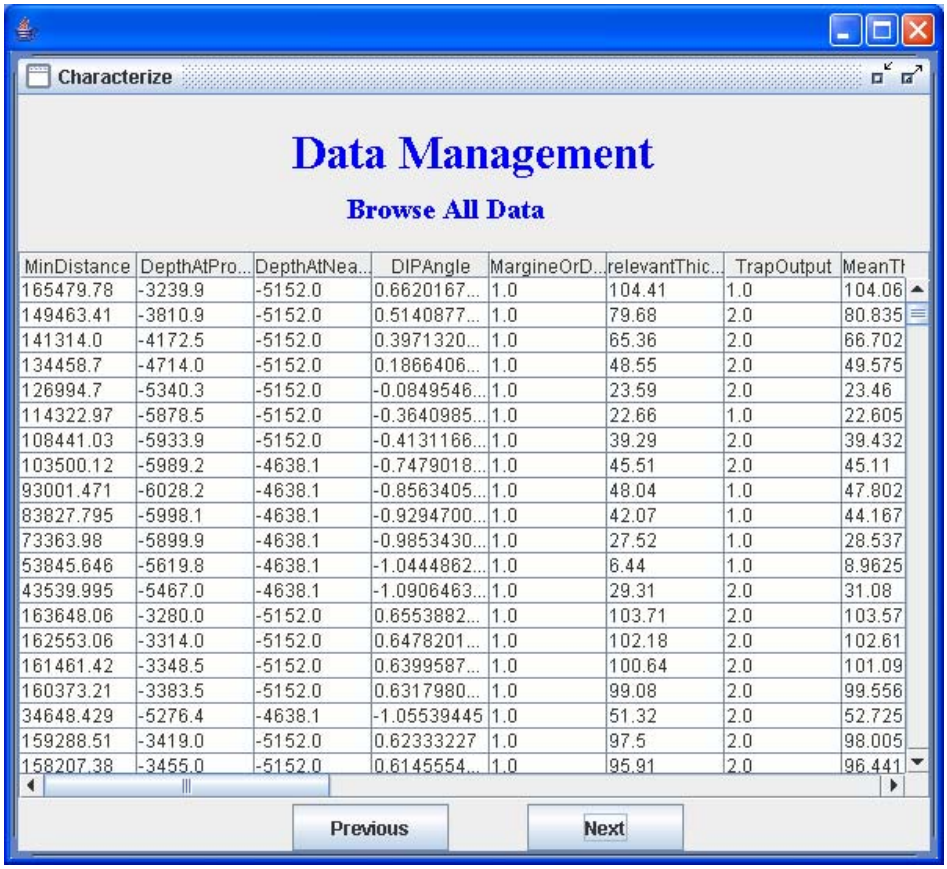


Fig. 2.36. Data Retrieval module, By Latitude/Longitude option.

The user can also elect to examine the entire data file at once using the “Browse All Data” option. The Browse All Data option is currently being re-implemented using Java pipe and thread concepts to optimize load times. The entire process is averaging 30–60 seconds depending on hardware. This is understandable as fetching the entire data set from the binary data file (60,000 records test data) is quite an enormous task. However, after the implementation of the pipe and thread concepts, the data should be retrieved much faster. Initially, the main application thread first retrieves the first 100 records. At that point a pipe is created. On the receiving end of the pipe is the main application, waiting for the next chunk of records (100 or less records). On the other end of the pipe another thread is created that sequentially requests the next chunk of 100 records (from the Data module) until the entire binary data file gets exhausted and all the records are retrieved. Fig 2.37 shows the screens through which the data is displayed using the Browse All Data option.



MinDistance	DepthAtPro...	DepthAtNea...	DIPAngle	MargineOrD...	relevantThic...	TrapOutput	MeanTh
165479.78	-3239.9	-5152.0	0.6620167...	1.0	104.41	1.0	104.06
149463.41	-3810.9	-5152.0	0.5140877...	1.0	79.68	2.0	80.835
141314.0	-4172.5	-5152.0	0.3971320...	1.0	65.36	2.0	66.702
134458.7	-4714.0	-5152.0	0.1866406...	1.0	48.55	2.0	49.575
126994.7	-5340.3	-5152.0	-0.0849546...	1.0	23.59	2.0	23.46
114322.97	-5878.5	-5152.0	-0.3640985...	1.0	22.66	1.0	22.605
108441.03	-5933.9	-5152.0	-0.4131166...	1.0	39.29	2.0	39.432
103500.12	-5989.2	-4638.1	-0.7479018...	1.0	45.51	2.0	45.11
93001.471	-6028.2	-4638.1	-0.8563405...	1.0	48.04	1.0	47.802
83827.795	-5998.1	-4638.1	-0.9294700...	1.0	42.07	1.0	44.167
73363.98	-5899.9	-4638.1	-0.9853430...	1.0	27.52	1.0	28.537
53845.646	-5619.8	-4638.1	-1.0444862...	1.0	6.44	1.0	8.9625
43539.995	-5467.0	-4638.1	-1.0906463...	1.0	29.31	2.0	31.08
163648.06	-3280.0	-5152.0	0.6553882...	1.0	103.71	2.0	103.57
162553.06	-3314.0	-5152.0	0.6478201...	1.0	102.18	2.0	102.61
161461.42	-3348.5	-5152.0	0.6399587...	1.0	100.64	2.0	101.09
160373.21	-3383.5	-5152.0	0.6317980...	1.0	99.08	2.0	99.556
34648.429	-5276.4	-4638.1	-1.05539445	1.0	51.32	2.0	52.725
159288.51	-3419.0	-5152.0	0.62333227	1.0	97.5	2.0	98.005
158207.38	-3455.0	-5152.0	0.6145554...	1.0	95.91	2.0	96.441

Fig. 2.37. Data Retrieval module, All Data result.

Data Modification Module

The Data Modification module is an important part of the data management subsystem for the CFS that allows users to make and save changes to the data file for their projects.

Add Variable/Column: The Add Variable/Column functionality provides the user with the ability to add more variables/columns to the existing data management subsystem binary data file. The effect resembles the addition of an attribute to an existing table in a relational database at run-time. A drop-down list of all the variables/columns along with their data types is pulled from a look-up file generated by another module. Only those variables/properties that are not a part of the current binary file are displayed in the drop down list of the Add Variable/Column module. If the user wants to add another column to the data file structure then the Add Variable option is selected on the main screen (Fig 2.38). This will take the user to the Add Variable main screen as shown in Fig 2.39.

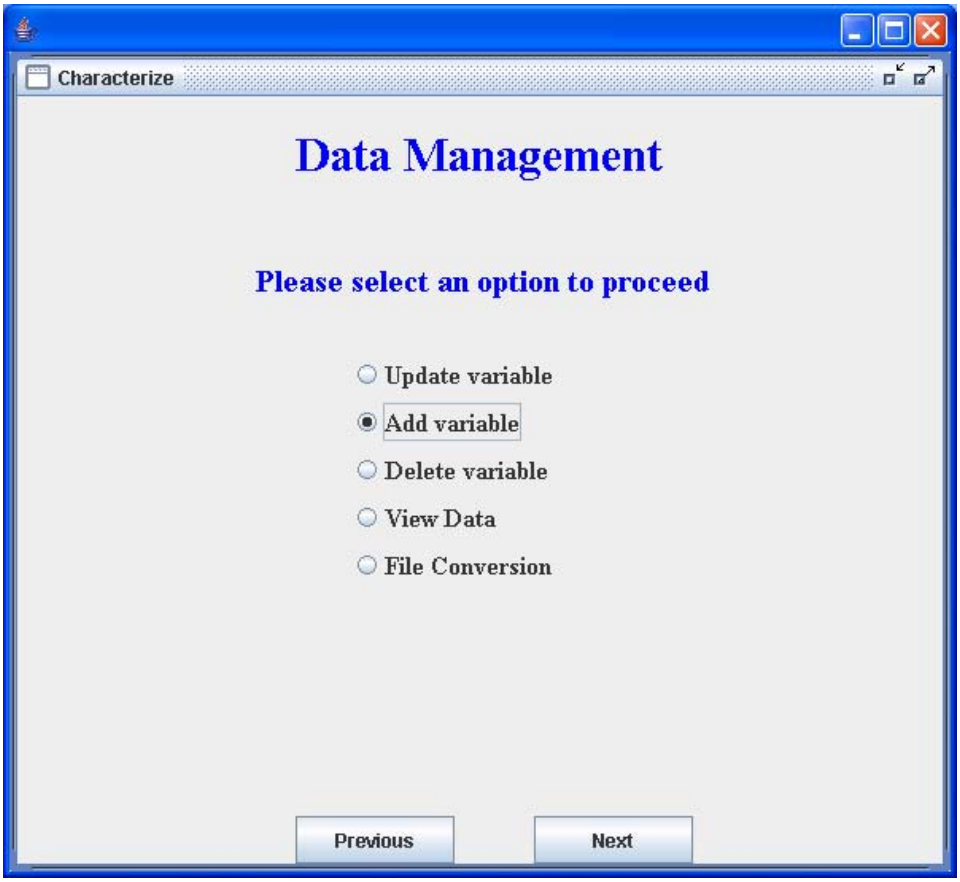


Fig. 2.38. Data Modification module, Add Variable option.

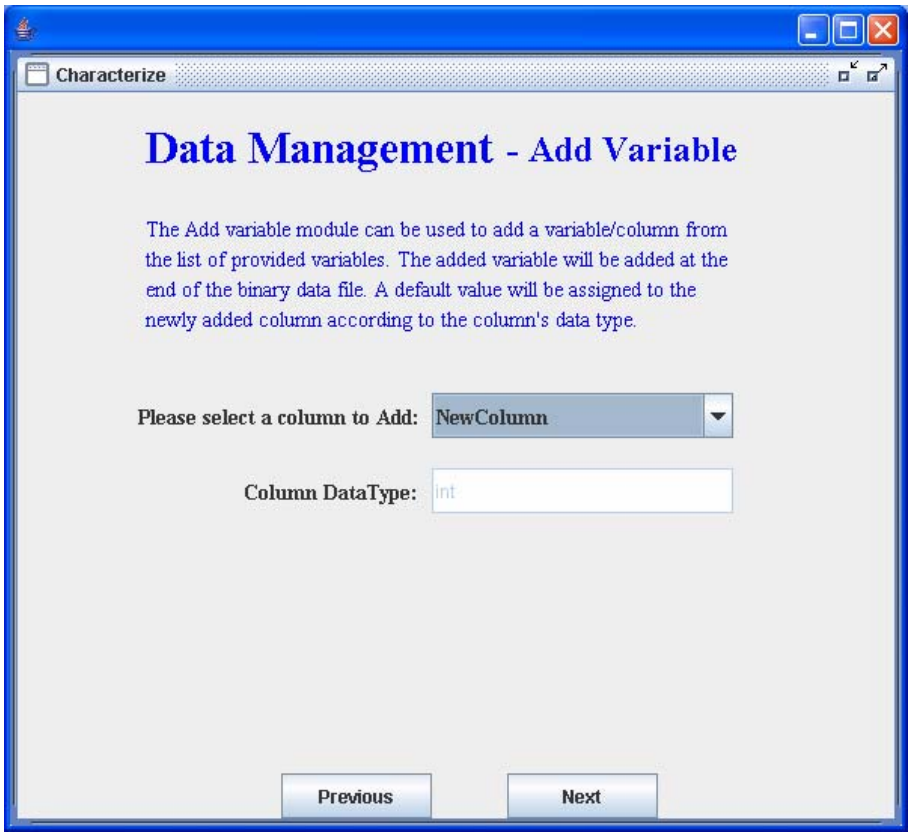


Fig. 2.39. Data Modification module, Add Variable screen.

In Fig 2.39 the user has elected to add the column "NewColumn" with data type "int". After the user presses the Next button the data file's initial data and format (Fig 2.33 and Table 2.12) change to that shown in Fig 2.40 and Table 2.13 respectively.

***NewColumn**: default value -99999*

165479.78	-3239.9	-5152.0	.....	4.533239	500	-99999
149463.41	-3810.9	-5152.0	.....	13.115721	500	-99999
141314.0	-4172.5	-5152.0	.....	1.1265149	500	-99999
134458.7	-4714.0	-5152.0	.....	0	500	-99999
126994.7	-5340.3	-5152.0	.....	0	500	-99999
114322.97	-5878.5	-5152.0	.....	0	500	-99999
108441.03	-5933.9	-5152.0	.....	0	500	-99999
103500.12	-5989.2	-4638.1	.....	0	1500	-99999
93001.471	-6028.2	-4638.1	.....	0	1500	-99999

Fig. 2.40. Data management subsystem, binary data file after Add.

Table 2.13. Data Management Subsystem, Binary File Format after Add

Column-Name	***Data Type***	***Size in Bytes***
Grid#	*Int*	*4*
X-Coordinate	*Int*	*4*
Y-Coordinate	*Int*	*4*
Latitude	*Double*	*8*
Longitude	*Double*	*8*
MinDistance	*Double*	*8*
DepthAtProspect	*Double*	*8*
DepthAtNearestWell	*Double*	*8*
. . .	. . .	. . .
STDDeviation	*Double*	*8*
HighPredProd	*Double*	*8*
NewColumn	*Int*	*4*

Update a Variable/Column: If the user wants to update the data contained within a column to the data file structure *Variable* then the user would select the Update Variable feature on the main screen (Fig. 2.41) before pressing the Next button. This functionality provides the user with a drop-down list of all the variables/columns that are currently available in the binary data file. In the Update variable screen, the user is prompted to select a column from the drop-down list before pressing the Next button. In Fig. 2.42 the user has selected the column “MinDistance” for updating.

The next screen is the data file selection screen (Fig. 2.43). The user in this case selects a text file whose data values need to be updated in the main binary data file. This text file can contain any number of records ranging from 1 to the maximum record size of the binary file. After the user makes this selection the data management module processes the request and the data in the binary data file is updated for the corresponding records in the text file. The system alerts the user by displaying an appropriate message that the modifications are completed.

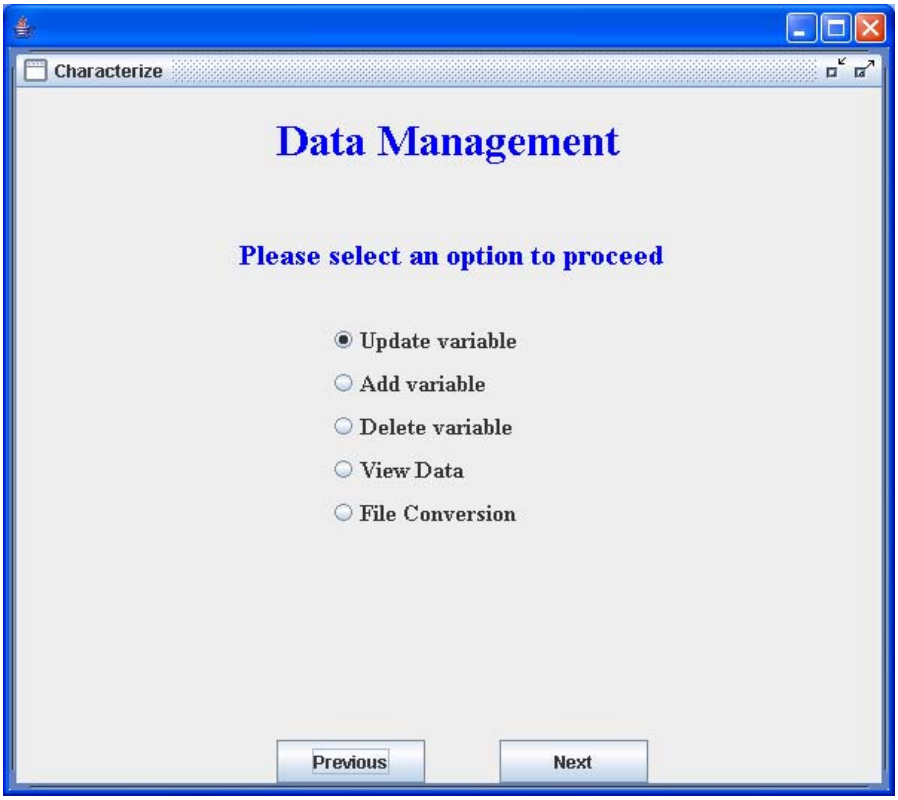


Fig. 2.41. Data management subsystem, Update option.

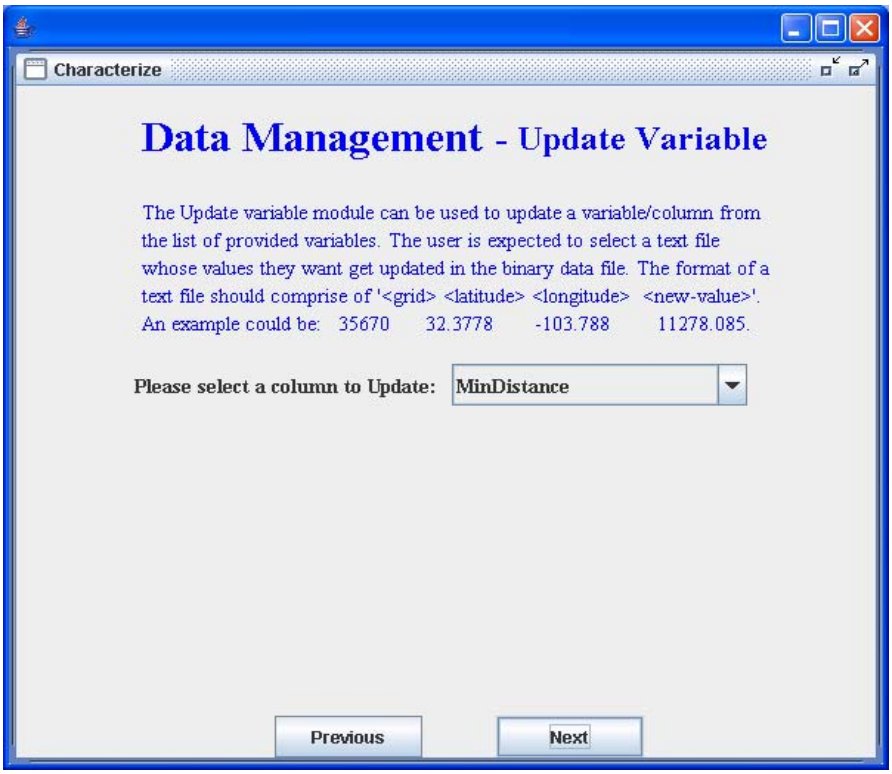


Fig. 2.42. Data management subsystem, Update Variable screen.

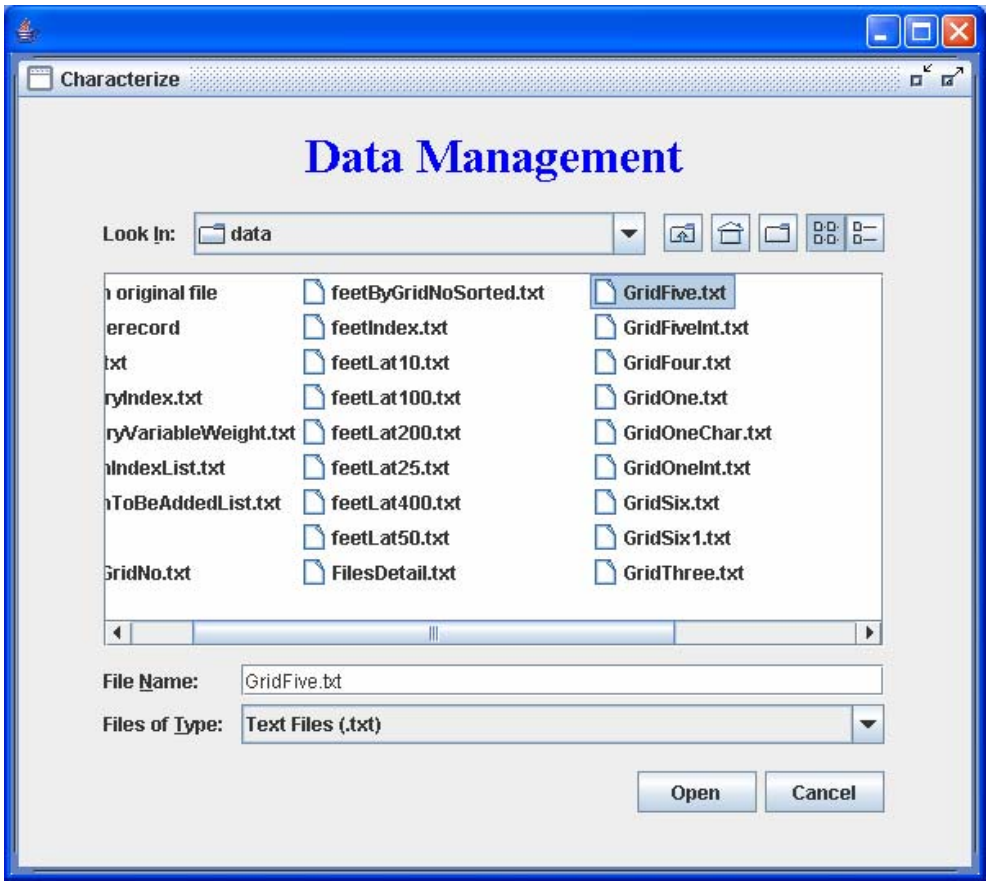


Fig. 2.43. Data management subsystem, Data File Selection.

Delete a Variable/Column: If the user wants to delete a column from the binary data file structure then the user would select the Delete Variable option on the main screen before pressing the Next button. In the Delete Variable main screen the user then selects a column that is to be deleted from the list of provided columns. When the user presses the Next button, the column is deleted and the user is notified about the modification. In Fig. 2.44 the user wants to delete the column "DepthAtProspect". After the user presses the Next button, the data file's data and format are changed to those shown in Fig. 2.45 and Table 2.14 respectively.

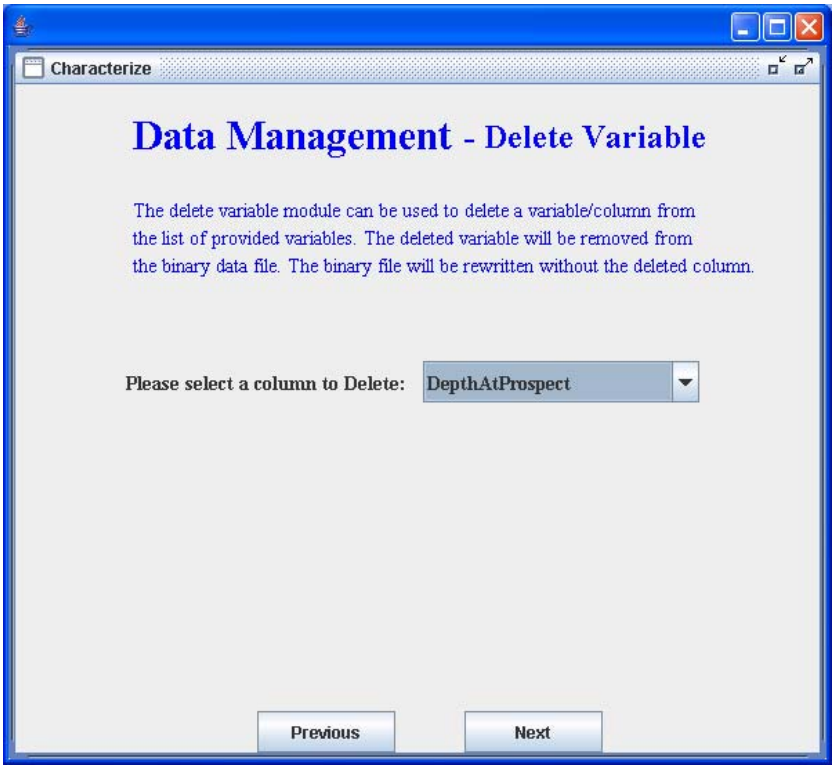


Fig. 2.44. Data management subsystem, Delete Variable screen.

```
165479.78   -5152.0 ........ 4.533239     500   -99999
149463.41   -5152.0 ........ 13.115721    500   -99999
141314.0    -5152.0 ........ 1.1265149    500   -99999
134458.7    -5152.0 ........ 0            500   -99999
126994.7    -5152.0 ........ 0            500   -99999
114322.97   -5152.0 ........ 0            500   -99999
108441.03   -5152.0 ........ 0            500   -99999
103500.12   -4638.1 ........ 0            1500  -99999
93001.471   -4638.1 ........ 0            1500  -99999
```

Fig. 2.45. Data management subsystem, binary data file after Delete.

Table 2.14. Data Management Subsystem, Binary File Format after Delete

Column-Name	***Data Type***	***Size in Bytes***
Grid#	*Int*	*4*
X-Coordinate	*Int*	*4*
Y-Coordinate	*Int*	*4*
Latitude	*Double*	*8*
Longitude	*Double*	*8*
MinDistance	*Double*	*8*
DepthAtNearestWell	*Double*	*8*
. . .	. . .	. . .
STDDeviation	*Double*	*8*
HighPredProd	*Double*	*8*
NewColumn	*Int*	*4*

File Conversion Module

The data management subsystem includes the function of converting the main binary data file to a text file (Figs. 2.46 and 2.47). this allows export of data into other programs, such as excel or a relational database.

Fig. 2.46. Binary to text conversion—Confirmation Dialog.

Fig. 2.47. Binary to text conversion—Notification Dialog.

Data Module Summary

The Data Module is local to the data management subsystem. It is not accessible from outside the data management subsystem. Only the Data Retrieval, Data Modification, File Conversion and Interface for the Customizable Fuzzy System modules can access the Data Module. It is generally expected the database will be populated automatically and the only need of a user would be to remove or add specific data streams. Wizards to accomplish these tasks will be available for less computer-savvy users.

Interface for the Customizable Fuzzy System (CFS): The data management subsystem is an independent application. However, the data management subsystem is developed for data management of the CFS. There has to be a means through which the CFS can com-

municate with the data management subsystem. For this purpose, APIs (application programming interfaces) have been designed to allow communication with the CFS.

Summary—Data Management Subsystem

The data management of the drilling data, a core component of the CFS, requires robust implementation for the success of the CFS. The data management subsystem, with its user interface, provides a platform from where operations on the CFS's binary data are performed efficiently and accurately. The binary data file is kept in a tabular format using user-defined variables/properties and default data. In order to maintain the integrity of the entire binary data file, white space is used in every user-defined variable/property of every record. The data management subsystem is designed in different modules where a particular functionality is restricted to that specific module.

The data management subsystem with all the current features in place is working as expected. The data retrieval module retrieves and displays the data for the requested grid#, latitude/longitude pair and the entire data file successfully.

The data modification's Add Variable/Column module successfully lists the columns that are not part of the data management system's existing binary data file. Once the user selects to add a particular column that column with its data type is added to the entire binary data file with a default value. The Update Variable/Column functionality updates a particular variable/column of the binary data file. The functionality accomplishes the task by first allowing the user select a variable/column from an existing binary data file. Second, the functionality allows the user to select a text file whose values need to be updated in the main binary data file. An appropriate message, confirming the modifications are complete, describes the success of the functionality. The Delete Variable/Column module successfully processes the user request and deletes a variable/column specified by the user from the entire binary data file.

The file conversion module successfully converts the data management subsystem's entire binary data file into a text file. The success of the module is verified by comparing the original text file with the text file generated by the file conversion module.

The data module, internal to the data management subsystem, deals with all the processing on the binary data file requested by the data retrieval, data modification and file conversion modules. The module is verified with the successful working of the data retrieval, data modification and file conversion modules.

The interface for the CFS is tested by integrating the data management subsystem with the entire customizable fuzzy system. All the data management subsystem's modules work as specified with the customizable fuzzy system after the integration.

2.4 Acquiring and Converting Mapped Data

The knowledgebase is built using the pre-existing answer base for play type, user provided data at specific locations and interpolated user provided data. An important feature is the ability to assimilate digital map data and to map both input data and CFS results. Figure 2.48 illustrates how the CFS utilizes digitally mapped data in the construction of a knowledgebase utilizing user-provided data interpolated over a geographic region. Addition of Wizard-based software for processing such data and requests is scheduled for the third project year. The algorithm development and associated code was completed in the second project year.

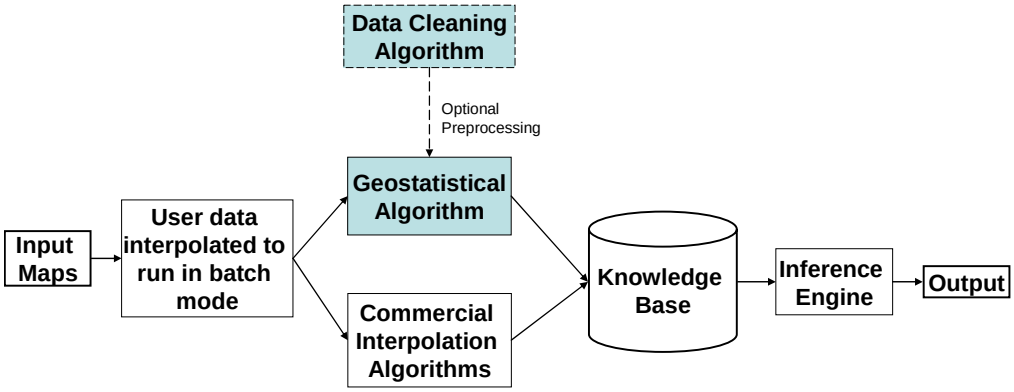


Fig. 2.48. Use of map data in knowledgebase generation for the CFS.

Geostatistical Analysis

The geostatistical algorithms—ordinary kriging and Inverse Distance Weighted (IDW)—were developed to interpolate the unknown values over the region based on the measured values at the specified locations. The IDW algorithm works by calculating the influence of a measured value on a grid point as a function of its distance. The kriging algorithm considers the spatial autocorrelation of the data by looking at the variogram, and uses properties of the variogram to calculate the weights. The weights, in turn, are used to calculate the values at the grid points.

In this tool, the IDW method is recommended when the input data consists of more than 30 locations. The Data Cleaning module is available for preprocessing this data prior to interpolating.

The first steps in the geostatistical algorithms are the same for either method, and involves reading in the user's values and then setting up the grid. A rectangular grid is used for this code, and can be set by default to be the minimum rectangle based on the user's data, or can be adjusted to any desired size. This allows the user to set up one grid for the entire process, and repeat the gridding algorithm on various variables to come up with one answer base.

Variables xk and yk should be in feet or meters; this makes a difference in the bin setup in the ordinary kriging algorithm. K total is the number of known points, zk is the measured value at the point (xk, yk) and can be any measured value in a variety of units (e.g. porosity, permeability, gas production, TOC). If K > 30, IDW is used, otherwise ordinary kriging is used.

The output grid is designed with six parameters:
xmin, xmax (the default here can be the smallest and largest xk values)
ymin, ymax (the default here can be the smallest and largest yk values)
xnum, ynum: number of cells in the x and y directions

xstep, ystep: dimensions (size) of the cell, which is calculated as follows:

$$xstep = \frac{x\max - x\min}{xnum}$$

$$ystep = \frac{y\max - y\min}{ynum}$$

Inverse Distance Weighted (IDW): The Inverse Distance Weighted algorithm calculates the z values at a location using the inverse distance weight measure. Figure. 2.49 describes the algorithm that interpolates the values of z value (i,j) for the given number of cells in x and y direction using the inverse distance weights. The values of xk, yk and zk are provided in the input file.

```
for i = 0 to xnum-1 do
   for j = 0 to ynum-1 do
      xcoordinate(i,j) = xmin + i*xstep
       ycoordinate(i,j) = ymin + j*ystep
      sumw = 0
      for ik = 1 to K do
```

$$DistK(ik) = \sqrt{(xcoordinate(i,j) - xk(ik))^2 + (ycoordiante(i,j) - yk(ik))^2}$$

$$w1(ik) = \frac{1}{DistK(ik)^2}$$

$$sumw = sumw + w1(ik)$$

```
            next
      zvalue(i,j) = 0
       for ik = 1 to K do
```

$$weight(ik) = {w1(ik)}/{sumw}$$

$$zvalue(i,j) = zvalue + weight(ik) * zk(ik)$$

```
            next
      next
next
```

Fig. 2.49. Algorithm: Inverse Distance Weighted (IDW).

Ordinary Kriging: This is a quick kriging algorithm, with a variogram calculated from a pre-defined set of bins and then fit with a simple estimate to get a value for sill, nugget and range. The sill, nugget and range values are then plugged into an exponential equation to calculate the C matrix between each pair of measured data. The matrix is then inverted and multiplied by the vector of covariance using distances from the gridpoint to the known data. This produces the weights, which are then normalized and applied as in the previous algorithm.

1. Calculate distances between data points:

```
for i = 1 to K do
     for j = 1 to K do
```

$$DK(i,j) = \sqrt{(xk(i) - xk(j))^2 + (yk(i) - yk(j))^2}$$

```
                next
           next
```

2. Bin data and compute points for a variogram:

a. Find max of the DK(i,j) to get the upper bound for the bins

```
MaxDK = 0
for i = 1 to K do
   for j = 1 to K do
          if DK(i,j) > MaxDK
          then MaxDK = DK(i,j)
     next
next
```

b. Calculate the number of bins and the binned variogram output:

```
lag_ft = 5000
binnumber = roundup(MaxDK/lag_ft)
for m = 1 to binnumber do
```

```
        H(m)=(lag_ft/2)+(m-1)(lag_ft)
        vardiff = 0
        for i = 1 to K do
           for j = 1 to K do
              if (m-1)(lag_ft) ≤ DK(i,j) < m(lag_ft)
              then N(m) = N(m) +1
              vardiff = (zk(i)-zk(j))^2 + vardiff
                          next
                   next
         gamma(m) = 1/(2N(m)) * vardiff
      next
```

c. Estimate sill, range and nugget:

```
      sum = 0
        for m = 1 to binnumber do
           sum = sum + gamma(m)
        next
        meangamma = sum/binnumber
        stsum = 0
        for m = 1 to binnumber
           stsum = (gamma(m) – meangamma)^2 + stsum
      next
        ssqr = stsum/(binnumber – 1)
        stdevgamma = √(ssqr)
        maxgamma = 0
        for m = 1 to binnumber do
                if gamma(m) > maxgamma
           then maxgamma = gamma(m)
                next
        if maxgamma < meangamma + 2* stdevgamma
         then sill = maxgamma
```

```
else sill = meangamma + 2* stdevgamma
end if
for m = 1 to binnumber do
     if gamma(m) ≥ 0.95* sill
   then range = H(m)
next
```

$$y\text{int} = gamma(2) - \left(\frac{gamma(3) - gamma(2)}{H(3) - H(2)} \right) * H(2)$$

```
if yint < 0 then nugget = 0
else nugget = yint
```

3. Build matrix C (using the covariance):

```
C1 = sill - nugget
for i = 1 to K do
   for j = 1 to K do
```

$$\text{if DK(i,j)} \neq 0 \text{ then C(i,j) = } C1 * e^{\frac{-3DK(i,j)}{range}}$$

```
      if DK(i,j) = 0 then C(i,j) = sill
      next
next
for i = 1 to K do
    C(i,K+1) = 1
next
for j = 1 to K do
    C(K+1,j) = 1
next
C(K+1,K+1) = 0
```

4. Invert the matrix C using Gauss-Jordan with full pivoting by LU decomposition.

5. Build the DCmat vector:

```
for i = 0 to xnum-1 do
    for j = 0 to ynum-1 do
```

```
        for m = 1 to K do
                    DP(m) =
```

$$\sqrt{(xcoordinate(i,j)-xk(m))^2+(ycoordinate(i,j)-yk(m))^2}$$

```
                    if DP(m) ≠ 0 then DCmat(m) =
```

$$C1*e^{\frac{-3DP(m)}{range}}$$

```
                    if DP(m) = 0 then DCmat(m) =
sill
            next
        DP(K+1) = 1
      next
next
```

6. Multiply the inverse of C by D to get the weights.

7. Calculate the normalized weights and the estimate:

```
sumw = 0
for m = 1 to K do
   sumw = w1(m) + sumw
          next
zvalue(i,j) = 0
for m = 1 to K do
```

$$weight(m) = {w1(m)}/{sumw}$$

$$zvalue(i,j) = zvalue + weight(m) * zk(m)$$

```
Next
```

2.5 Summary and Future Work

During the extension of year 2, several large software tasks were undertaken:

- All interfaces have been changed from menu-drive into a work-flow style, which separates the defining functions into eight stages. Screenshots of major steps will be presented in a normal sequence at the end of this section.

1. **Start** - Project Manager allows users to create a new Fuzzy Expert System, modify a defined Fuzzy Expert System or delete an existing project. (Figs. 2.50 and 2.51).
2. **Characterize** - Default Data Manager allows users to input, modify, update and delete default data of variables (Figs 2.52–2.54 and 2.34–2.46)
3. **Variables** - Variable Manager allows users to define new variables, modify existing variables and delete variables (Fig. 2.55).
4. **Fuzzy Sets** - Fuzzy Sets Manager allows users to define, modify and delete fuzzy sets and membership shapes of Fuzzy Variables (Figs. 2.56 and 2.57)
5. **Interfaces** - Interface Manager allows users define, modify and delete the interfaces of Fuzzy Expert System in the interface style of the original FEE tool software (Figs. 2.58–2.60)
6. **Rules** - Rules Manager allows users to define, modify and delete the rules related to defined variables and flags (Figs. 2.61 and 2.62).
7. **Parameters** - Weights Manager allows users to adjust the weight value of each variable for inference engine of Fuzzy Expert System (Fig. 2.63).

- Each stage is divided into a few simple steps that users can easily understand and follow. "Previous" and "Next" buttons on most steps allow users to navigate to previous step or navigate to a later process from the current step. Each step contains a title, a short instruction and one/more components (Radio buttons, text fields, combo box, etc.) which allow users to input, view, modify data or select options, or data.

- The entire software package has been re-implemented in JAVA 1.60 with a noted increase in software performance.

- CFS source codes have grown to 1,545,588 bytes (1.5MB) with a total of 51,165 lines since the end of the first project year (1,123,492 bytes with a total of 35,992

lines). This corresponds to a 40% growth in software, primarily in wizards and interfaces.

Some software issues remain to be determined through the use of beta testers. A summary of these issues follows:

- Because the number of most data items is not fixed in the CFS, it is hard to decide how to determine when a process can be checked as complete on the sidebar with regard to present project state. For example, the number of variables, rules, questions, and fuzzy sets are all variable. We are considering a new way to display the state of the system to the user, or the answer may become obvious once users are testing the system under a broader set of circumstances.

- Due to interface improvements with JAVA 1.60, which allows users to have multiple open windows or stages, some data needs to be simultaneously shared by more than one application. This requires real time data update (communication) among processes (stages) and may require that the data management software needs to be improved.

Generally speaking, the CFS is in a stage of development which requires beta user input for improving, integrating and debugging software and to accomplish the desired level of wizard functionality and user-friendliness.

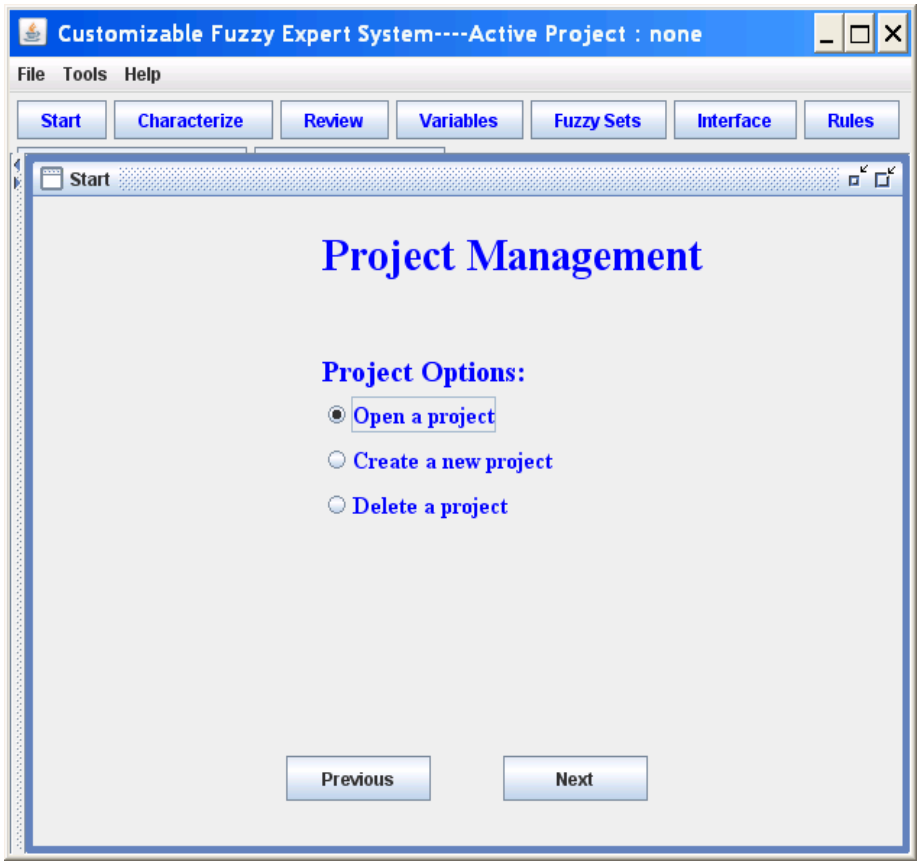


Fig. 2.50. Application screen with Project Management application running.

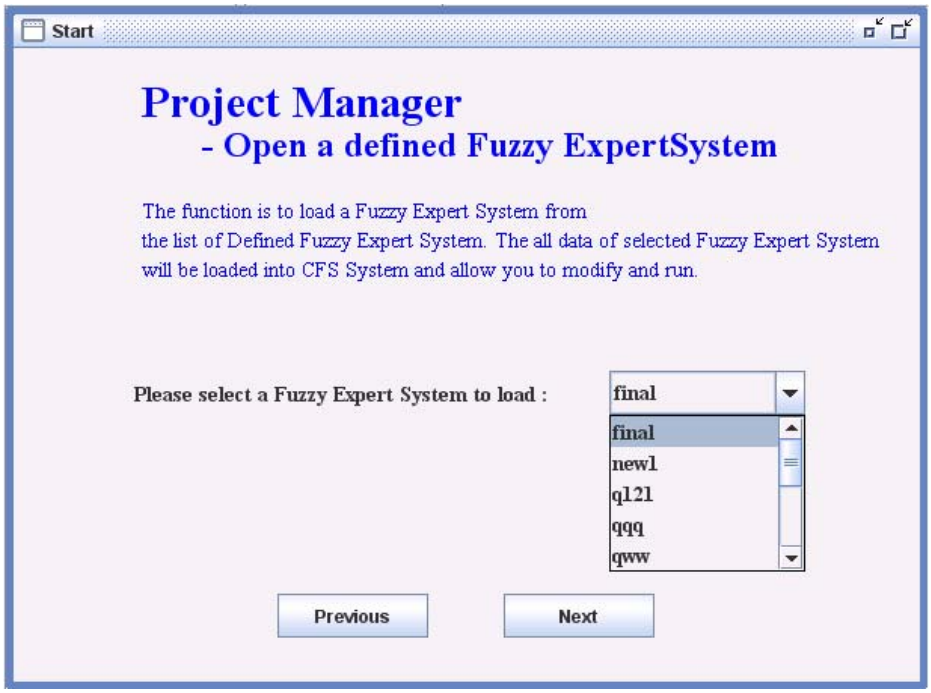


Fig. 2.51. Sample screen for loading an existing project.

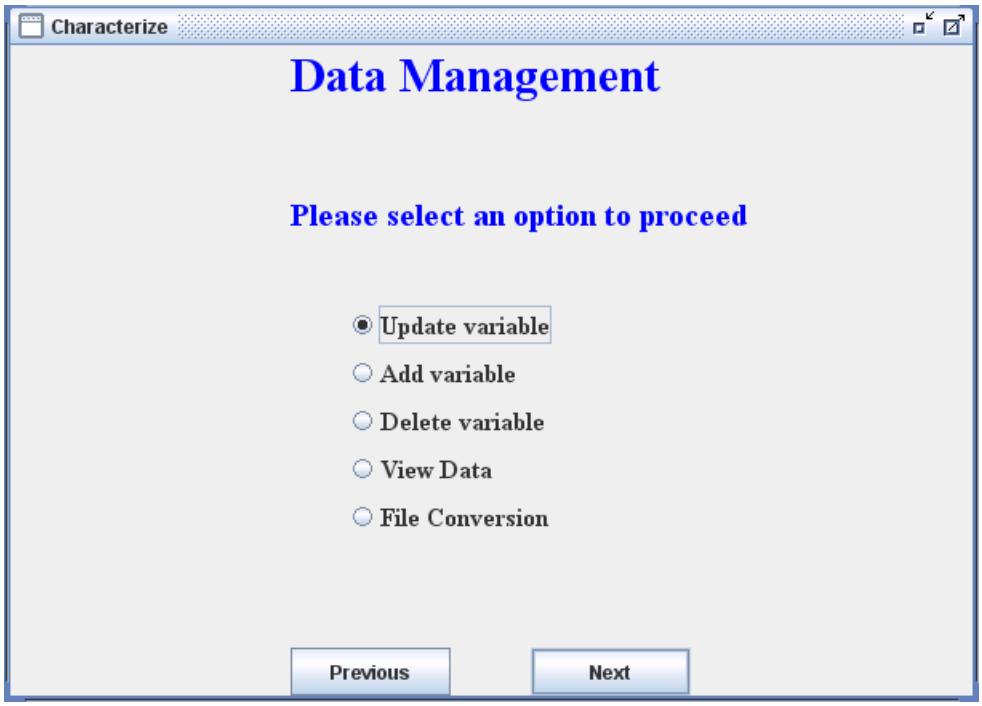


Fig. 2.52. Data Management application options.

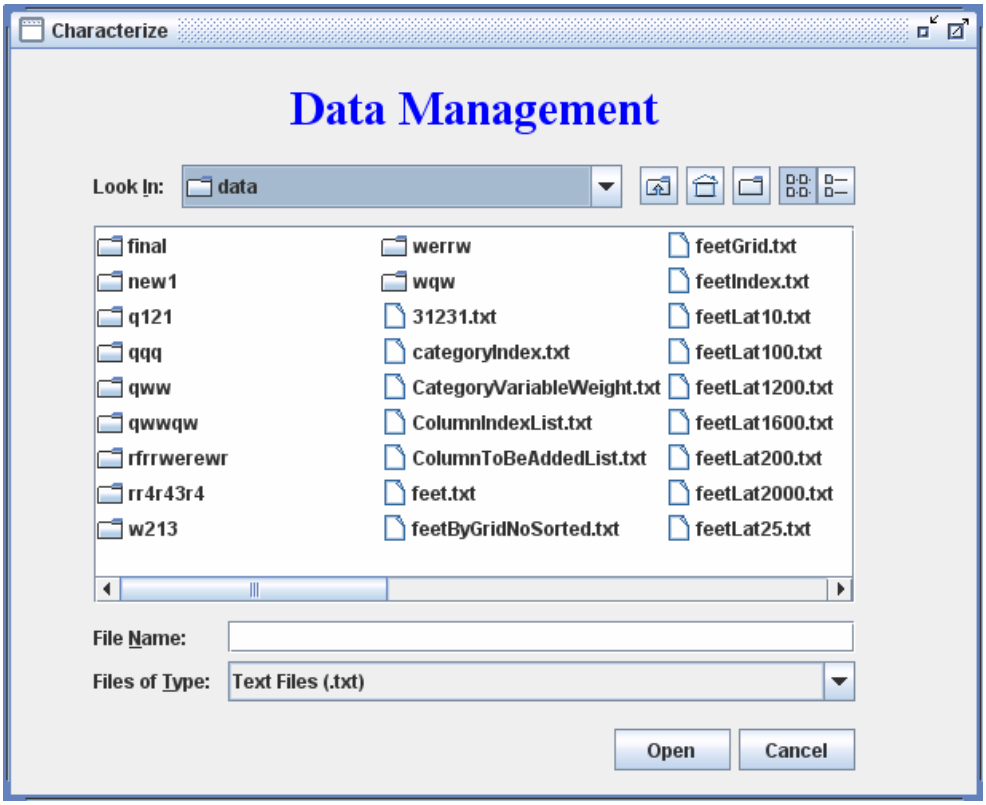
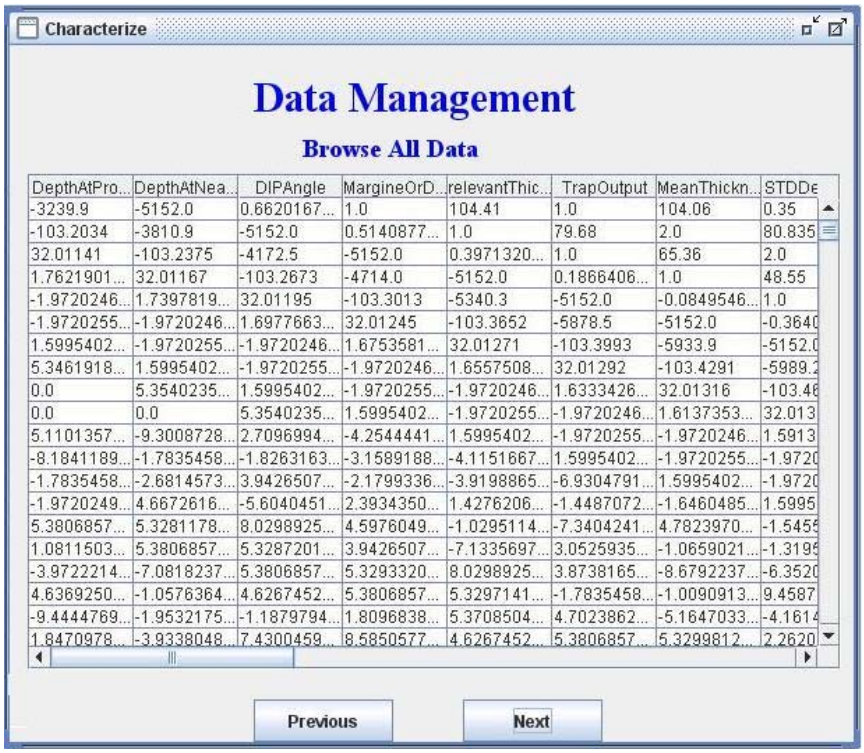


Fig. 2.53. Select file screen.



DepthAtPro...	DepthAtNea...	DIPAngle	MargineOrD...	relevantThic...	TrapOutput	MeanThickn...	STDDe
-3239.9	-5152.0	0.6620167...	1.0	104.41	1.0	104.06	0.35
-103.2034	-3810.9	-5152.0	0.5140877...	1.0	79.68	2.0	80.835
32.01141	-103.2375	-4172.5	-5152.0	0.3971320...	1.0	65.36	2.0
1.7621901...	32.01167	-103.2673	-4714.0	-5152.0	0.1866406...	1.0	48.55
-1.9720246...	1.7397819...	32.01195	-103.3013	-5340.3	-5152.0	-0.0849546...	1.0
-1.9720255...	-1.9720246...	1.6977663...	32.01245	-103.3652	-5878.5	-5152.0	-0.3640
1.5995402...	-1.9720255...	-1.9720246...	1.6753581...	32.01271	-103.3993	-5933.9	-5152.0
5.3461918...	1.5995402...	-1.9720255...	-1.9720246...	1.6557508...	32.01292	-103.4291	-5989.2
0.0	5.3540235...	1.5995402...	-1.9720255...	-1.9720246...	1.6333426...	32.01316	-103.46
0.0	0.0	5.3540235...	1.5995402...	-1.9720255...	-1.9720246...	1.6137353...	32.013
5.1101357...	-9.3008728...	2.7096994...	-4.2544441...	1.5995402...	-1.9720255...	-1.9720246...	1.5913
-8.1841189...	-1.7835458...	-1.8263163...	-3.1589188...	-4.1151667...	1.5995402...	-1.9720255...	-1.9720
-1.7835458...	-2.6814573...	3.9426507...	-2.1799336...	-3.9198865...	-6.9304791...	1.5995402...	-1.9720
-1.9720249...	4.6672616...	-5.6040451...	2.3934350...	1.4276206...	-1.4487072...	-1.6460485...	1.5995
5.3806857...	5.3281178...	8.0298925...	4.5976049...	-1.0295114...	-7.3404241...	4.7823970...	-1.5455
1.0811503...	5.3806857...	5.3287201...	3.9426507...	-7.1335697...	3.0525935...	-1.0659021...	-1.3195
-3.9722214...	-7.0818237...	5.3806857...	5.3293320...	8.0298925...	3.8738165...	-8.6792237...	-6.3520
4.6369250...	-1.0576364...	4.6267452...	5.3806857...	5.3297141...	-1.7835458...	-1.0090913...	9.4587
-9.4444769...	-1.9532175...	-1.1879794...	1.8096838...	5.3708504...	4.7023862...	-5.1647033...	-4.1614
1.8470978...	-3.9338048...	7.4300459...	8.5850577...	4.6267452...	5.3806857...	5.3299812...	2.2620

Fig. 2.54. Browse data application.

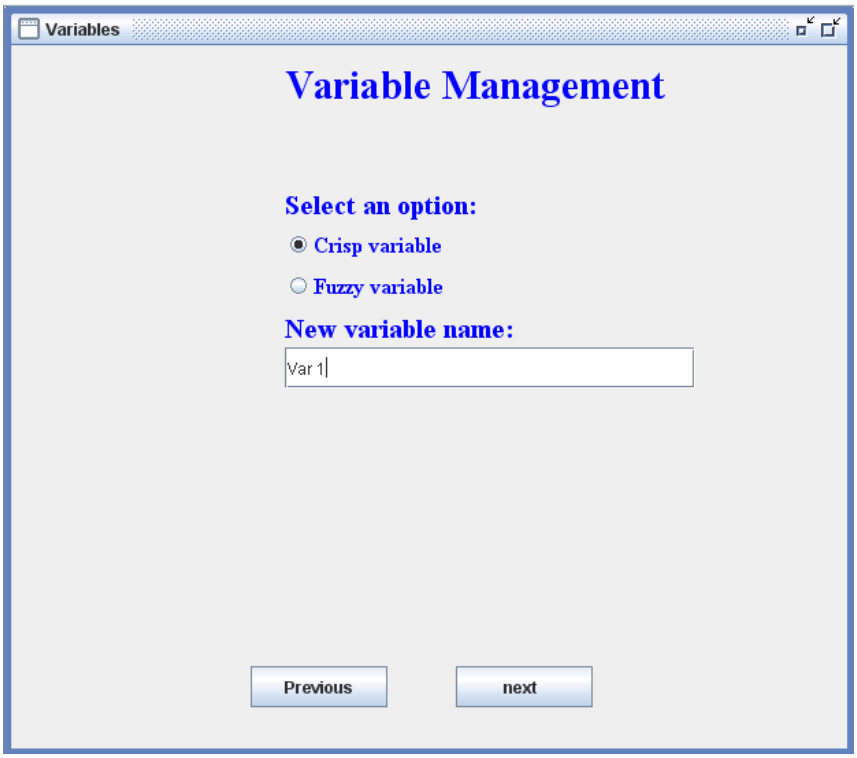


Fig. 2.55. Application screen for managing fuzzy variables.

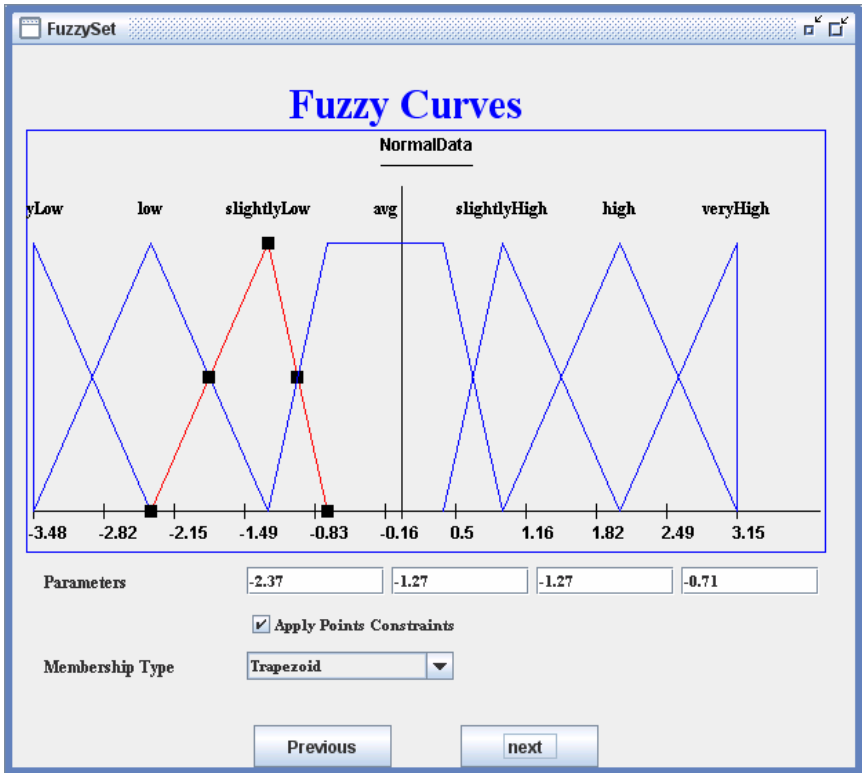


Fig. 2.56. Displayed fuzzy curves for a defined variable.

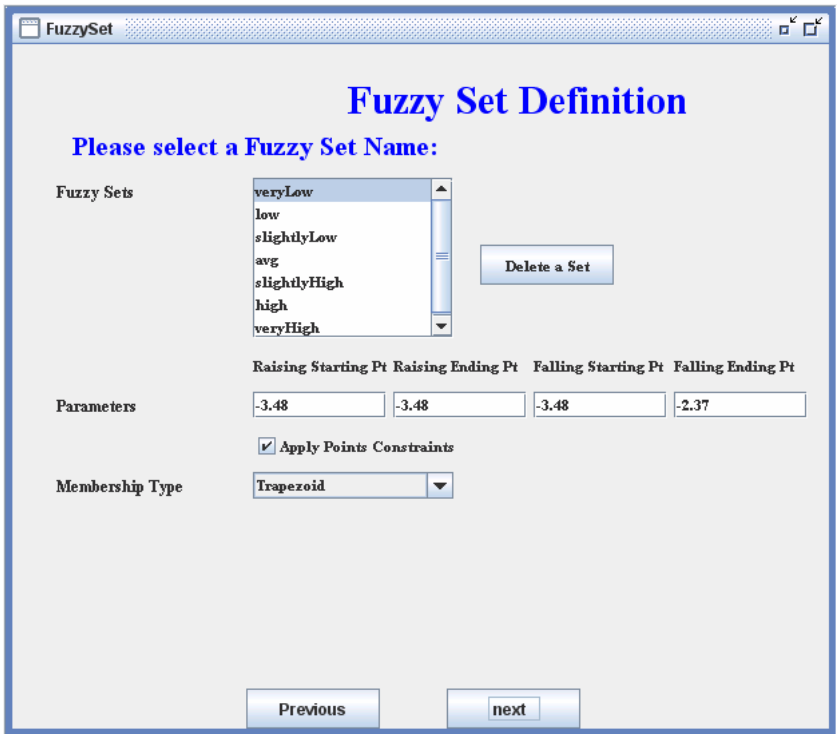


Fig. 2.57. Set definition screen for fuzzy variables.

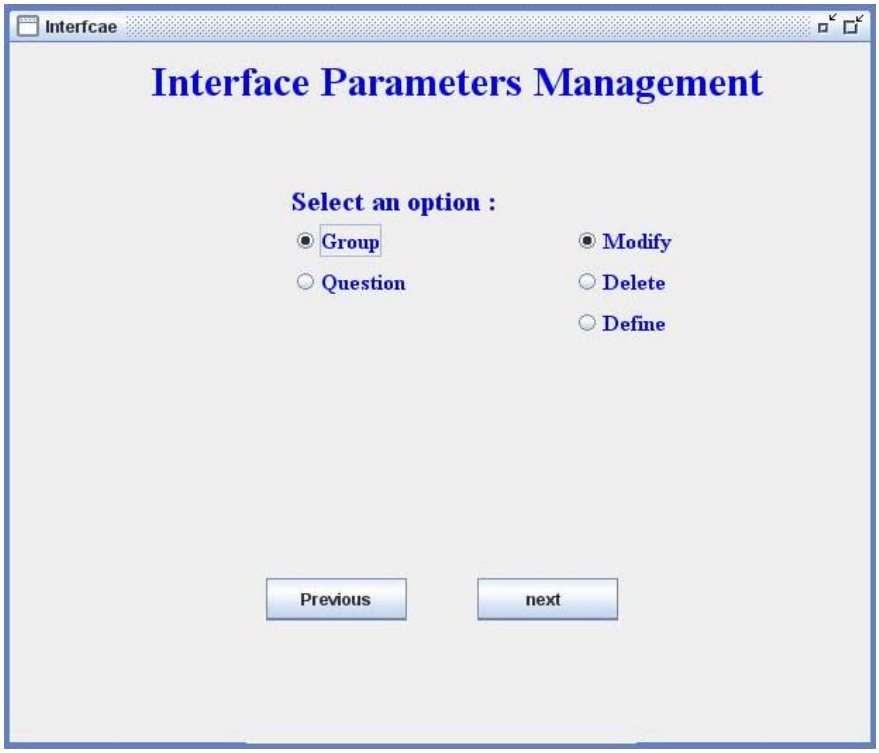


Fig.2.58. Interface management screen.

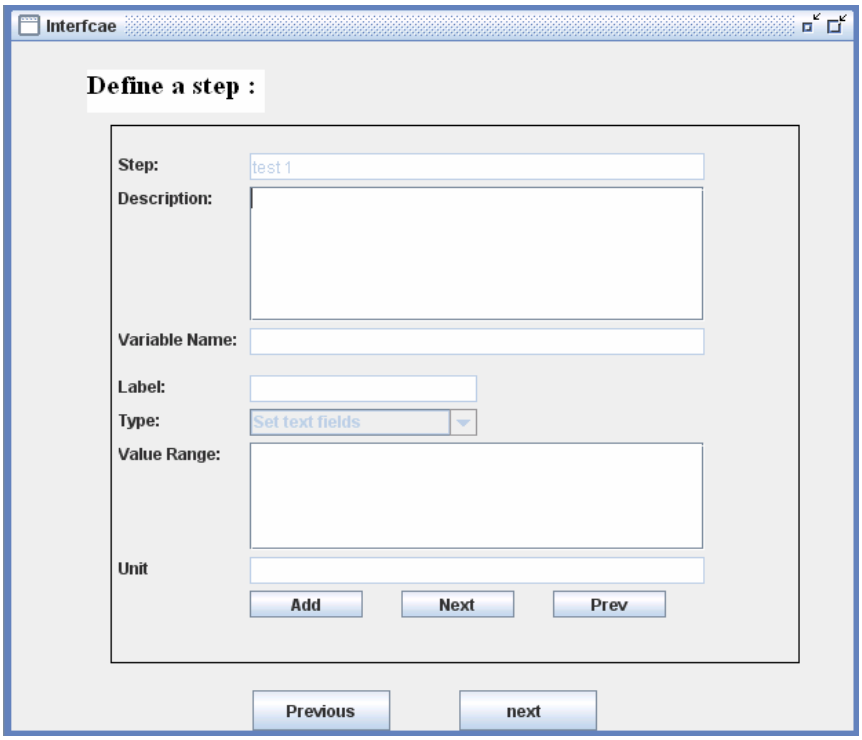


Fig. 2.59. In this application, text of questions can be entered.

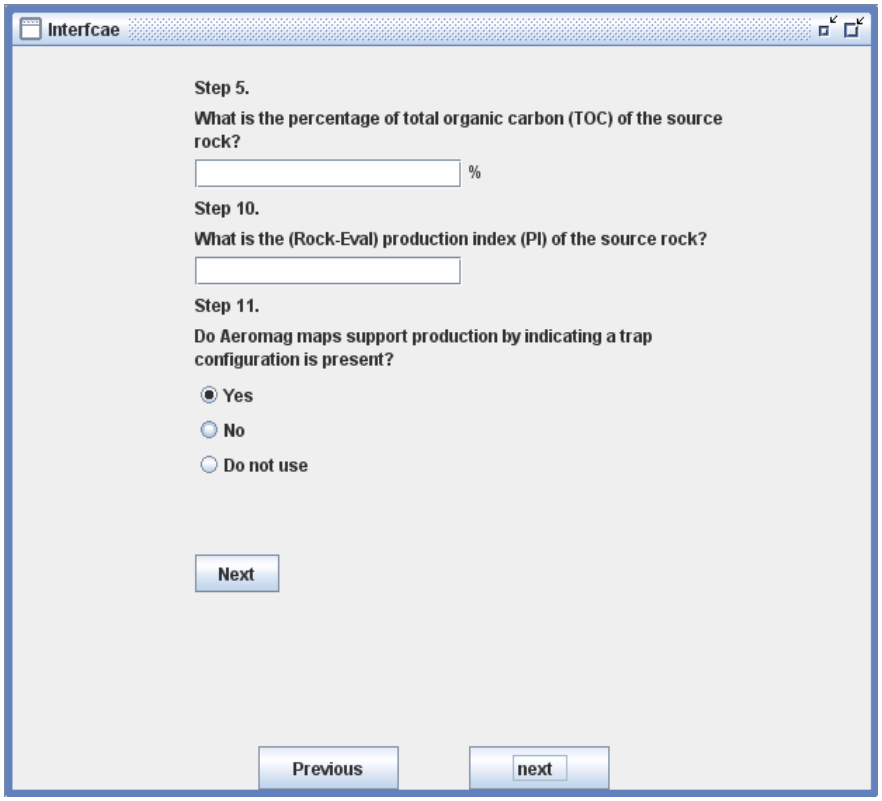
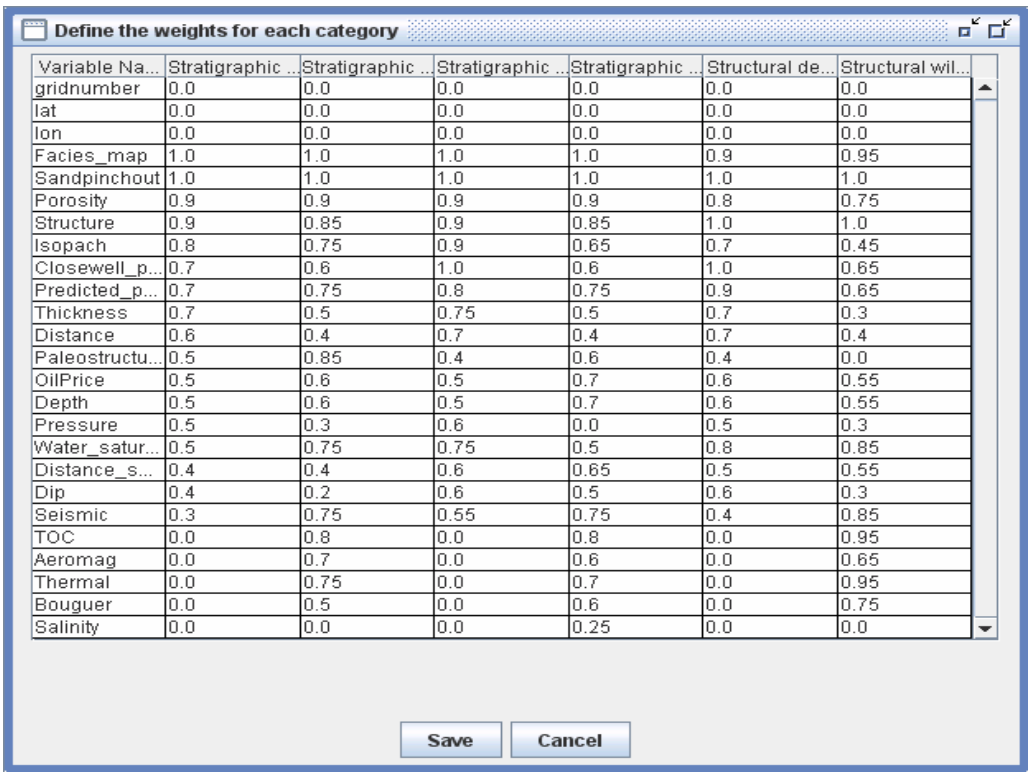


Fig. 2.60. Sample output for entered questions.

Fig. 2.61. Inference parameter setup application introduction screen.

Fig. 2.62. Boolean logic can be used to define inference parameters.



Define the weights for each category

Variable Na...	Stratigraphic ...	Stratigraphic ...	Stratigraphic ...	Stratigraphic ...	Structural de...	Structural wil...
gridnumber	0.0	0.0	0.0	0.0	0.0	0.0
lat	0.0	0.0	0.0	0.0	0.0	0.0
lon	0.0	0.0	0.0	0.0	0.0	0.0
Facies_map	1.0	1.0	1.0	1.0	0.9	0.95
Sandpinchout	1.0	1.0	1.0	1.0	1.0	1.0
Porosity	0.9	0.9	0.9	0.9	0.8	0.75
Structure	0.9	0.85	0.9	0.85	1.0	1.0
Isopach	0.8	0.75	0.9	0.65	0.7	0.45
Closewell_p...	0.7	0.6	1.0	0.6	1.0	0.65
Predicted_p...	0.7	0.75	0.8	0.75	0.9	0.65
Thickness	0.7	0.5	0.75	0.5	0.7	0.3
Distance	0.6	0.4	0.7	0.4	0.7	0.4
Paleostructu...	0.5	0.85	0.4	0.6	0.4	0.0
OilPrice	0.5	0.6	0.5	0.7	0.6	0.55
Depth	0.5	0.6	0.5	0.7	0.6	0.55
Pressure	0.5	0.3	0.6	0.0	0.5	0.3
Water_satur...	0.5	0.75	0.75	0.5	0.8	0.85
Distance_s...	0.4	0.4	0.6	0.65	0.5	0.55
Dip	0.4	0.2	0.6	0.5	0.6	0.3
Seismic	0.3	0.75	0.55	0.75	0.4	0.85
TOC	0.0	0.8	0.0	0.8	0.0	0.95
Aeromag	0.0	0.7	0.0	0.6	0.0	0.65
Thermal	0.0	0.75	0.0	0.7	0.0	0.95
Bouguer	0.0	0.5	0.0	0.6	0.0	0.75
Salinity	0.0	0.0	0.0	0.25	0.0	0.0

Save Cancel

Fig. 2.63. Defined weights of variables for a sample projects.

3. Geologic Data Acquisition and Analysis

Introduction

During the period from January 2006 through December 2006, geologic data acquisition and analysis continued and was completed on the Upper Pennsylvanian and Lower Permian carbonate reservoirs within the project areas in the Permian Basin of southeastern New Mexico (Figs. 3.1–3.3). During 2006, work on the Bough intrashelf project area was completed. Geologic data was acquired, mapped and analyzed on a-second project area, the shelf-margin project area that includes the Upper Pennsylvanian Dagger Draw oil reservoirs.

The second project area was added to diversify the expert system's approach to prospecting for oil in carbonate stratigraphic traps. Although reservoirs in both areas are formed primarily by phylloid algal mounds that have seen multiple episodes of diagenesis, the reservoirs within the Bough intrashelf area grew on bathymetrically high paleostructures and the reservoirs within the shelf-margin area grew primarily as bioherms on a constructional shelf-margin; their geographic location does not appear to be related to paleostructures.

Both of the project areas have significant production. The 58 Permo-Pennsylvanian reservoirs that are at least partially present within the Bough intrashelf project area have produced a combined total of 329 million bbls oil (MMBO; Table 1). The 24 Upper Pennsylvanian reservoirs within the shelf-margin project area have produced a combined total of 83 million bbls oil (Table 3). The reservoirs from the intrashelf project area are located within the Permo-Pennsylvanian age Bough zone. The Bough has been subdivided informally by industry geologists into four widely recognized members (descending; Bough A member, Bough B member, Bough C member, Bough D member). The Bough zones have long been recognized as straddling the boundary between the Upper Pennsylvanian and the Lower Permian (see Cys, 1986) with the Bough A, B, and C placed within the lowermost part of the Permian by most workers and the Bough D placed in the uppermost part of the Pennsylvanian. The most recent work, based on international conodont zonation (see Wahlman, 2001), raises the boundary between the Per-

mian and the Pennsylvanian and would apparently place the entire Bough section in the uppermost Pennsylvanian. The Bough reservoirs are productive from phylloid algal mounds and associated flanking grainstones (Cys and Mazzullo, 1985; Malek-Aslani, 1985; Cys, 1986; Broadhead, 1999a, 1999b; Wahlman, 2001). The reservoirs from the shelf-margin project area are productive primarily from Canyon and Cisco strata within the Upper Pennsylvanian and therefore are mostly somewhat older than the reservoirs within the intrashelf project area. However, they are also productive from phylloid algal mounds and associated flanking deposits (see Cox et al., 1998) and therefore should have similar reservoir characteristics.

Our work for the first year indicated that the Bough C member (Figs. 3.1, 3.4) dominates production within the Bough intrashelf project area. During the first project year (calendar year 2005), basic geologic data pertaining to structure, paleostructure, and the stratigraphic environments relevant to oil and natural gas production in the intrashelf project area were obtained from well records and by analyzing and correlating electric logs from 300 wells drilled within the 1300 mi^2 project area. It is estimated that approximately one-half of the geologic data required for the project were obtained during this first year of the project.

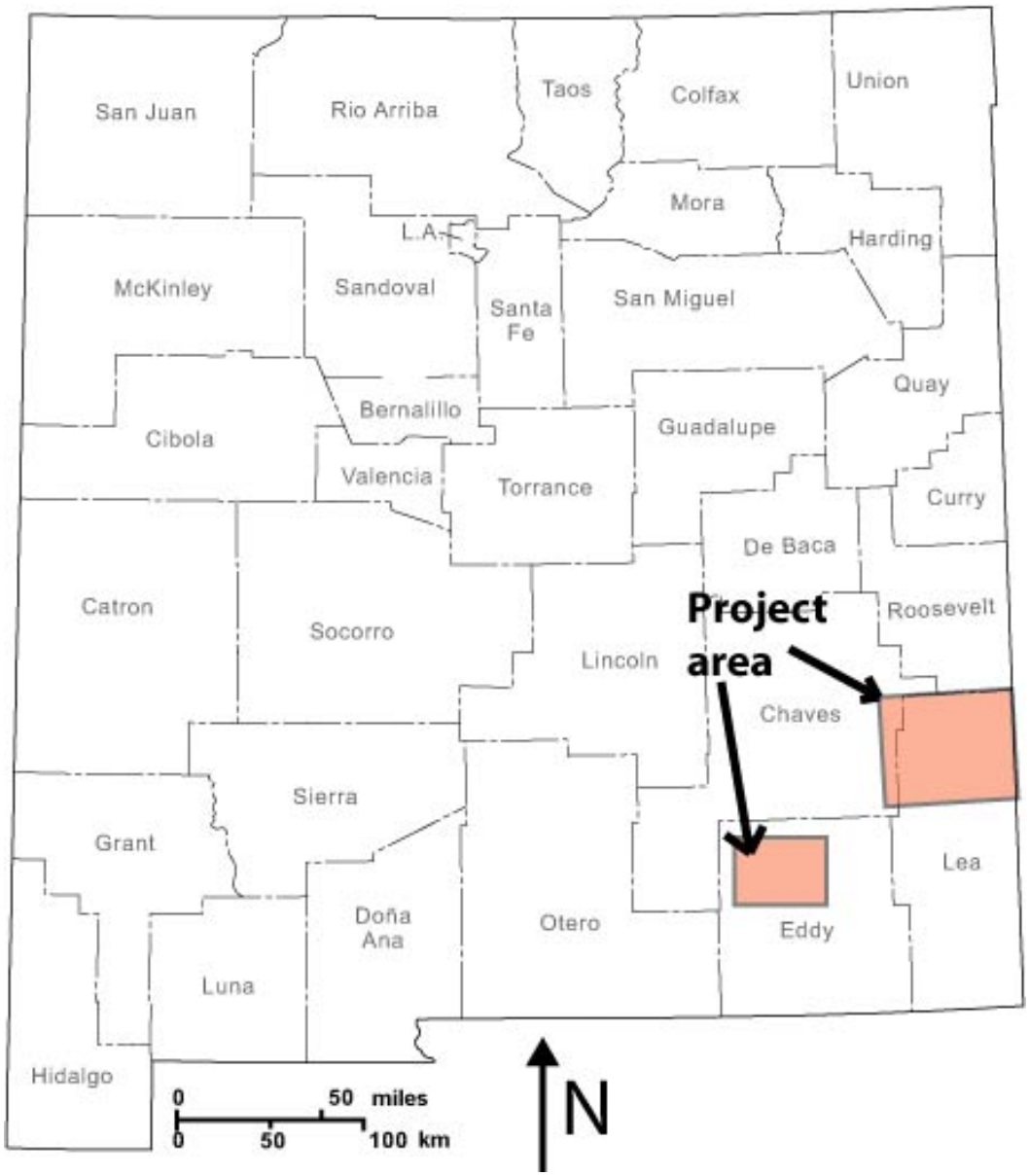


Fig. 3.1. Location of project areas in New Mexico.

Table 3.1. Permo-Pennsylvanian Carbonate Reservoirs Productive within the Bough Intrashelf Project Area (See Fig. 3.5 for Locations of Reservoirs)

Reservoir name	Productive unit	2003 oil production (bbls oil)	Cumulative oil production 2003 (million bbls)
Allison	Upper Pennsylvanian	30645	23.92
Alston Ranch	Upper Pennsylvanian	0	0.14
Anderson Ranch North	Wolfcamp	27785	6.74
Austin	Permo Penn	101	0.06
Austin Northwest	Permo Penn	0	0.61
Austin Southwest	Wolfcamp	0	0.03
Bagley	Pennsylvanian	998	4.34
Bagley East	Upper Pennsylvanian	2195	0.25
Bagley East	Wolfcamp	0	0.02
Bagley North	Permo Penn	138830	53.40
Bar-U	Upper Pennsylvanian	25667	1.46
Baum	Upper Pennsylvanian	21159	15.30
Baum North	Wolfcamp	1221	0.03
Baum South	Wolfcamp	98	0.04
Bluitt	Wolfcamp	5577	0.58
Bough	Permo Penn	0	6.33
Bronco Southwest	Wolfcamp	11203	0.24
Caprock North	Wolfcamp	1775	0.09
Caprock	Wolfcamp	0	0.00
Caprock East	Wolfcamp	5213	0.43
Caudill	Permo Upper Penn	9198	2.01
Caudill Northeast	Wolfcamp	0	0.32
Cerca	Upper Pennsylvanian	0	1.98
Cindy	Wolfcamp	0	0.02
Crossroads	Upper Pennsylvanian	0	2.17
Denton	Wolfcamp	219118	42.44
Denton East	Wolfcamp	0	0.02
Echols	Wolfcamp	0	0.01
Eight Mile Draw	Permo Upper Penn	0	0.09
Flying M South	Bough	0	1.21
Flying M	Pennsylvanian	4498	0.07
Feather East	Upper Pennsylvanian	1390	0.06
Four lakes	Upper Pennsylvanian	8520	2.77
Gladiola	Wolfcamp	13565	4.19
Gladiola South	Wolfcamp	1284	0.17
High Plains	Permo Upper Penn	7988	1.06
Hightower East	Upper Pennsylvanian	7824	1.08
Hightower	Permo Penn	4109	0.71
Hightower	Wolfcamp	0	0.00
Inbe	Permo Upper Penn	8730	16.47
Jenkins	Cisco	0	2.10
King	Wolfcamp	57482	1.50
King	Pennsylvanian	0	0.09
King West	Pennsylvanian	0	0.01
Lane	Wolfcamp	0	1.03
Llano	Upper Pennsylvanian	0	0.00

Reservoir name	Productive unit	2003 oil production (bbls oil)	Cumulative oil production 2003 (million bbls)
McDonald	Upper Pennsylvanian	2091	0.16
Mescalero North	Upper Pennsylvanian		0.00
Mescalero North	Cisco	50446	0.61
Mescalero	Permo-Pennsylvanian	2115	0.97
Mescalero West	Pennsylvanian	0	0.00
Mescalero Northeast	Cisco	657	0.02
Mescalero North	Wolfcamp	0	0.06
Milnesand	Pennsylvanian	0	1.00
Milnesand East	Pennsylvanian	0	0.00
Milnesand West	Pennsylvanian	21799	0.36
Moore	Permo Penn	270	0.23
Morton	Wolfcamp	6784	2.63
Morton East	Wolfcamp	16858	1.84
Morton North	Permo Upper Penn	0	0.86
Nonombre	Upper Pennsylvanian	0	1.08
Nonombre North	Upper Pennsylvanian	613	0.04
Pollock	Wolfcamp	0	0.23
Prairie South	Cisco	0	2.91
Prairie South	Wolfcamp	0	0.04
Ranger Lake	Upper Pennsylvanian	7396	5.11
Ranger Lake	Bough	2583	0.27
Ranger Lake East	Cisco	0	0.00
Saunders	Permo-Upper Penn	71081	39.20
Saunders East	Permo Penn	4932	2.73
Saunders South	Permo Upper Penn	2115	0.29
SRR	Upper Pennsylvanian	0	0.04
Tatum	Upper Pennsylvanian		0.20
Tatum	Wolfcamp	0	0.68
Tobac	Upper Pennsylvanian	11501	9.27
Todd	Wolfcamp	24599	1.19
Tres Papalotes	Upper Pennsylvanian	43773	2.11
Tres Papalotes West	Upper Pennsylvanian	0	1.24
Tulk	Pennsylvanian	12624	1.85
Tulk	Wolfcamp	16957	2.48
Tulk North	Wolfcamp	0	0.02
Tulk Southwest	Wolfcamp	2957	0.05
Vada	Upper Pennsylvanian	34252	53.43

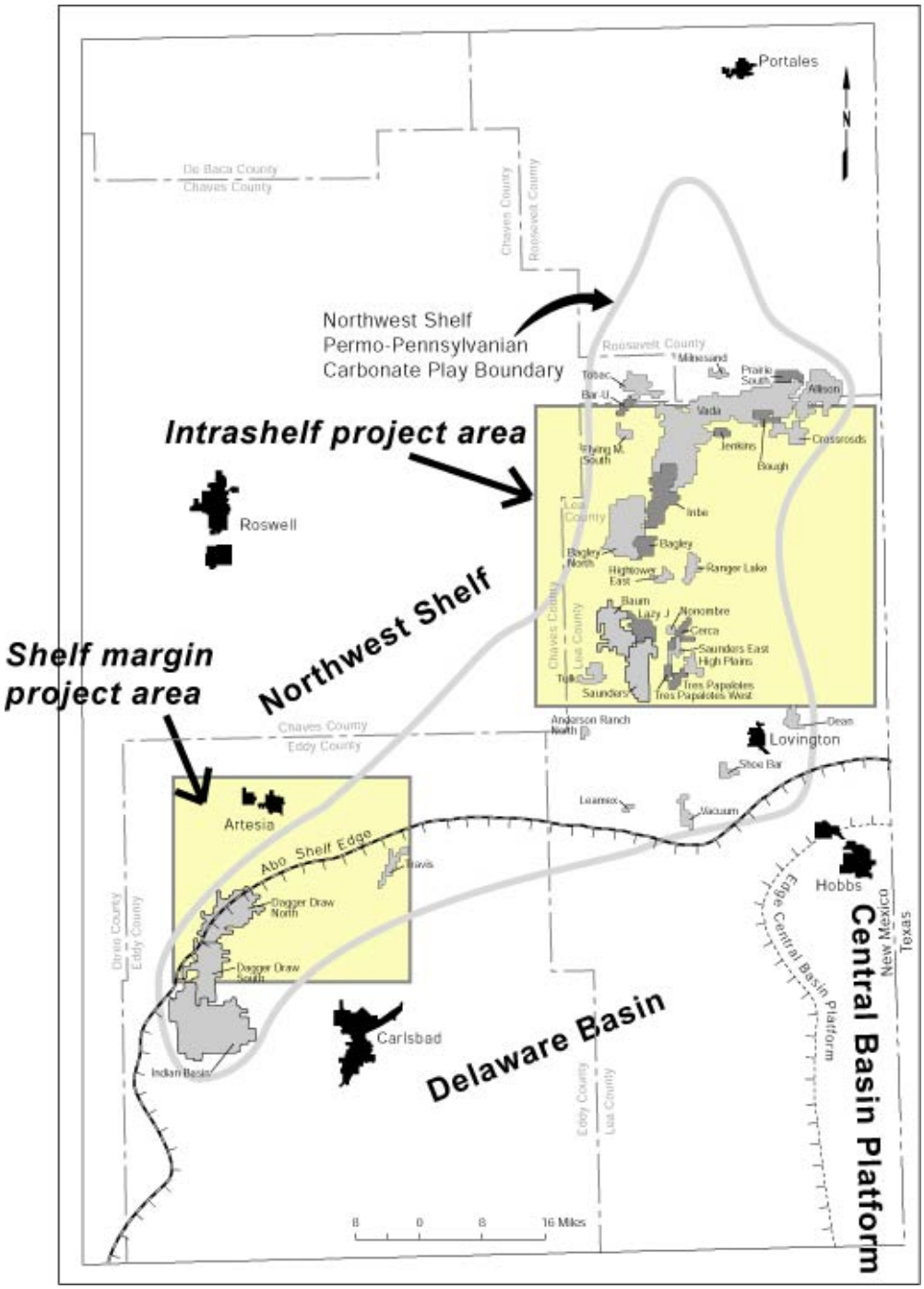


Fig. 3.2. Location of project areas in southeastern New Mexico in relationship to boundaries of the Permo-Pennsylvanian carbonate reservoir oil play and major tectonic elements. Shown in gray are oil reservoirs that have produced more than 1 million bbls oil from Upper Pennsylvanian and Lower Permian carbonate reservoirs. Modified from Broadhead and others (2004).

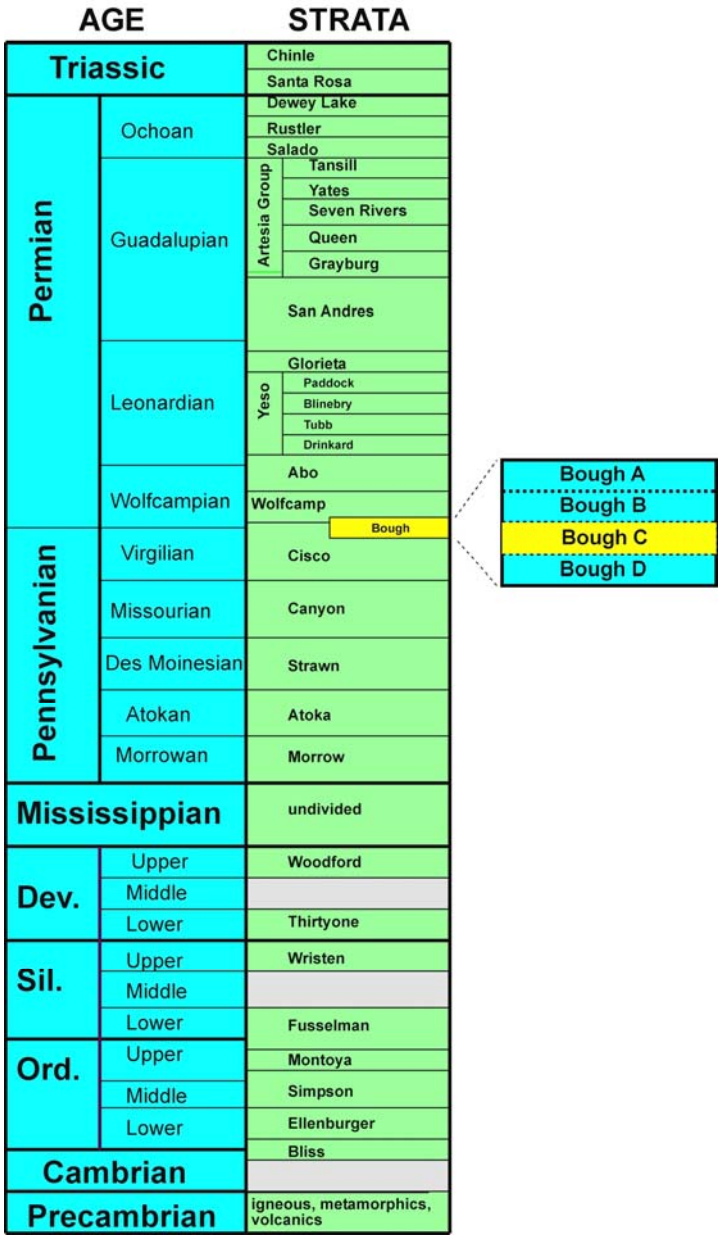


Fig. 3.3. Stratigraphic chart of the Northwest Shelf of the Permian Basin emphasizing the productive Bough carbonate reservoirs. The Bough C member (highlighted in yellow) is the major productive Bough member within the intrashelf project area. Canyon strata contain the major productive reservoirs within the shelf-margin project area with secondary production provided by reservoirs in Cisco strata.

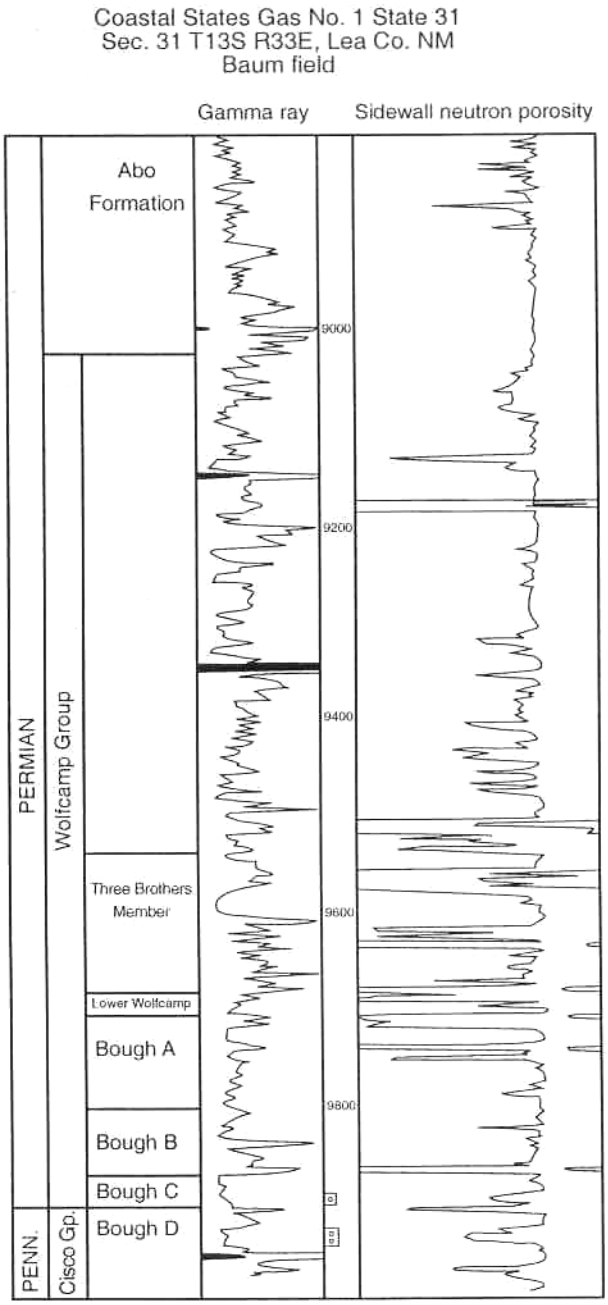


Fig. 3.4. Typical well log through the Bough members within the Bough intrashelf project area. From Broadhead (1999a).

During the second project year (calendar year 2006), additional stratigraphic data were acquired for the intrashelf project area as well as detailed production data for individual wells. Additionally, a small amount of petrographic data from reservoir cores were acquired so that relationships of diagenetic reservoir development to paleostructure could be discerned.

During the second project year (calendar year 2006) geologic data pertaining to structure, paleostructure, and the stratigraphic environments relevant to oil and natural gas production in the shelf-margin project area were obtained from well records and by analyzing and correlating electric and radioactivity logs from 312 wells within the 720 mi^2 region. Although additional data acquisition and analysis would no doubt result in improved operation of the expert system in reservoir types of the shelf-margin project area, termination of the project at the end of 2006 necessitated an early finish to the project and therefore a premature end to data acquisition and analysis. Further discussion on data acquisition, mapping and analysis is subdivided below into sub-discussions on the Bough intrashelf area and the Dagger Draw shelf-margin area.

Bough Intrashelf Project Area

Project Tasks – Geologic Data and Acquisition

Geologic data, production data, and well data for the work performed in the Bough intrashelf project area may be found in the following attached databases:

1. ***Bough greater area.xls***: contains well data and geologic data for the 300 wells used to map and analyze the entire project area (Fig. 3.5)
2. ***Bough detailed area.xls***: contains well data, geologic data and production data for the portion of the project area selected for more detailed and comprehensive work.

The following tasks were enjoined during the second project year:

1. ***Compiled cumulative production data*** on Bough (Permo-Pennsylvanian) carbonate reservoirs within the project area and mapped the data.
2. ***Stratigraphic data acquisition in wells:*** Refined correlations of productive Bough carbonate zones in 224 wells within the detailed part of the project area.
3. ***Stratigraphic data mapping and analysis*:** Made final isopach maps of productive Bough carbonate reservoir zones in the detailed part of the project area and related production to the isopach maps.
4. ***Structural data mapping and analysis*:** Constructed structure contour maps of the upper surface of the Bough A, Bough B, Bough C, and Bough D zones within the project area and related structure to production from Bough carbonate reservoirs.
5. ***Paleostructural mapping and analysis*:** During the first project year, structural data on multiple formations were used to create a paleostructure map of positive tectonic elements that were formed concurrently with deposition of Bough reservoirs. Positive paleostructural elements were related to production from Permo-Pennsylvanian carbonate reservoirs. Additional, significant data analysis was performed during 2006.
6. ***Reservoir analysis:*** Described and analyzed cores and thin sections of cores from productive and non-productive reservoir strata in the Bough zones in order to determine factors that control porosity development and to better relate reservoir locations to paleostructures within the project area.

Each of these tasks is discussed more fully below and maps produced as a result of tasks are presented.

1. Compiled production and stratigraphic data on Bough (Permo-Pennsylvanian) carbonate reservoirs within the project area. Reservoir-wide cumulative and annual oil production data and stratigraphic data were compiled for the 83 designated oil and gas reservoirs that produce from Permo-Penn carbonate reservoirs within the boundaries of the project area (Fig. 3.5; Table 3.1). Because New Mexico cumulative production figures are not always accurate after the year 1993, the annual production data after 1993 were added to the cumulative total for reser-

voirs that are either wholly or partially present within the Bough intrashelf project area. Details of the problems with published post-1993 cumulative production data are given in Broadhead et al. (2004) and Dutton et al. (2005). A database was also produced for 224 wells within a more detailed subsection of the project area (Figs. 3.6, 3.7) that includes cumulative oil, gas and water production data for each well (again calculated by adding annual production data for years after 1993 to the published 1993 cumulative numbers), depth of perforated productive reservoirs, a summary of pertinent completion information for nonproductive wells, which Bough zone is productive in the well, and geologic data obtained from the correlations. Contour maps of cumulative oil production per well were prepared for the entire Bough formation (Fig. 3.8) and for the Bough C member (Fig. 3.9), which is the most productive Bough member (Table 3.2).

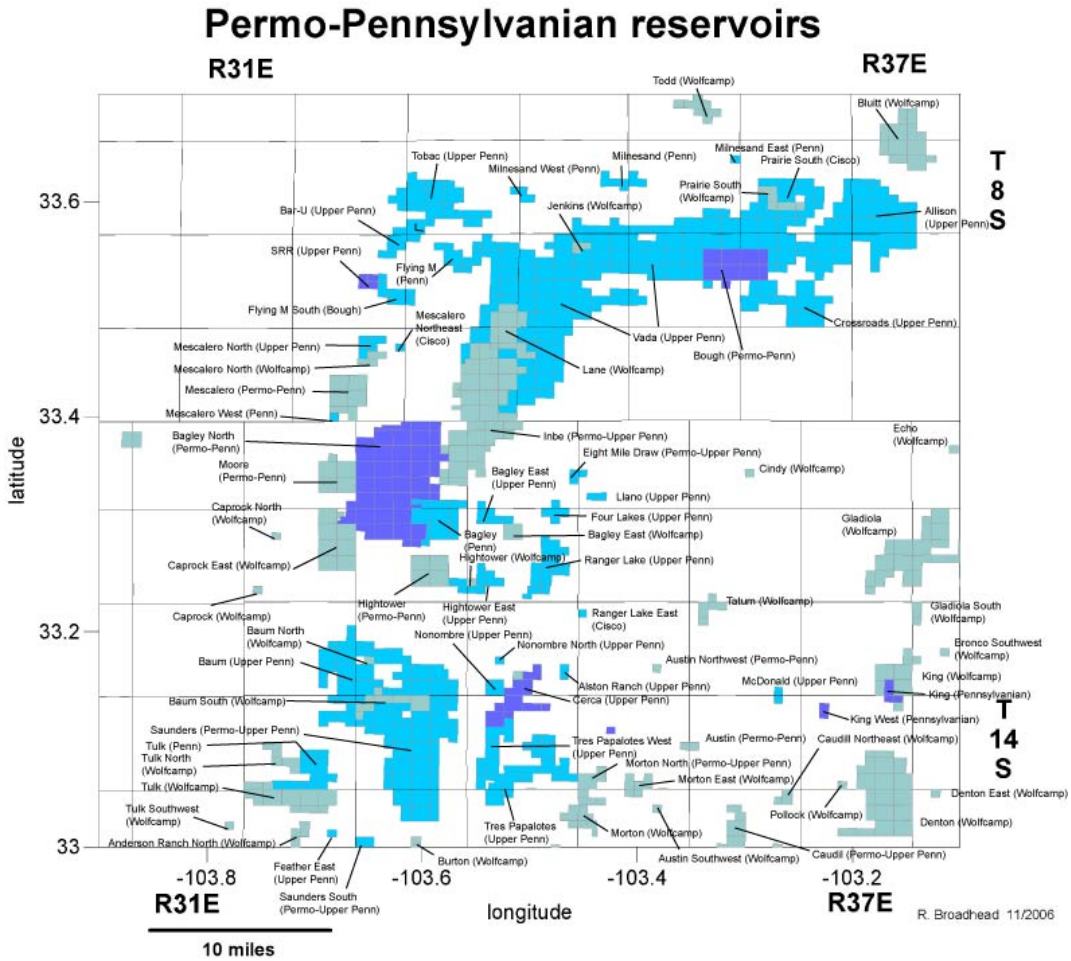


Fig. 3.5. Reservoirs productive from Upper Pennsylvanian and Lower Permian carbonates in the Bough intrashelf project area.

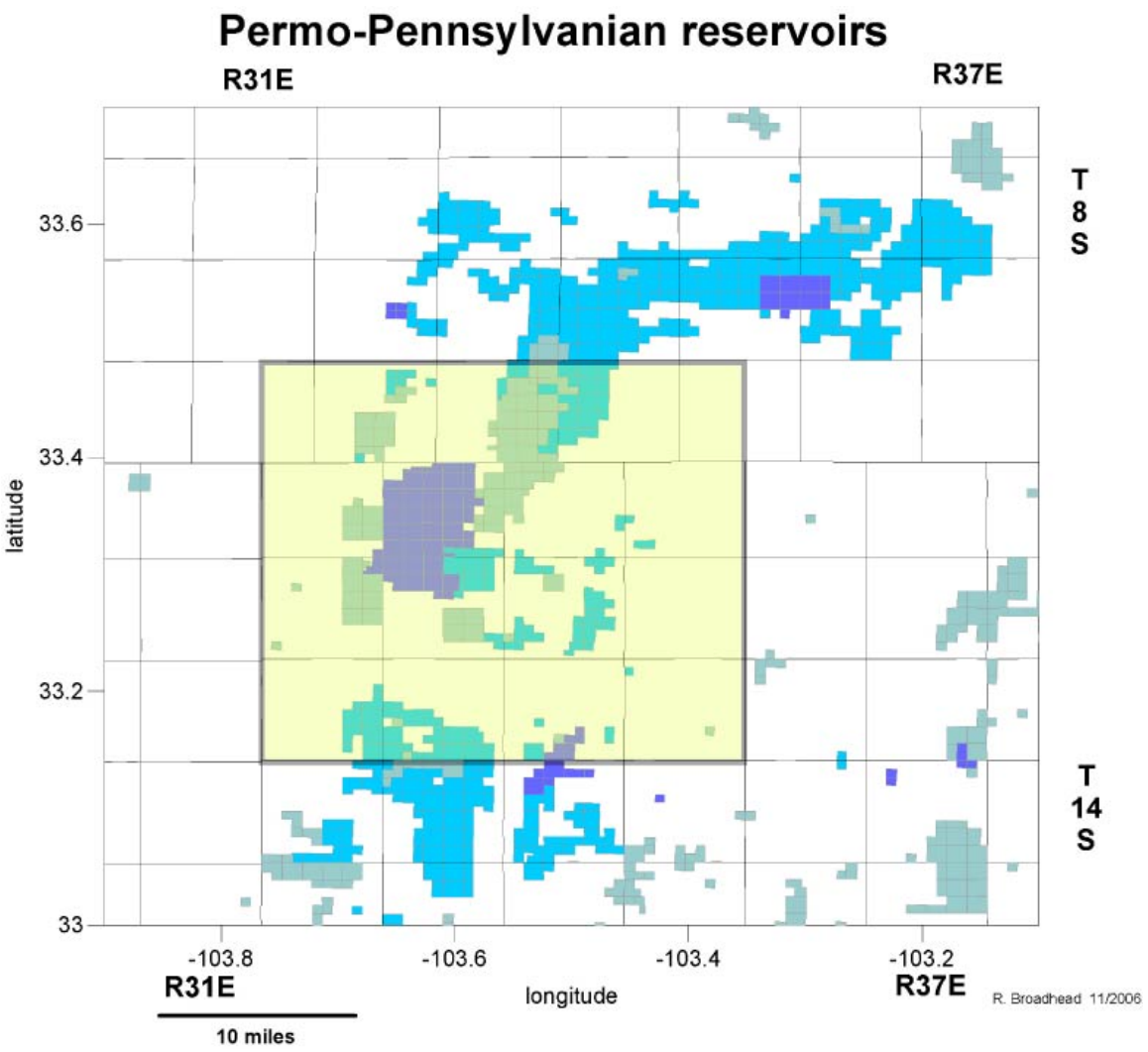


Fig. 3.6. Location of the area selected for more detailed analysis of the Bough formation. The detailed work in this area was undertaken by graduate student Jake Sharp.

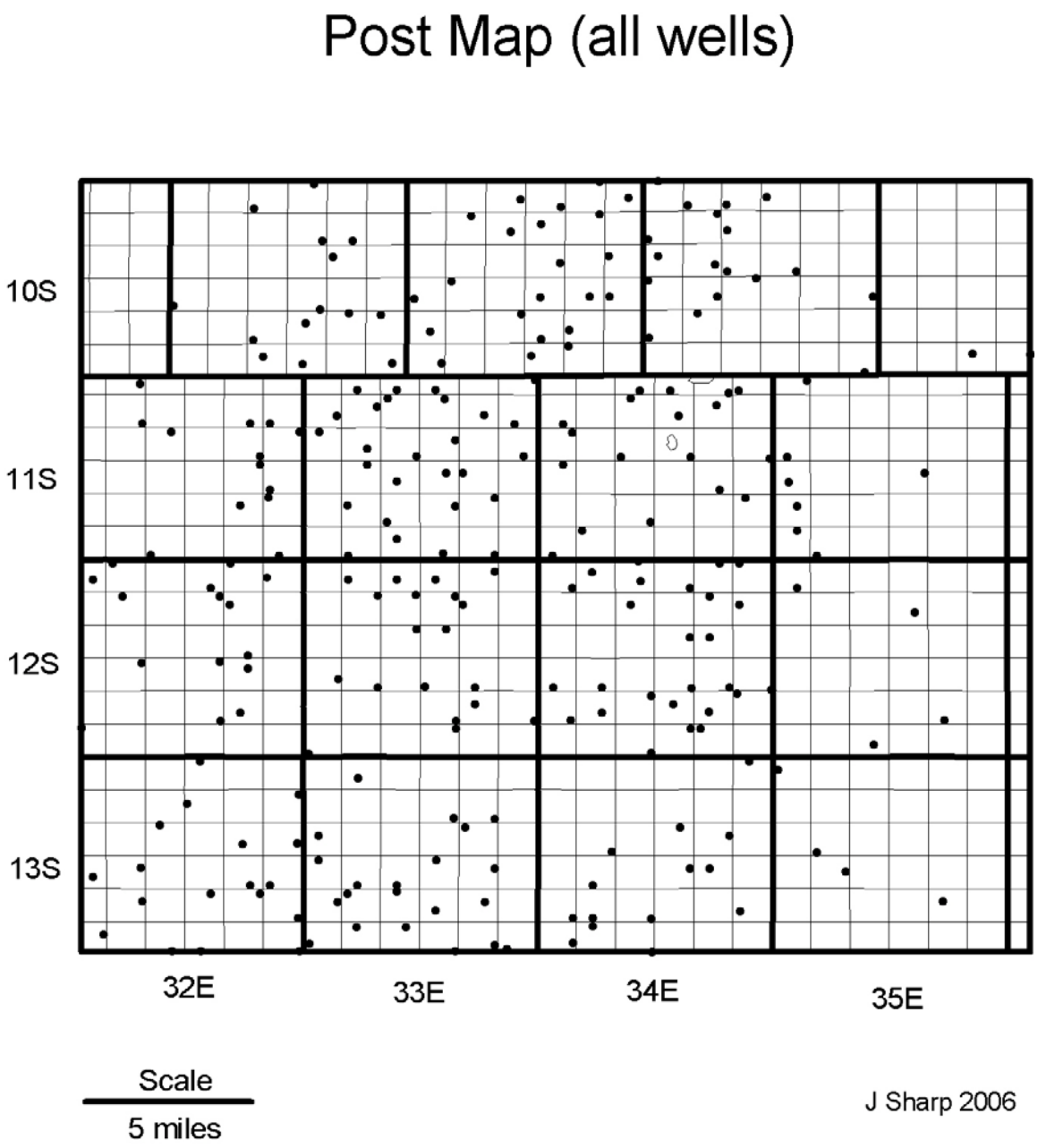


Fig. 3.7. Location of wells used for geologic and production analysis in the detailed project area.

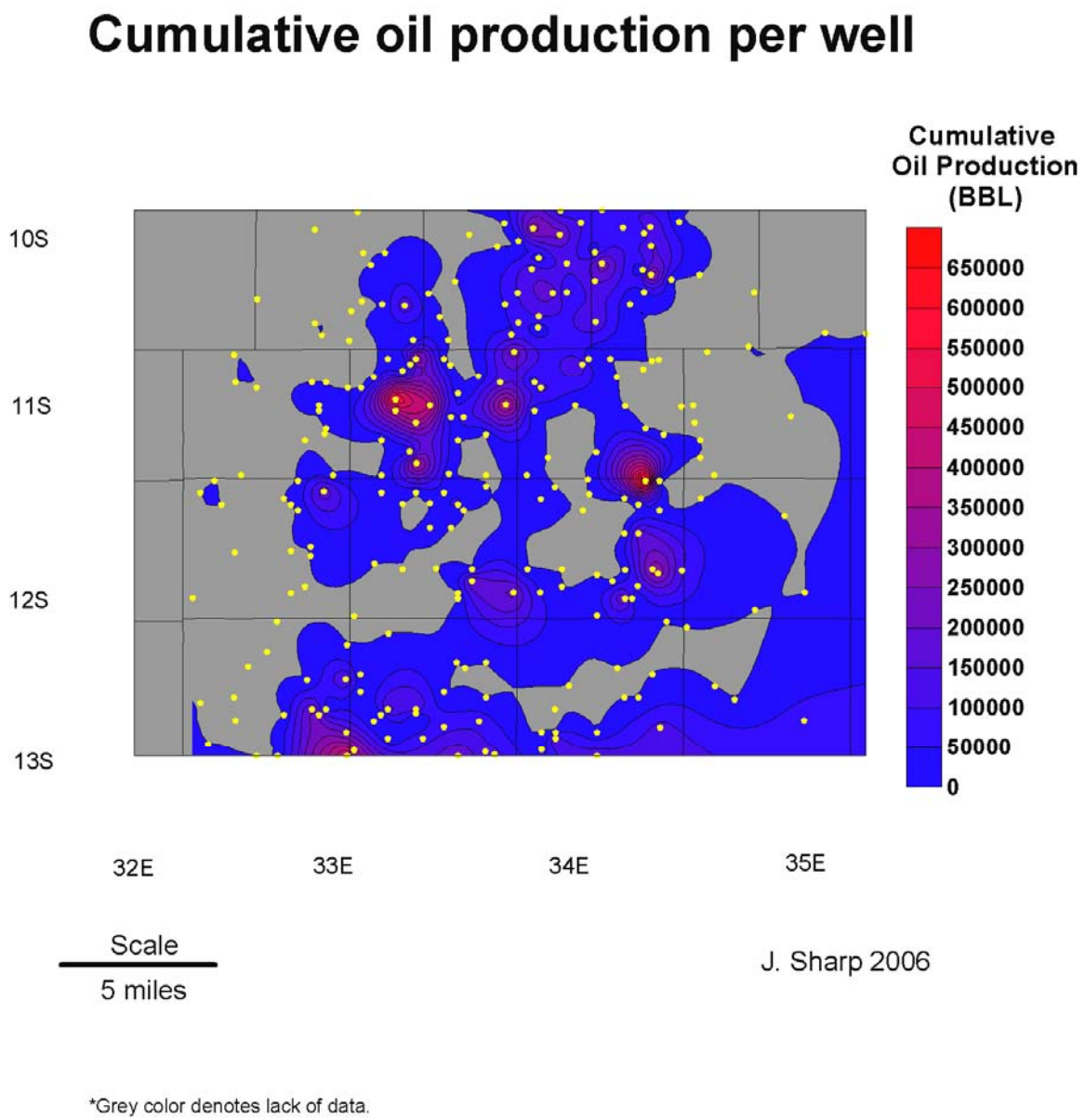


Fig. 3.8. Contour map of cumulative oil production from wells productive from the Bough A, Bough B, Bough C, and Bough D zones in the detailed project area.

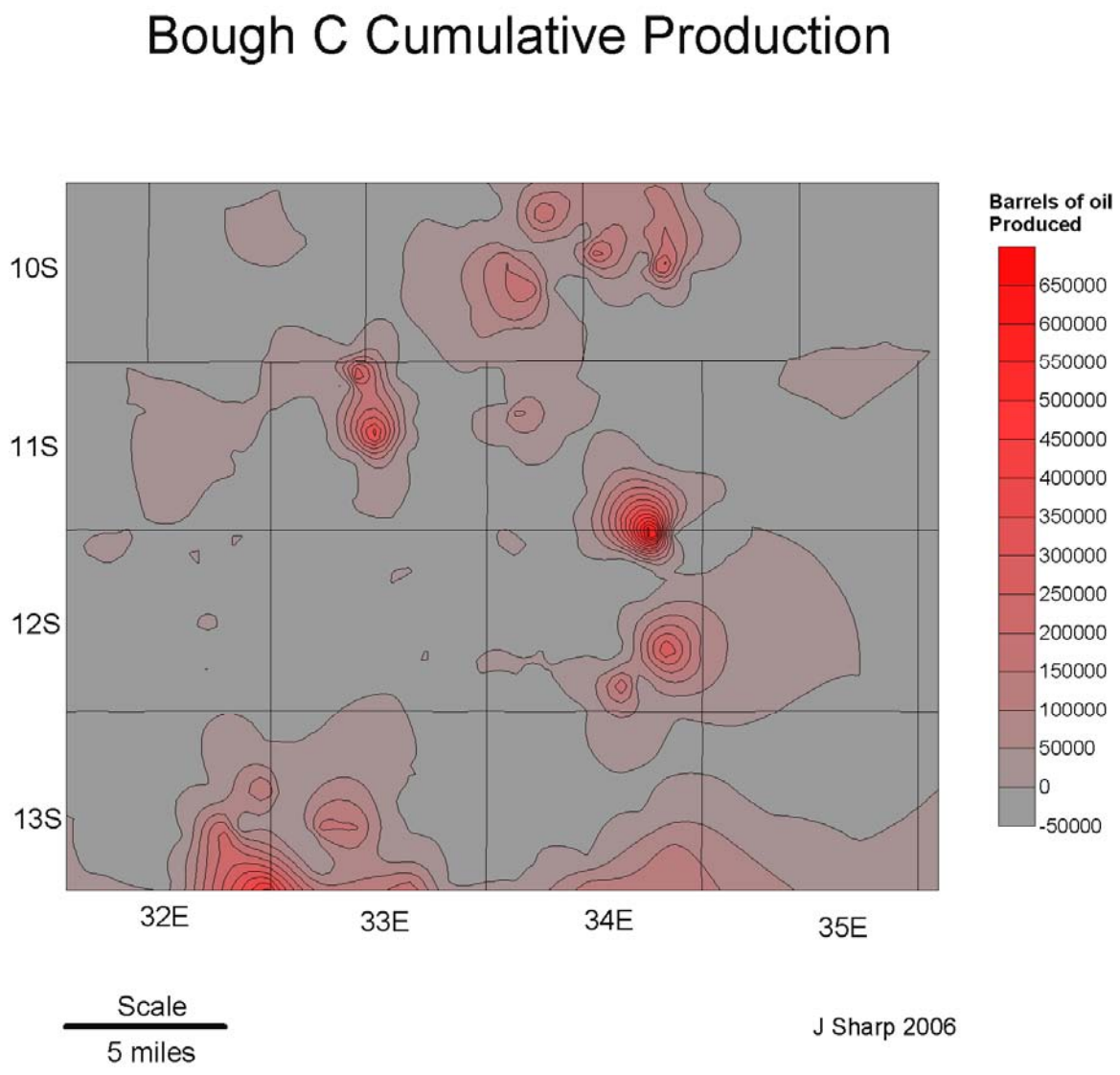


Fig. 3.9. Contour map of cumulative oil production from wells productive from the Bough C zone in the detailed project area.

Table 3.2. Cumulative Oil, Gas and Water Production Data for the 241 Wells Analyzed in the Detailed Study Area*

	Bough A	Bough A + B	Bough B	Bough B + C	Bough C	Bough C + D	Bough D	Bough A + B + C + D	Totals
Cumulative oil, bbls	554323	340623	419428	719475	4032341	2439081	1282871	169222	***9957364***
% oil	5.57%	3.42%	4.21%	7.23%	40.50%	24.50%	12.88%	1.70%	***100.00%***
Cumulative gas, MCF	1637370	391337	896582	3692646	6850480	3165094	4942738	165689	***21741936***
% gas	7.53%	1.80%	4.12%	16.98%	31.51%	14.56%	22.73%	0.76%	***100.00%***
Cumulative water, bbls	704903	865800	394348	1474241	6924880	2523138	1984326	170960	***15042596***
% water	4.69%	5.76%	2.62%	9.80%	46.04%	16.77%	13.19%	1.14%	***100.00%***
Number of wells	5	4	5	4	30	8	14	1	***71***

* includes cumulative production and percent of total production obtained from the Bough A, Bough B, Bough C and Bough D zones. Cumulative production totals are only applicable to the 241 wells (71 of which were productive from the Bough), but percentages of production from each of the Bough members are thought to be representative for all Bough reservoirs in the project area.

2. ***Stratigraphic data acquisition in wells:*** Productive Bough carbonate zones were correlated in 224 wells throughout a subsection of the project area. Where penetrated by wells, the tops of the Bough A, Bough B, Bough C and Bough D members (Fig. 3.4) were correlated with gamma ray, resistivity, and other borehole logs. Inasmuch as the traps that form oil reservoirs in Bough strata are largely stratigraphic (see Broadhead, 1999a, b; Cys, 1986; Cys and Mazzullo, 1985; Malek-Aslani, 1985; Wahlman, 2001), the acquisition of reliable, consistent, and accurate stratigraphic data is essential to the geologic analysis of reservoirs and traps. In order to ensure adequate, consistent and correct correlations, 14 reference cross sections had been produced throughout the project area utilizing 126 wells during the first (2005) project year (Fig. 3.10). The wells were rigorously correlated into closed loops so as to eliminate correlation inconsistencies and errors; these correlations were reevaluated and revised where necessary in 2006 and additional wells were correlated to obtain as even a density of data as existing (e.g. already drilled) wells allow. The latitude and longitude of the well locations were calculated using a digital land grid and *Geographix* software (*Geographix* is a registered trademark of Landmark Graphics, Inc.) that uses the digital land grid to

convert surveyed footage measurements of wells from section boundaries into latitude and longitude. Well names, locations and depths to the tops of the Bough A, Bough B, Bough C and Bough D members were entered into an Excel database. Maps made with data from the additional wells will reveal variations in stratigraphy that may be related to the localization of hydrocarbon traps. In total, the tops of the Bough A, Bough B, Bough C and Bough D were correlated in 224 wells.

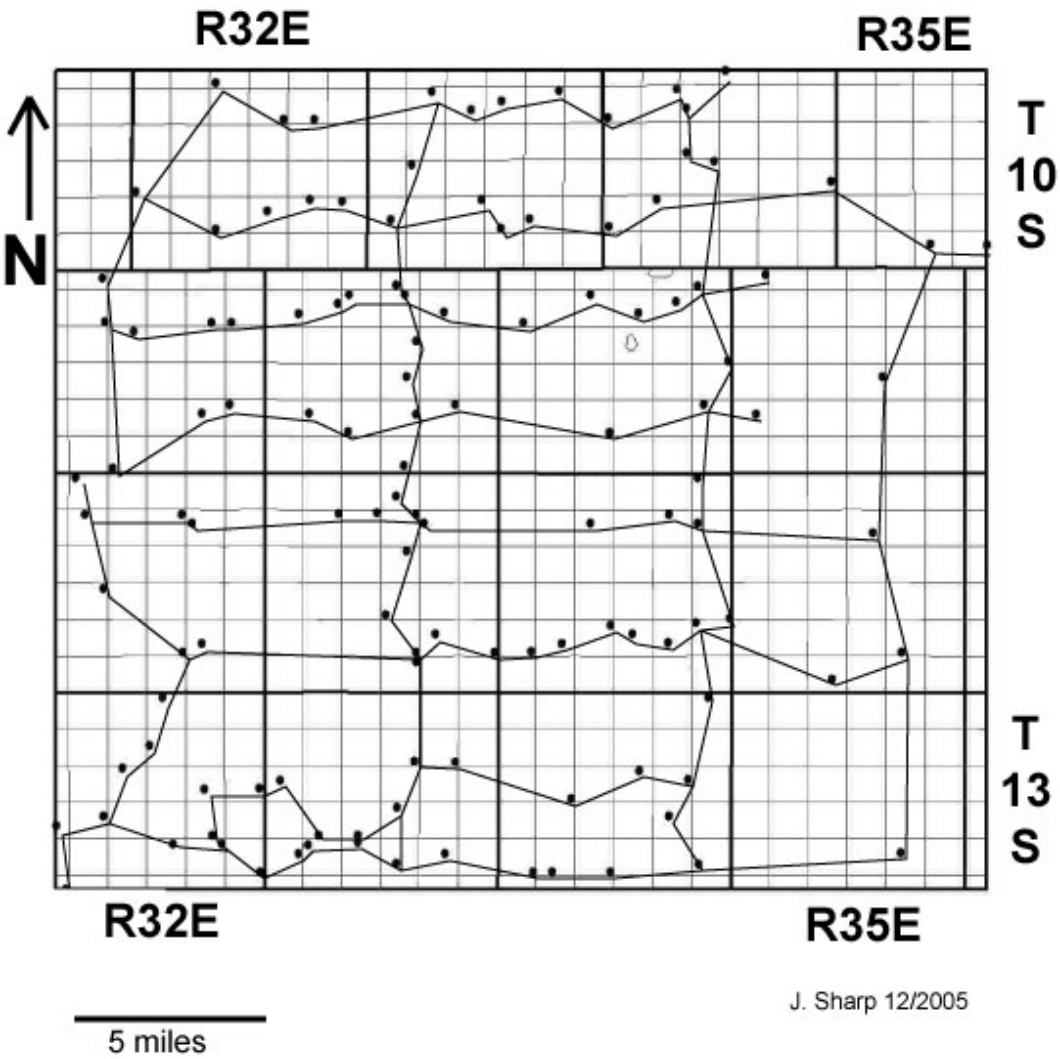


Fig. 3.10. Locations of cross sections and wells utilized in constructing the cross sections in the detailed study area.

3. ***Stratigraphic data mapping and analysis*:** Digital isopach maps were made of the Bough A, Bough B and Bough C members using all wells within the area of detailed investigation (Figs. 3.11–3.13). These maps were produced using *Surfer 8*, a modern and sophisticated contouring program (*Surfer 8* is a registered trademark of Golden Software, Inc.). Data were gridded with a kriging method. The boundaries of oil pools productive from the Permo-Penn carbonate reservoirs were superimposed upon the isopach maps (Figs. 3.14–3.16). Additional isopach maps were prepared that indicate which of the 224 wells have been productive from the mapped Bough member (Figs. 3.17–3.19).

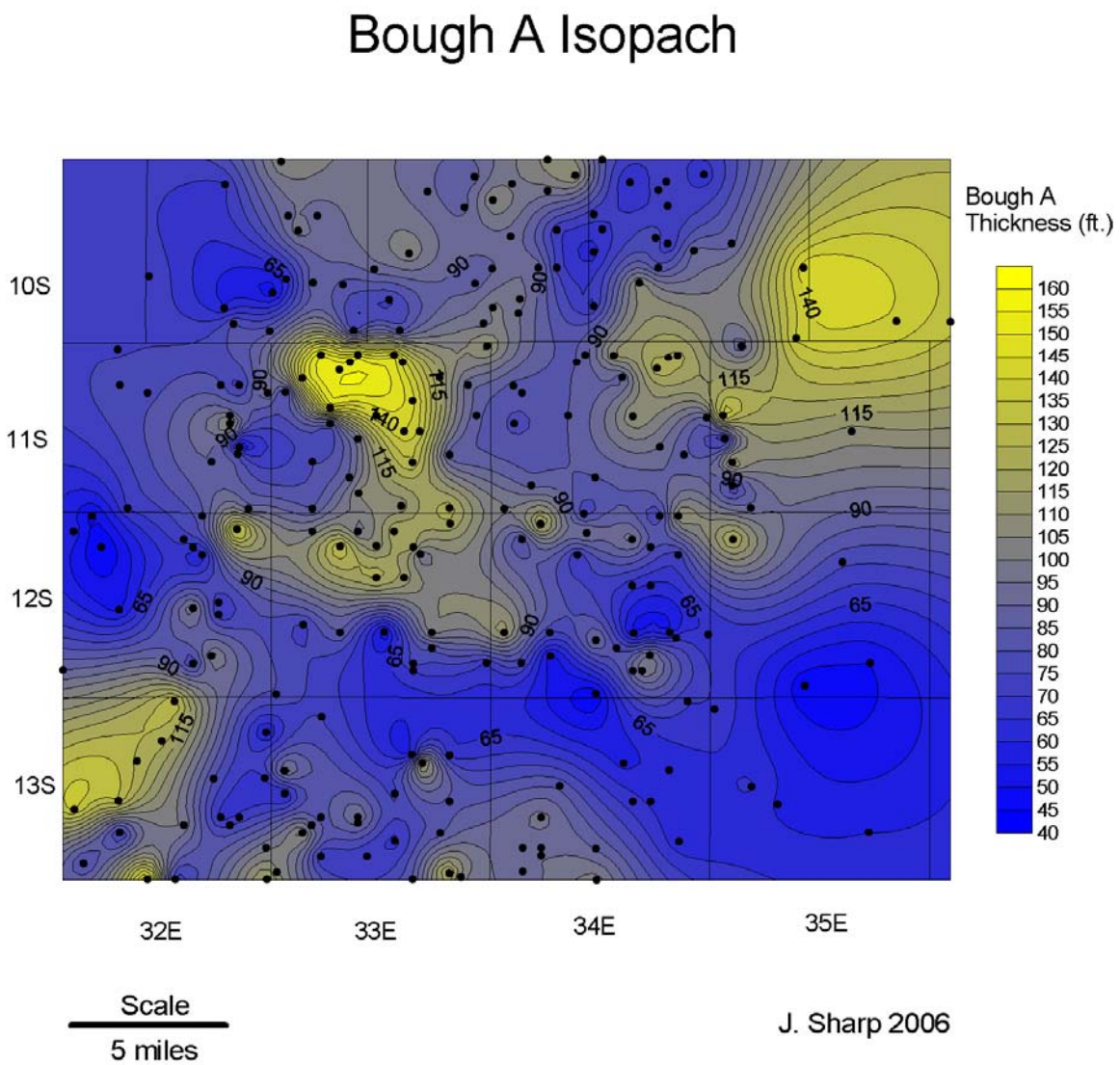


Fig. 3.11. Isopach map of the Bough A member, detailed project area. Contours in feet.

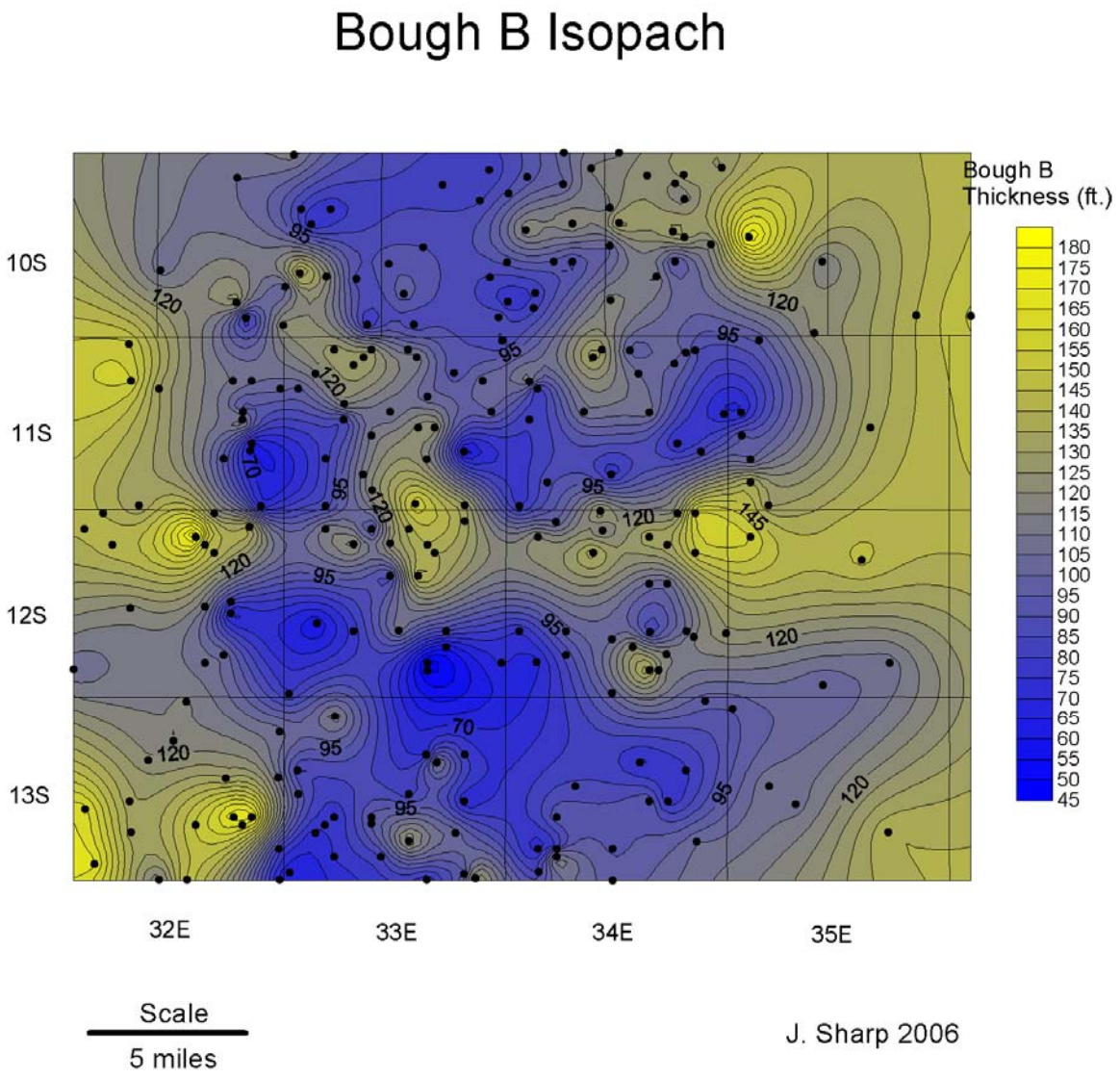


Fig. 3.12. Isopach map of the Bough B member, detailed project area. Contours in feet.

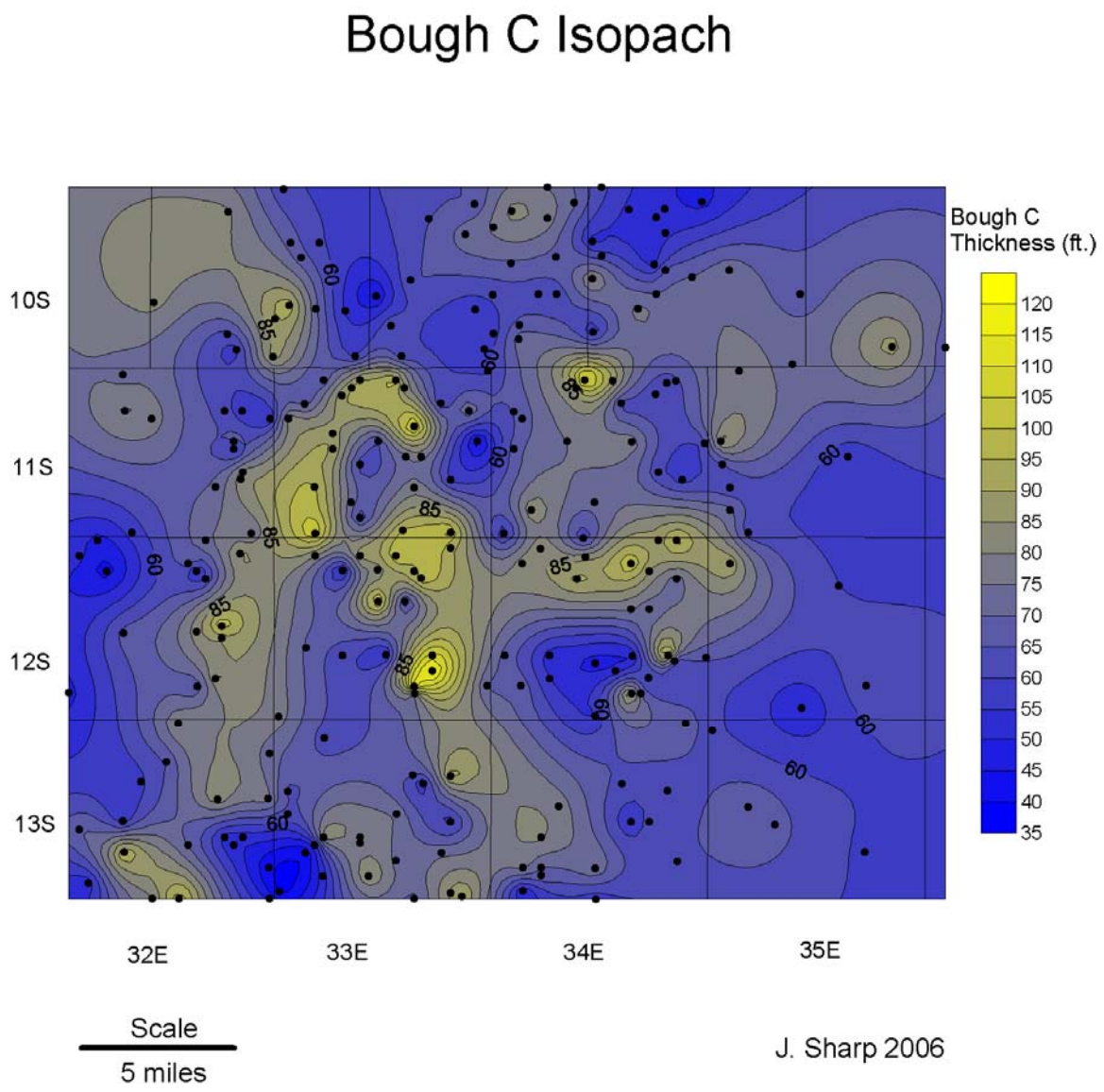


Fig. 3.13. Isopach map of the Bough C member, detailed project area. Contours in feet.

Bough A Isopach

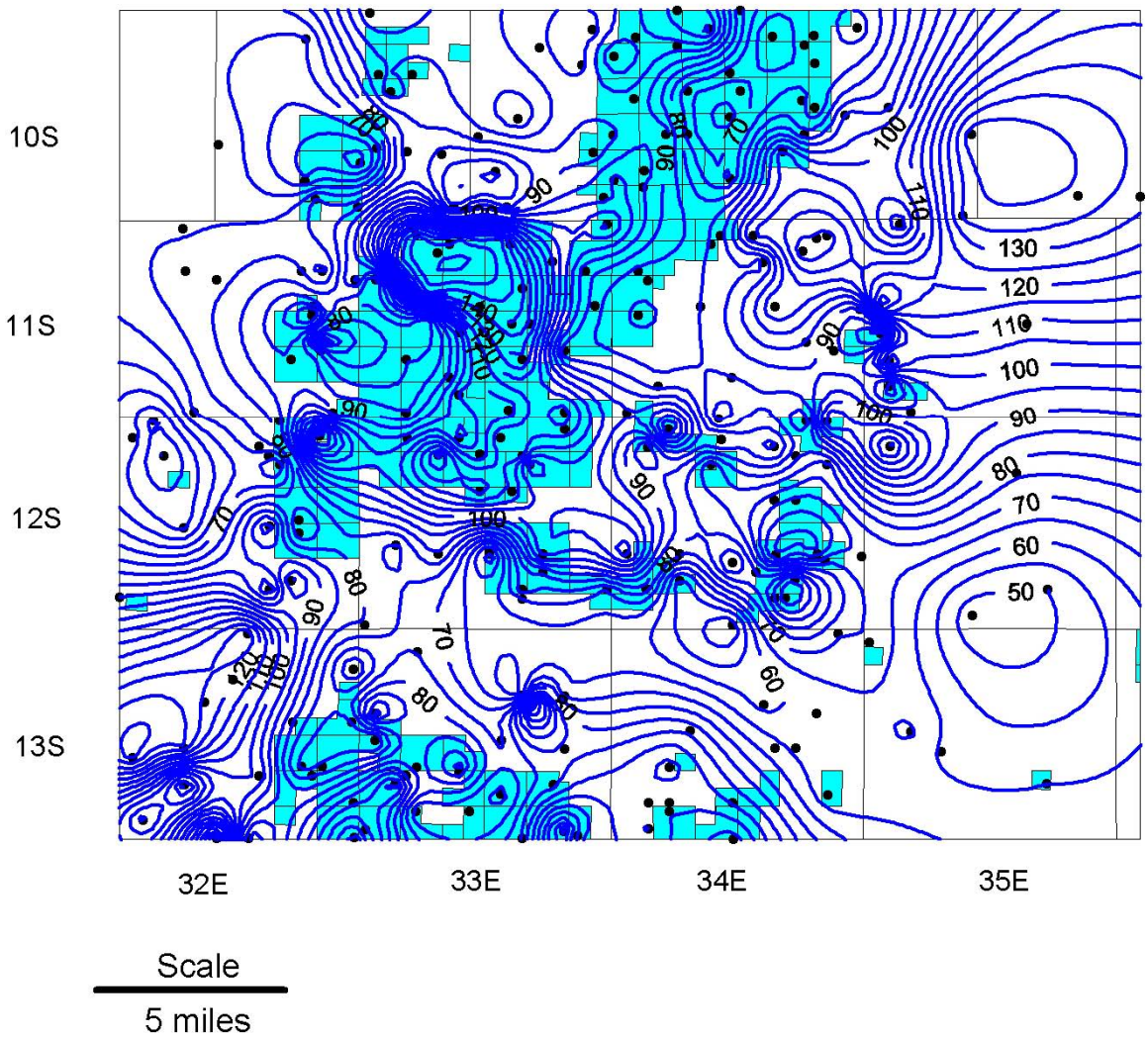


Fig. 3.14. Isopach map of the Bough A member, detailed project area with boundaries of reservoirs productive from Upper Pennsylvanian and Lower Permian strata superimposed. Contours in feet.

Bough B Isopach

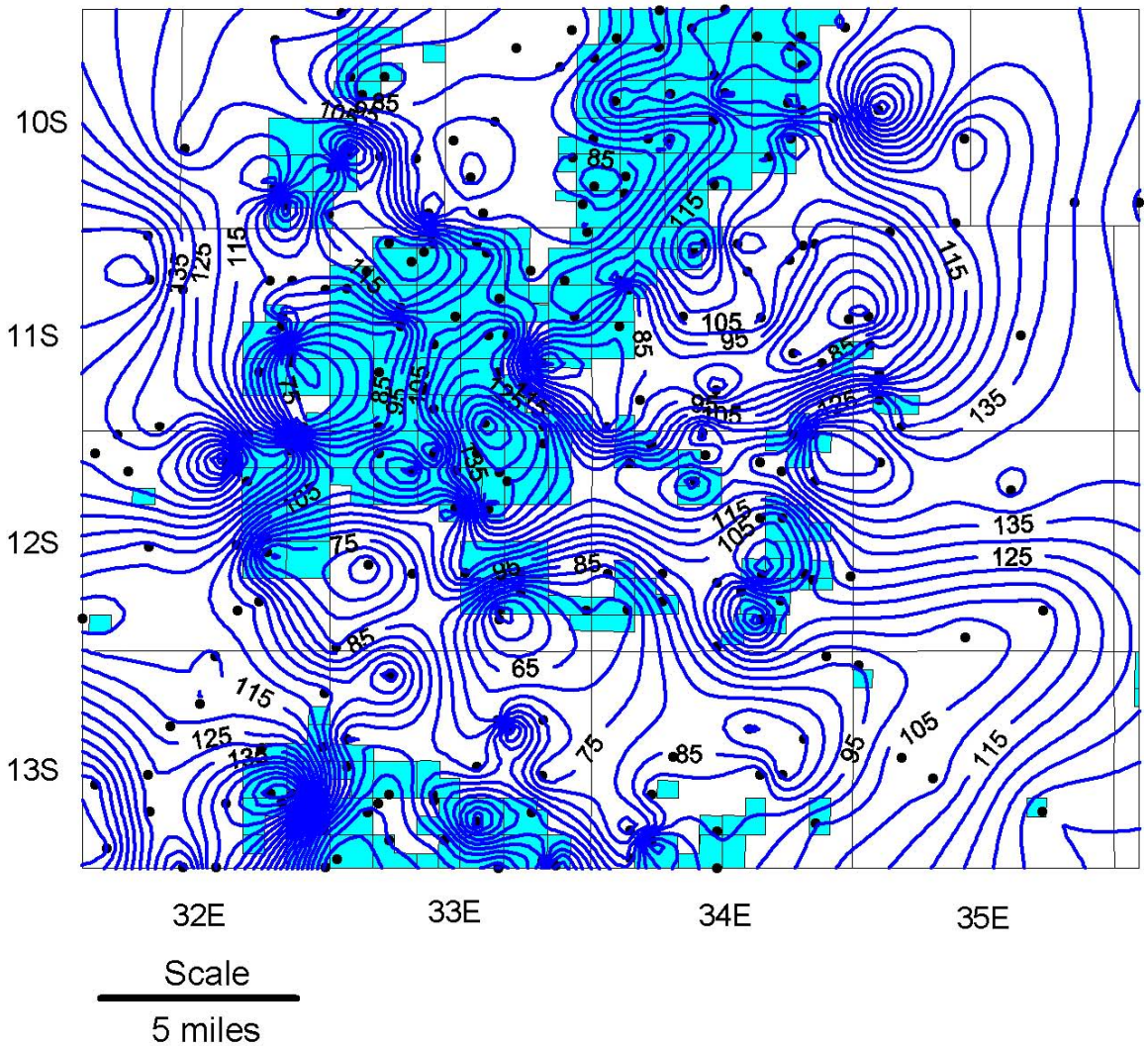


Fig. 3.15. Isopach map of the Bough B member, detailed project area with boundaries of reservoirs productive from Upper Pennsylvanian and Lower Permian strata superimposed. Contours in feet.

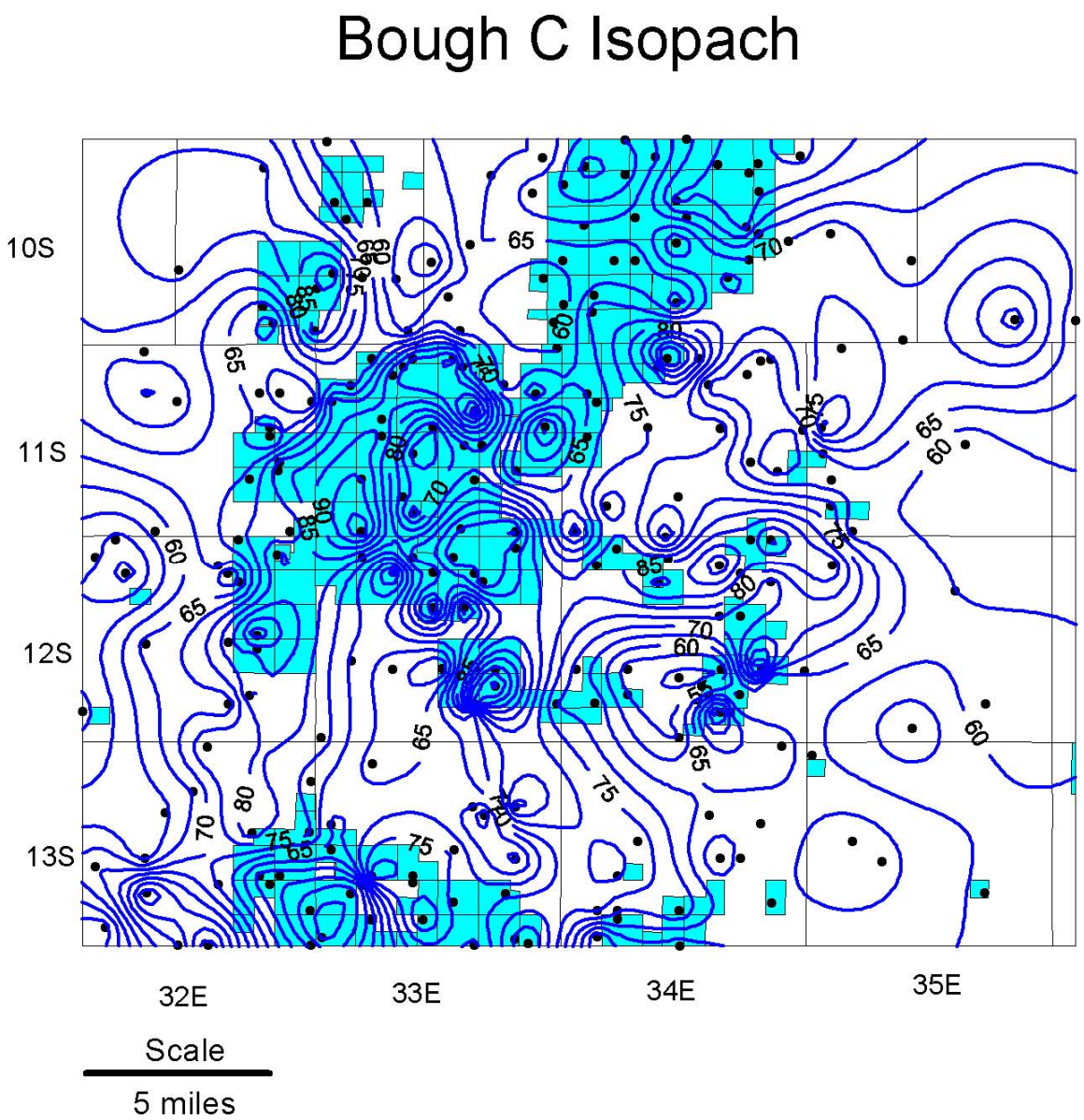


Fig. 3.16. Isopach map of the Bough C member, detailed project area with boundaries of reservoirs productive from Upper Pennsylvanian and Lower Permian strata superimposed. Contours in feet.

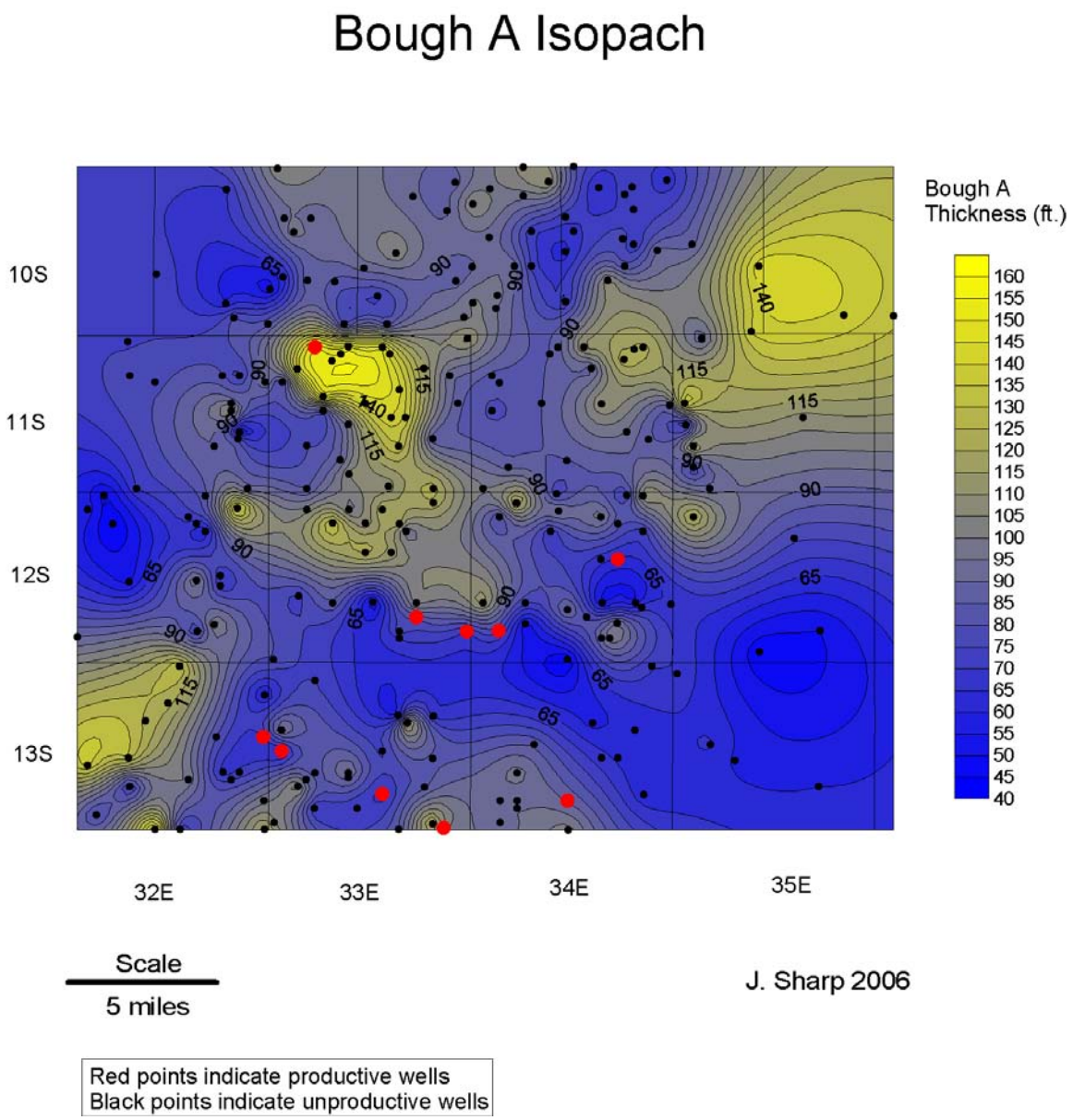


Fig. 3.17. Isopach map of Bough A member. Dots indicate wells used to construct the map. Red symbols indicate wells productive from the Bough A; black symbols indicate wells that are not productive from the Bough A. Contours in feet.

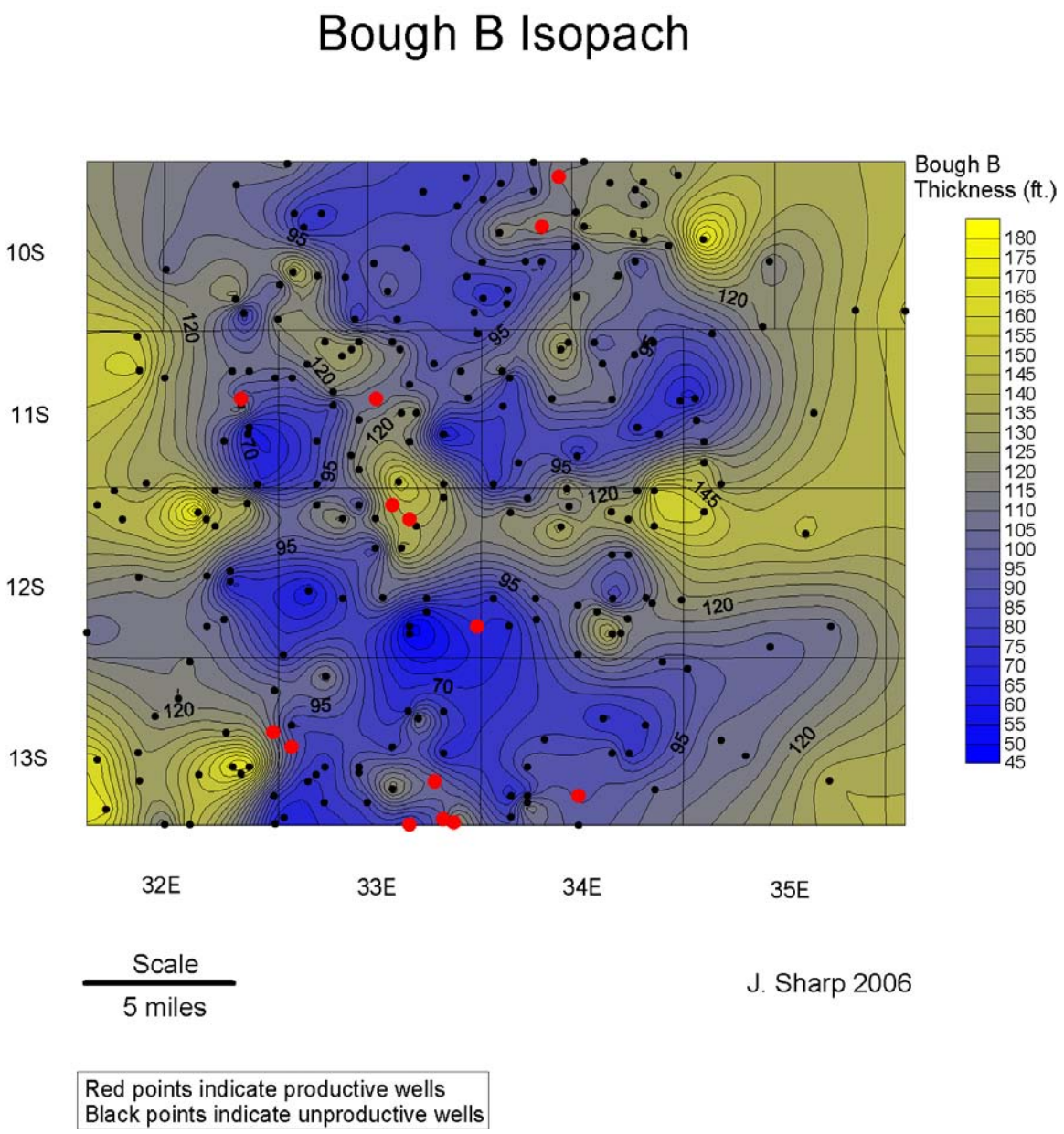


Fig. 3.18. Isopach map of Bough B member. Dots indicate wells used to construct the map. Red symbols indicate wells productive from the Bough B; black symbols indicate wells that are not productive from the Bough B. Contours in feet.

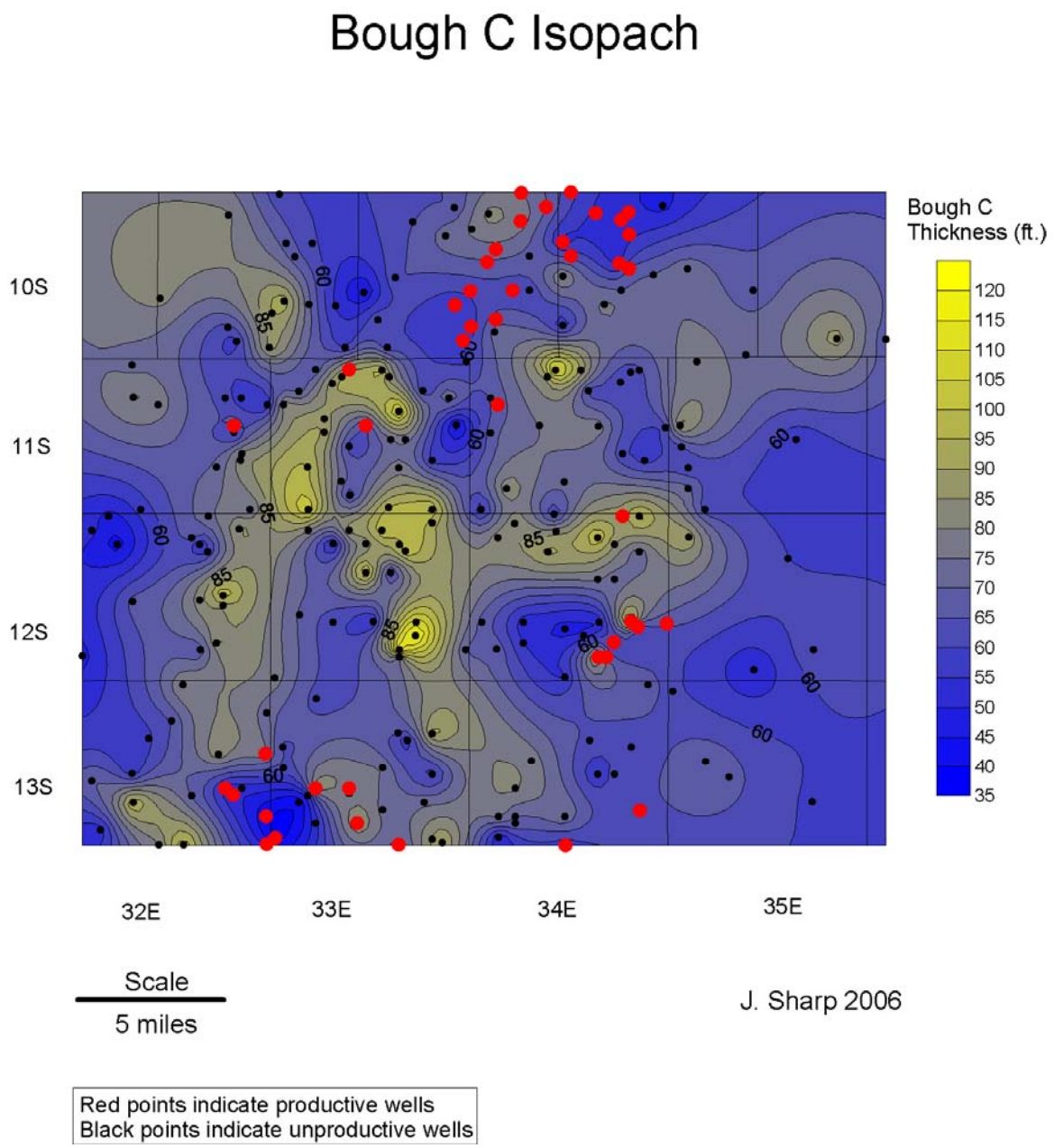


Fig. 3.19. Isopach map of Bough C member. Dots indicate wells used to construct the map. Red symbols indicate wells productive from the Bough C; black symbols indicate wells that are not productive from the Bough C. Contours in feet.

4. ***Structural data mapping and analysis*:** Digital structure contour maps were prepared with *Surfer 8* for the upper surfaces of the Bough A, B, C and D members in the area selected for more detailed investigation (Fig. 3.6). Structure maps of the Abo Formation (Fig. 3.20) and the Mississippian System (Fig. 3.21) had been prepared during the first project year (2005). The boundaries of oil pools productive from the Permo-Pennsylvanian carbonate reservoirs were superimposed upon the structure maps (Figs. 3.22–3.23). Additional structure maps of the Bough A, B and C members were prepared that indicate which of the 224 wells have been productive from the mapped Bough member (Figs. 3.24–3.26). No exact correlations between the Abo structure map and the locations of oil reservoirs are apparent. However, there is a general correlation between the locations of oil reservoirs and closed contours delimiting positive areas on the Mississippian structure map.

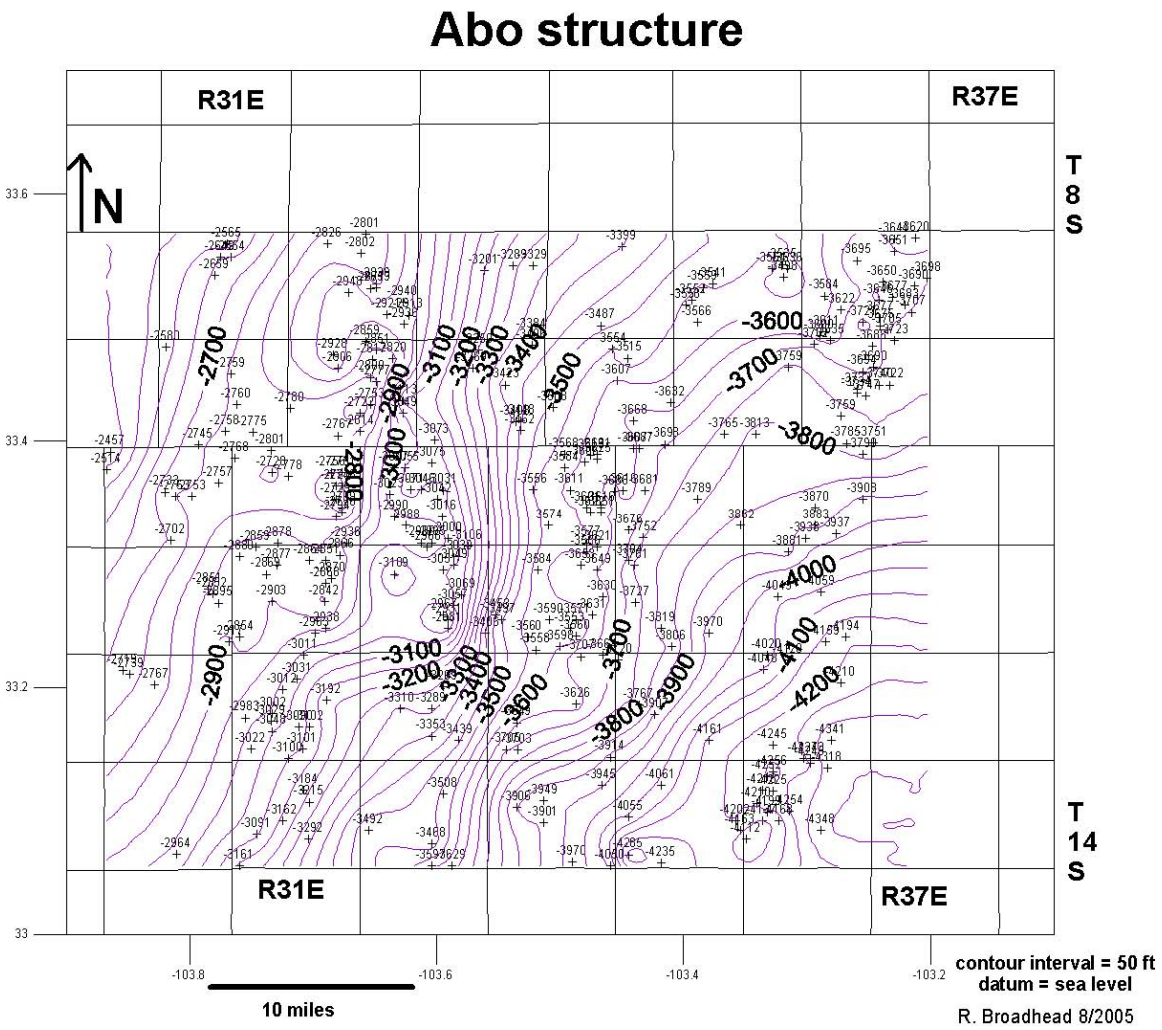


Fig. 3.20. Structure contours on top of Abo Formation (Lower Permian), Bough intrashelf project area.

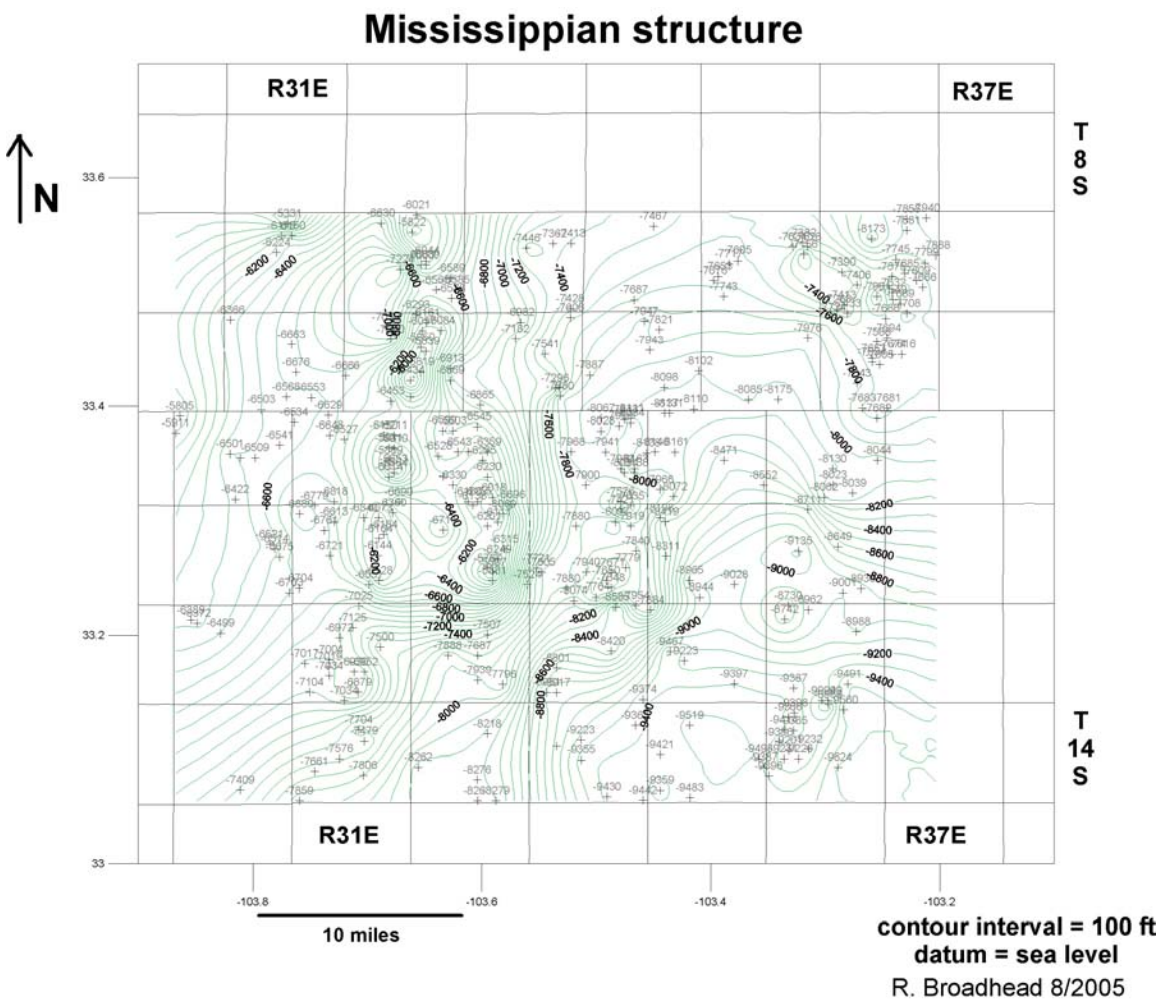


Fig. 3.21. Structure contours on top of the Mississippian System, Bough intrashelf project area.

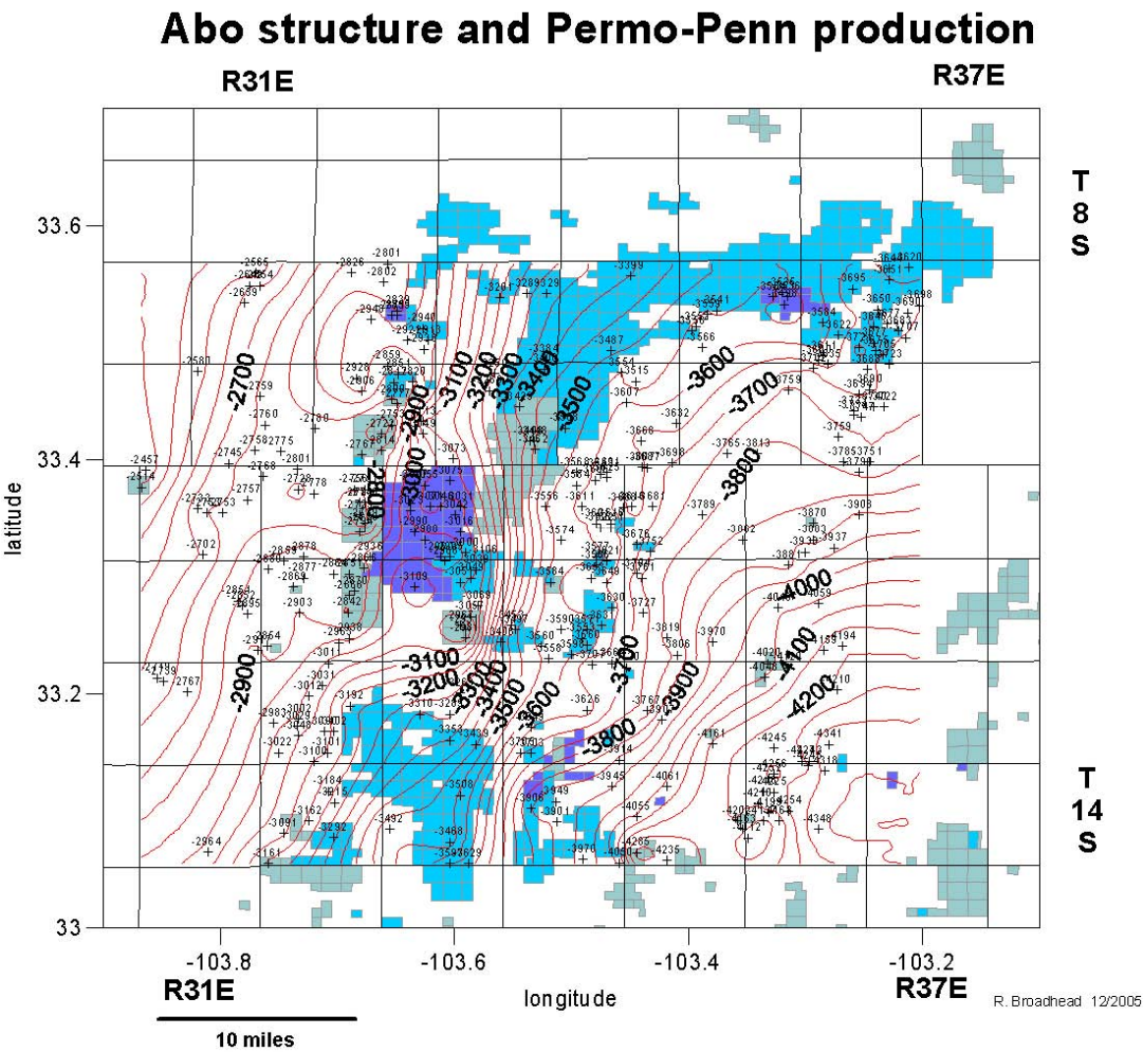


Fig. 3.22. Structure contours on top of Abo Formation (Lower Permian), Bough intrashelf project area with boundaries of reservoirs productive from Upper Pennsylvanian and Lower Permian strata superimposed. Contours in feet. Datum = sea level.

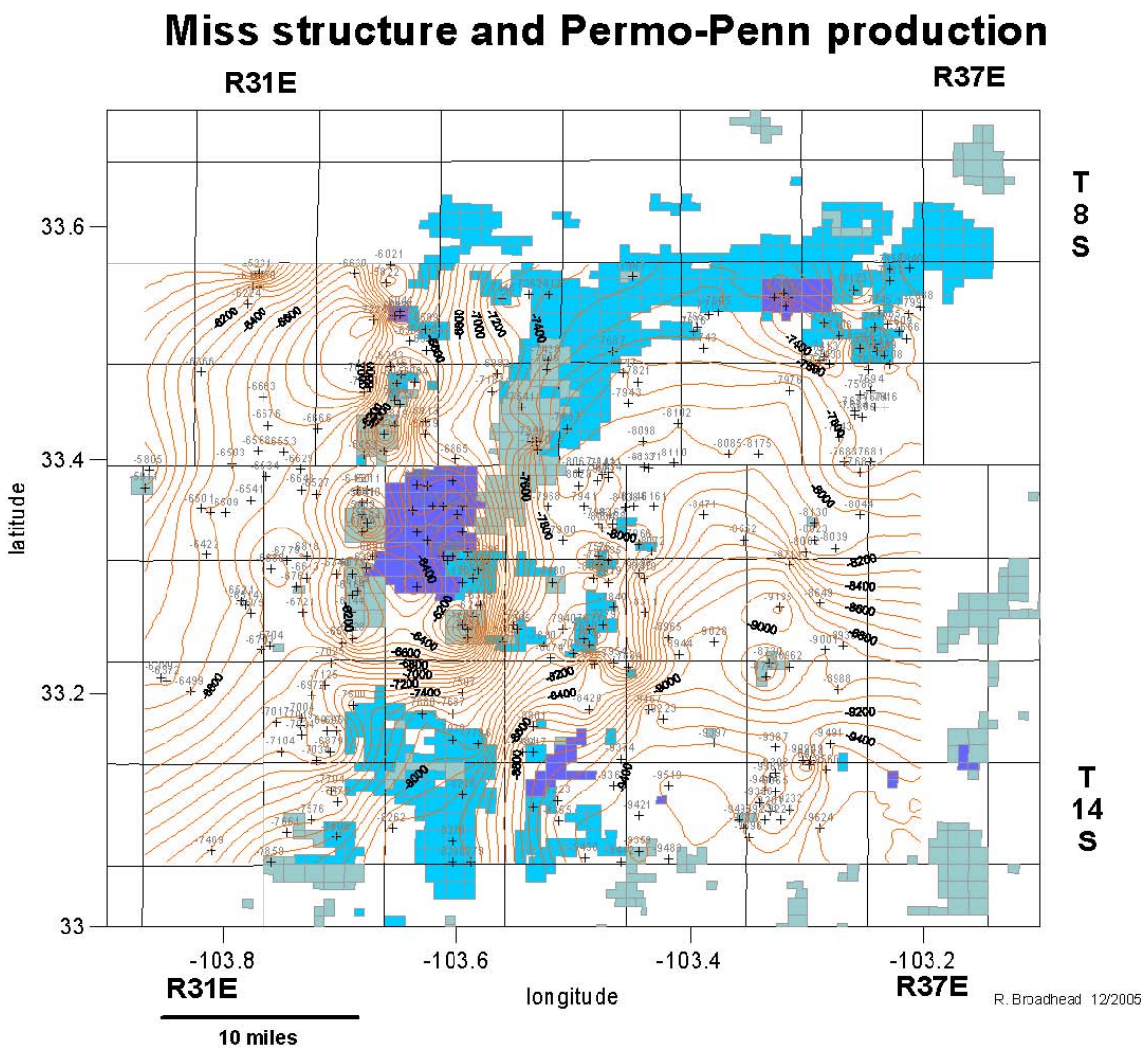


Fig. 3.23. Structure contours on top of the Mississippian System, Bough intrashelf project area with boundaries of reservoirs productive from Upper Pennsylvanian and Lower Permian strata superimposed. Contours in feet. Datum = sea level.

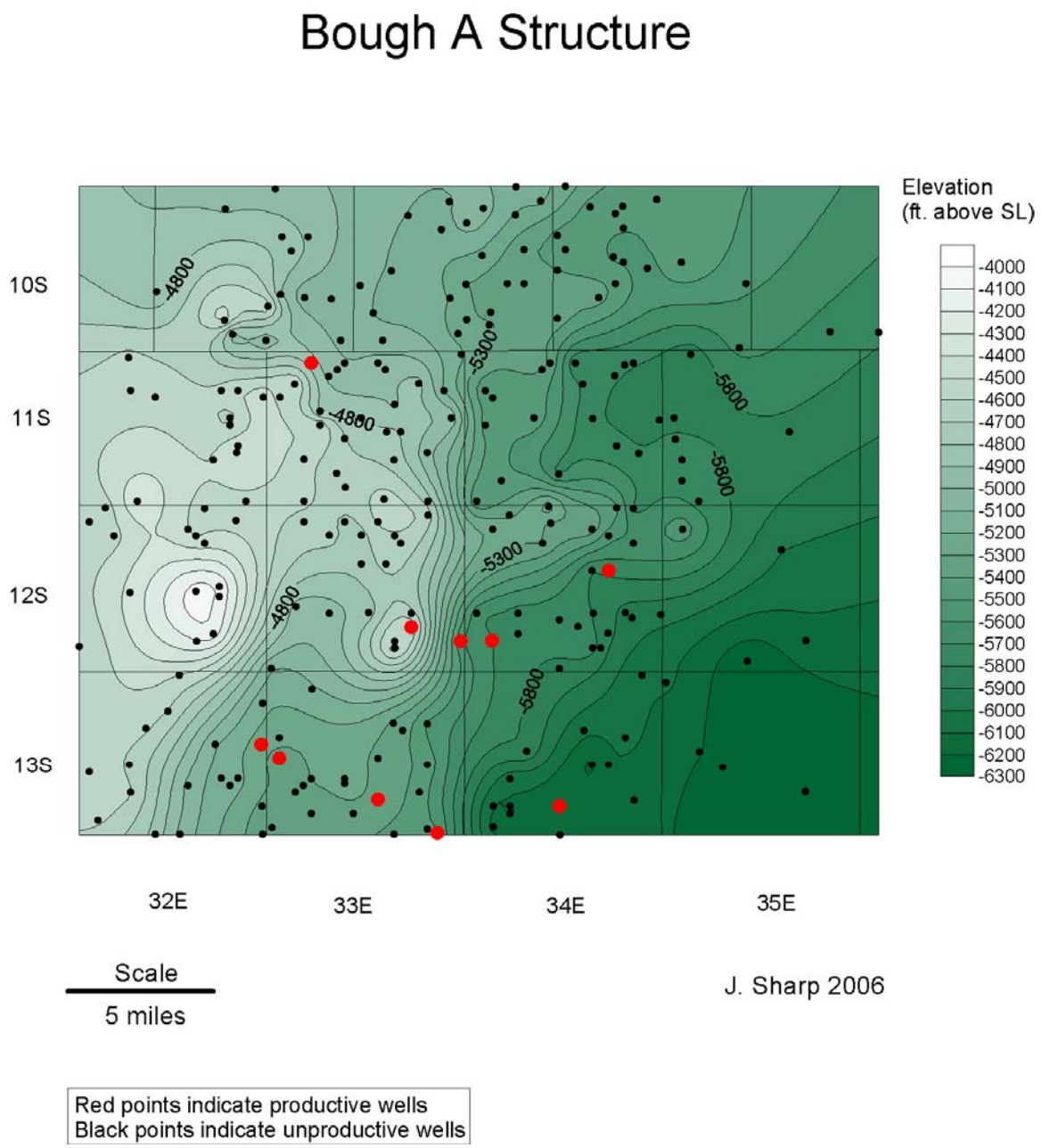


Fig. 3.24. Structure contour map of Bough A member. Dots indicate wells used to construct the map. Red symbols indicate wells productive from the Bough A; black symbols indicate wells that are not productive from the Bough A. Contours in feet. Datum = sea level.

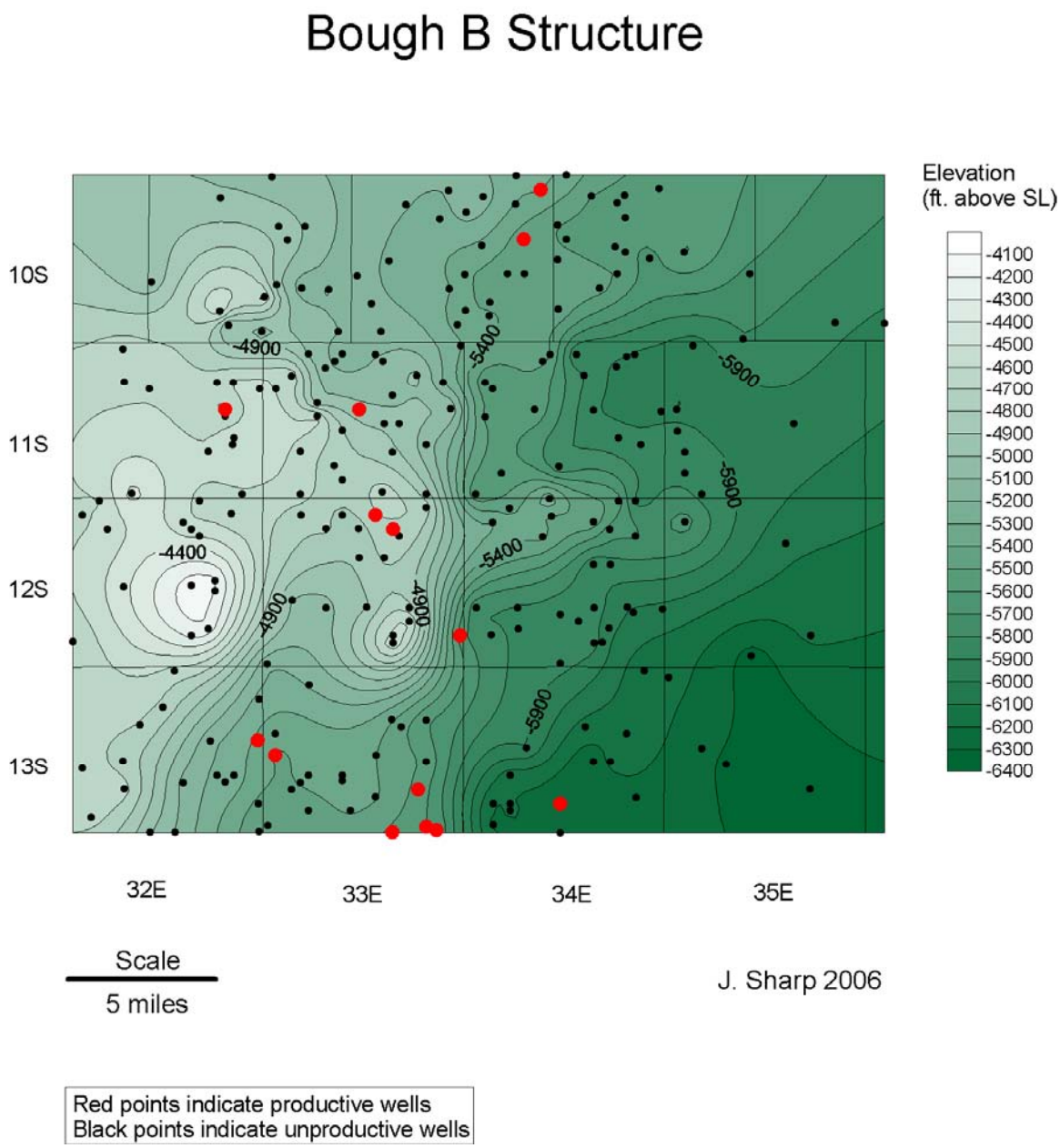


Fig. 3.25. Structure contour map of Bough B member. Dots indicate wells used to construct the map. Red symbols indicate wells productive from the Bough B; black symbols indicate wells that are not productive from the Bough B. Contours in feet. Datum = sea level.

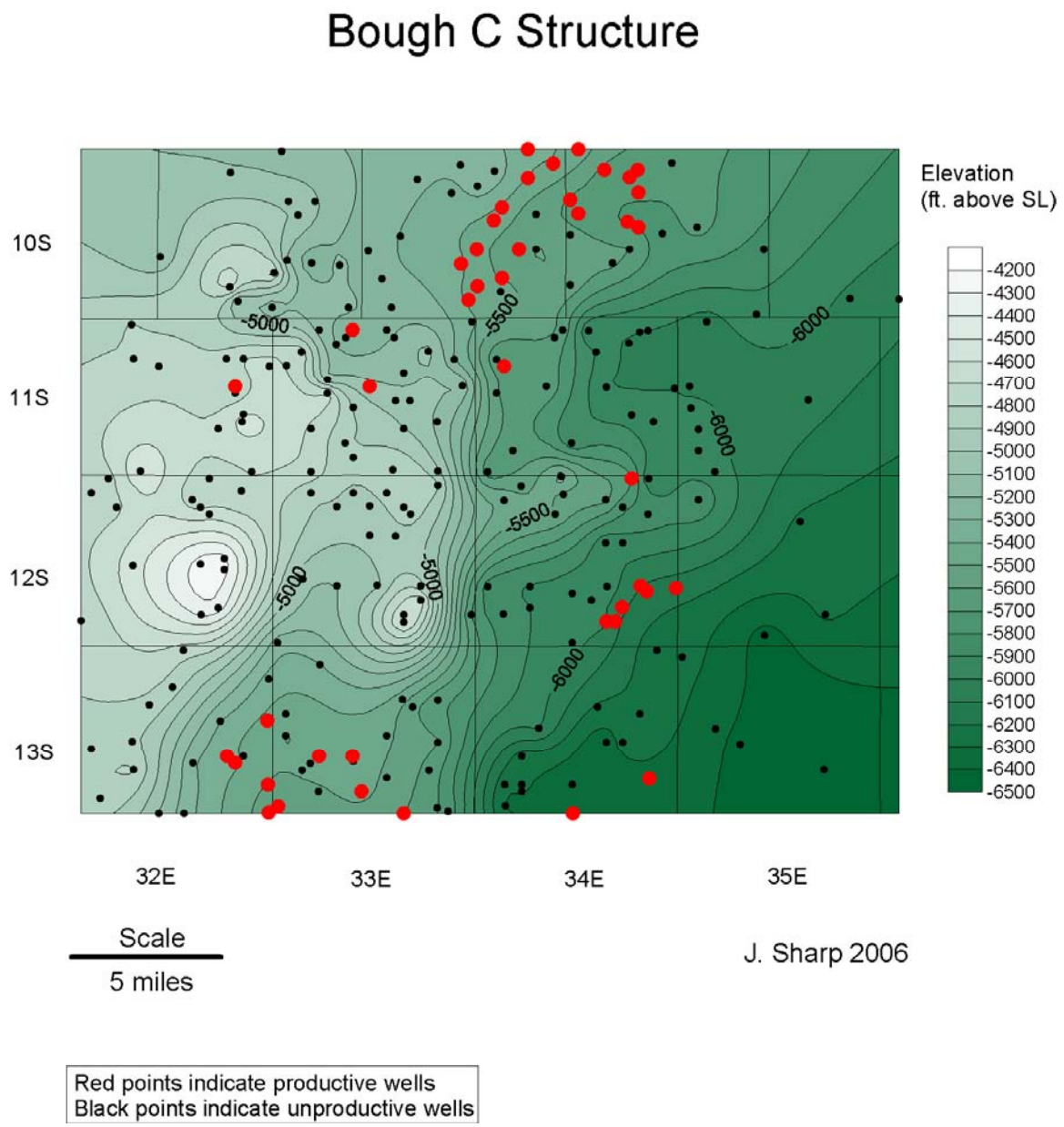


Fig. 3.26. Structure contour map of Bough C member. Dots indicate wells used to construct the map. Red symbols indicate wells productive from the Bough C; black symbols indicate wells that are not productive from the Bough C. Contours in feet. Datum = sea level.

5. ***Paleostructural mapping and analysis*:** This task was mostly completed during 2005 but some additional maps were produced during 2006 from data produced during 2005. The Pennsylvanian and Early Permian were times of active tectonism and structural deformation in New Mexico (e.g. Broadhead, 2001a-c; Kues and Giles, 2004). During this period of tectonism, numerous structures that together formed the Ancestral Rocky Mountains arose out of the Late Paleozoic seas that covered New Mexico and surrounding areas forming upthrown fault blocks and anticlines that were either emergent islands (if they rose to a sufficient elevation) or bathymetrically high spots on the sea floor. Growth of algal mounds and associated strata that form oil reservoirs in Permo-Pennsylvanian strata is thought to be controlled primarily by location on the tops and flanks of structures that were rising during the deposition of the algal mound reservoirs (Malek-Aslani, 1985; Cys and Mazzullo, 1985; Cys, 1986; Wahlman, 2001). Therefore, the mapping and analysis of paleostructures of Pennsylvanian and Early Permian age should be a primary factor in the analysis and prediction of the trends and locations of oil reservoirs in Permo-Pennsylvanian strata.

   Timing of the formation of structures and therefore the location of paleostructures is reflected in the thinning of strata deposited during the time period during which the structure was formed. For Ancestral Rocky Mountain paleostructures, this is apparent in the thinning of Pennsylvanian through Lower Permian strata over positive paleostructural elements and the thickening of Pennsylvanian and Lower Permian strata over negative paleostructural elements (Fig. 3.27). Therefore, an isopach map of the interval between the top of the Abo Formation (see Fig. 3.22) and the top of the Mississippian System (see Fig. 3.21) will reveal positive paleostructural elements as enclosed thin areas and negative paleostructural elements as thick areas.

   An isopach map of the interval between the top of the Abo Formation (Lower Permian) and the top of the Mississippian System was prepared with *Surfer 8* using the structural data previously acquired (Fig. 3.28). Thin areas denote positive paleostructural elements. Based upon the isopach maps, the axes

and directions of plunge of positive paleostructural elements were interpreted and mapped (Fig. 3.29). The Abo-Mississippian isopach map was overlain on a wire-frame image (see Fig. 3.30) of Mississippian structure (Fig. 3.31). The resulting map clearly shows an excellent correlation between Mississippian structure and the thickness of the interval between the top of the Abo and the top of Mississippian, indicating major structural movement between the end of Mississippian time and the end of Abo time; where at appropriate relation to sea level, the phylloid algal mounds grew on top of the emerging structures. The Abo-Mississippian isopach map was then overlain on the locations of oil reservoirs that produce from Permo-Penn carbonate strata (Fig. 3.32). As predicted, most, but not all, areas characterized by Permo-Pennsylvanian production coincide with thin areas on the Abo-Mississippian isopach map. Inasmuch as the thin areas represent positive paleostructural elements, it appears that the Abo-Mississippian isopach map, which acts as a proxy for an Ancestral Rocky Mountains paleostructure map, should be a key ingredient for the FEE tool when it comes to predicting the trends and locations of oil reservoirs formed by carbonate stratigraphic traps.

The correlation between paleostructure and the locations of oil reservoirs in Permo-Penn strata is not absolute. Examination of Fig. 3.32 reveals indicated paleostructures that have been proven by drilling to be barren of production. Conversely, there are some productive areas that are not associated firmly with paleostructures. It is believed several factors, principally paleo-water depth over the structures, may be responsible for some structures being barren of production. If the paleostructure caused the bottom of the sea floor to rise up too high then water may have been too shallow to sustain significant algal mound growth in the area above the structure; however, in this case, water may have been sufficiently deep over the flanks of the structure to sustain algal mound growth. If the area over the structure was emergent, then it would have been subject to erosion rather than deposition. On the other hand if water depth was too great, conditions such as nutrient and oxygen supply and the amount of sunlight reaching bottom waters would have been incorrect for the growth of algal mounds and associated reservoirs.

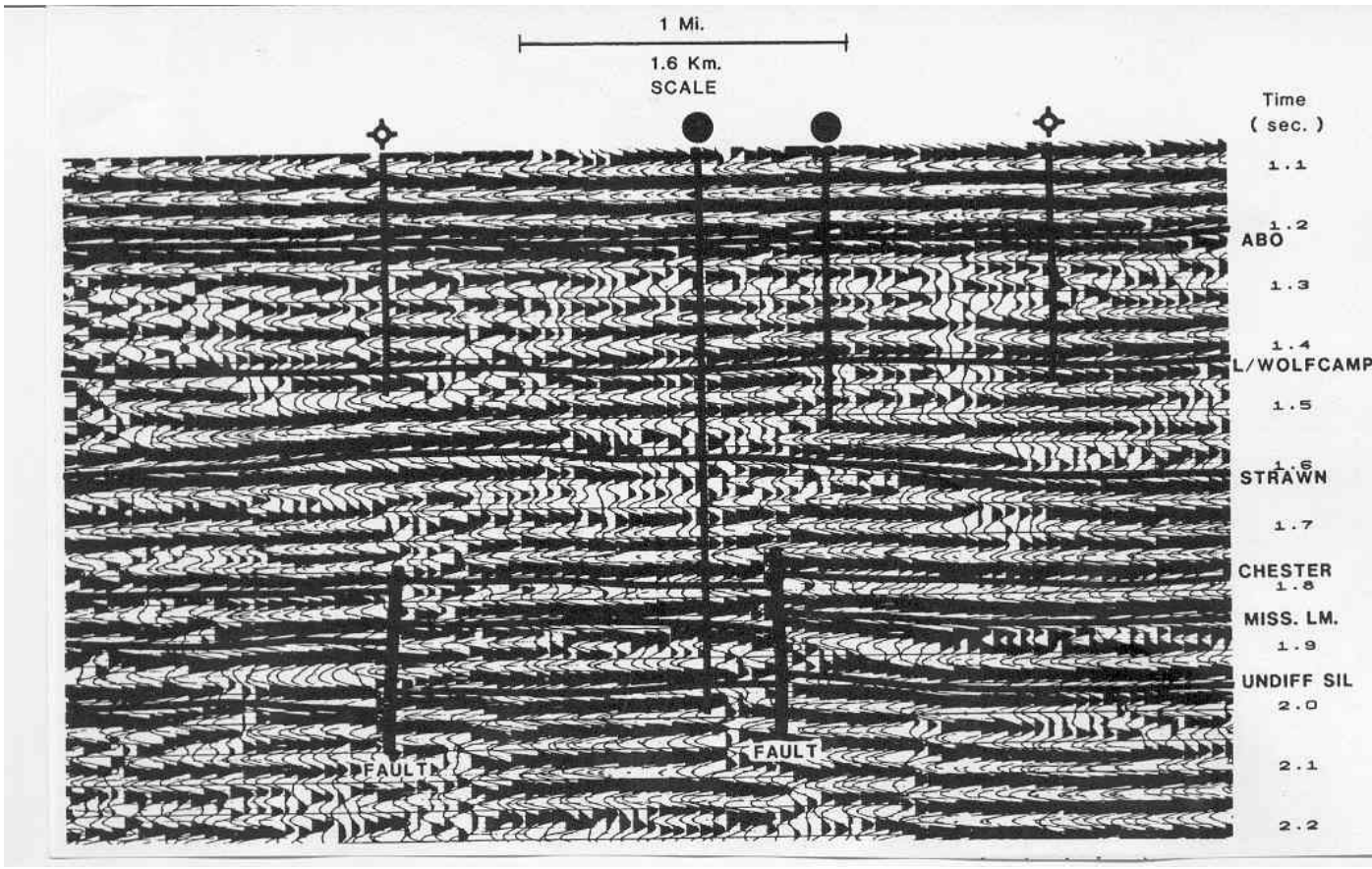


Fig. 3.27. Reflection seismic line across a Permo-Pennsylvanian phylloid algal mound, showing thinning of Upper Pennsylvanian and Lower Permian strata over a paleostructure. From Cys (1986).

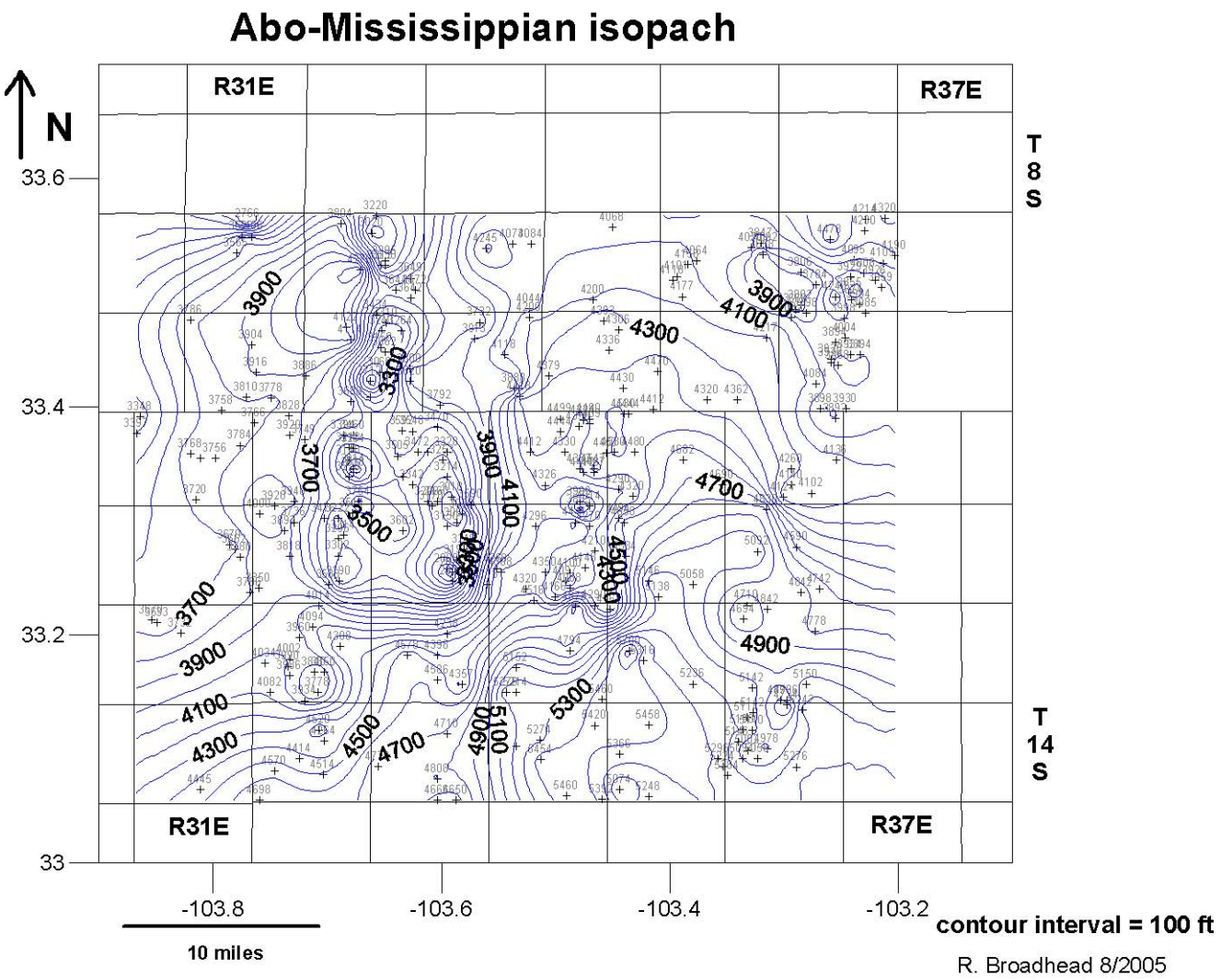


Fig. 3.28. Isopach map of the stratigraphic interval between the top of the Abo Formation (Lower Permian) and the top of the Mississippian System, Bough intrashelf project area.

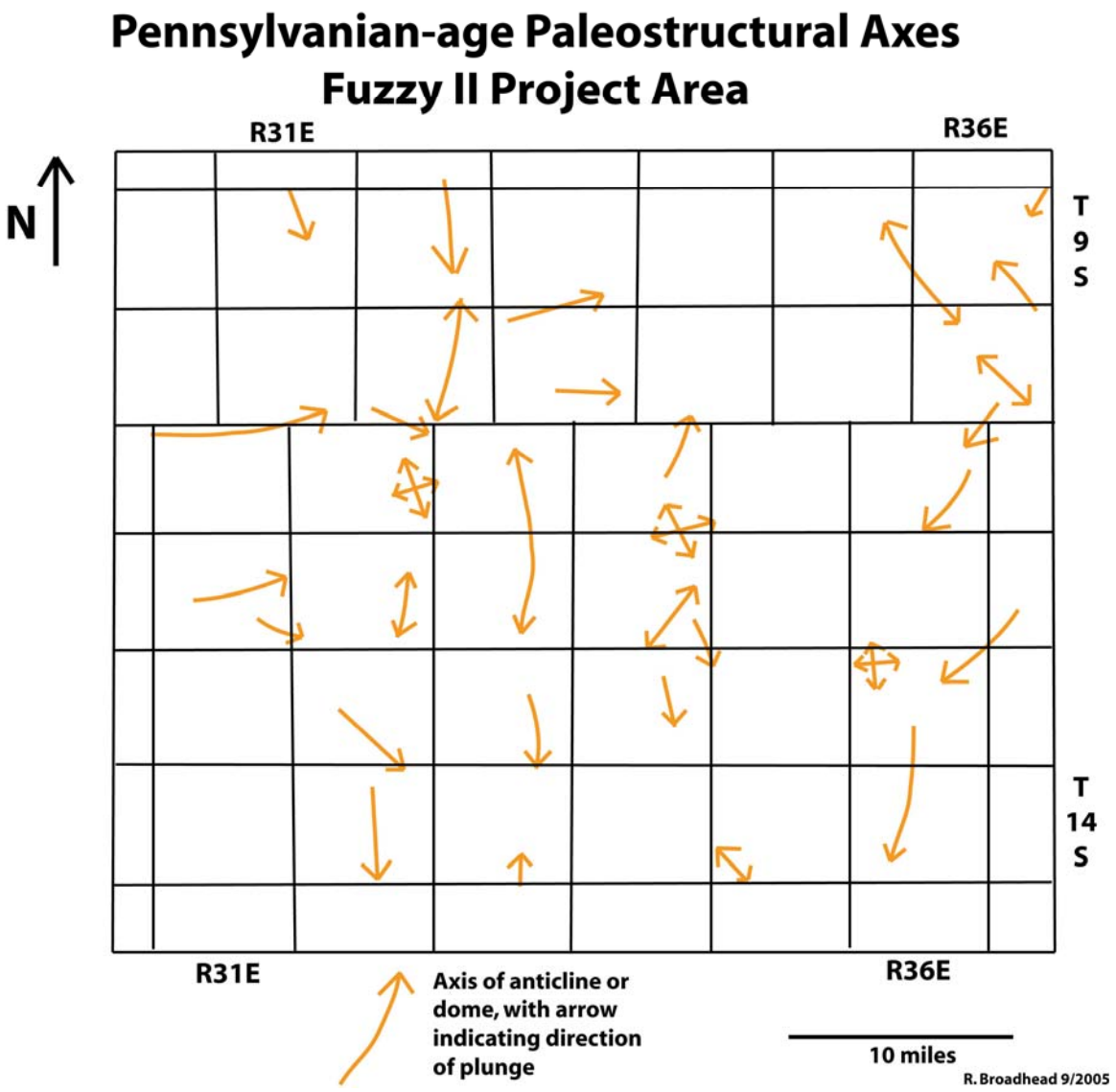


Fig. 3.29. Location of Late Pennsylvanian paleostructural positive elements as determined from the Abo-Mississippian isopach map.

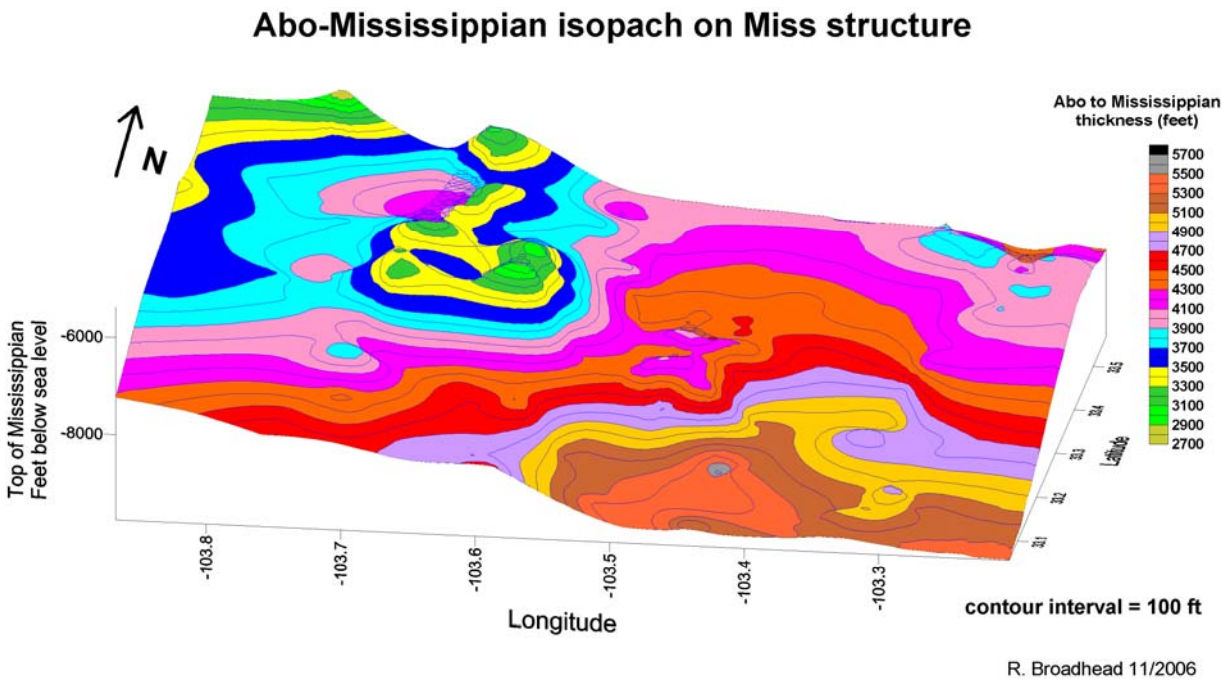


Fig. 3.30. Abo-Mississippian isopach map superimposed on a wireframe relief map of Mississippian structure, Bough intrashelf project area.

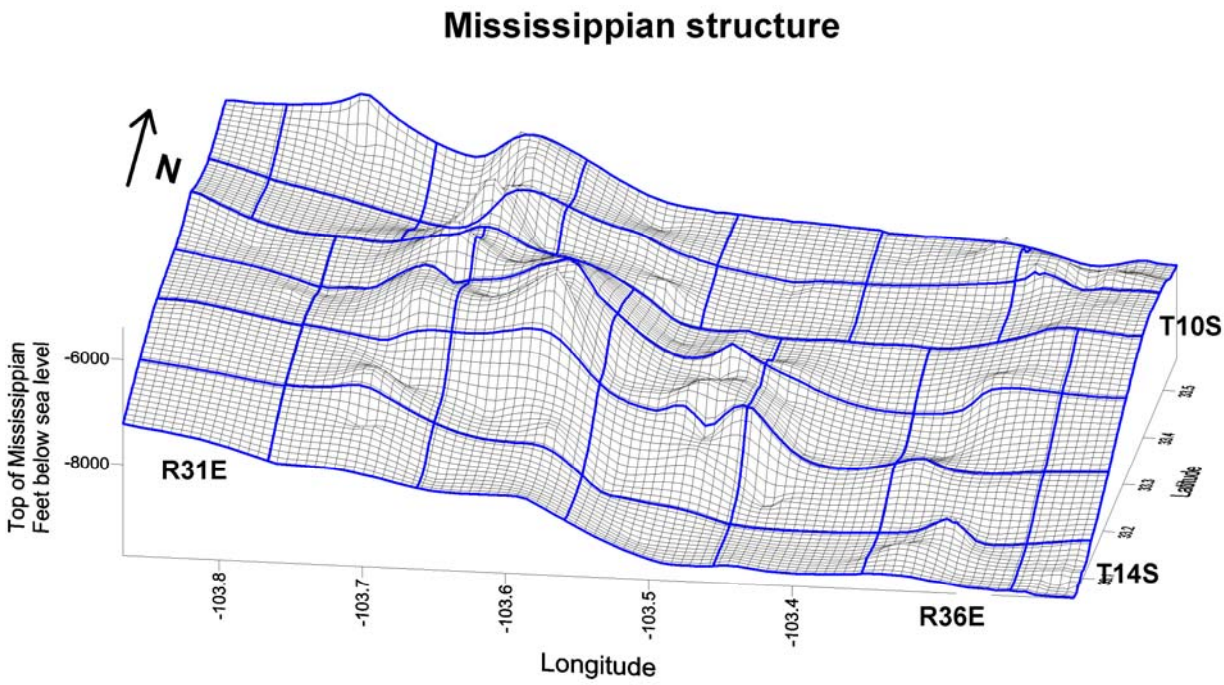


Fig. 3.31. Wireframe relief map of Mississippian structure, Bough intrashelf project area.

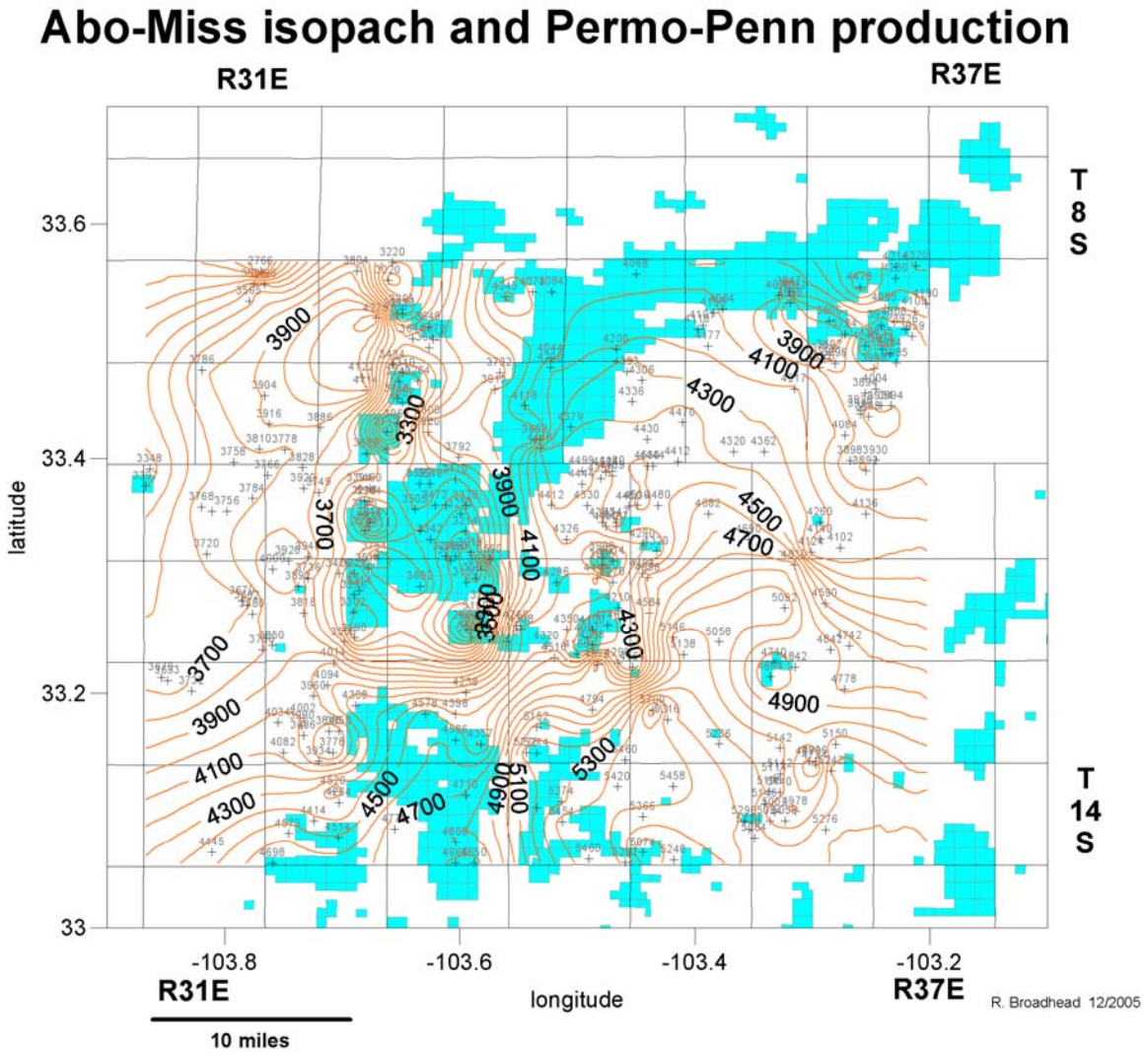


Fig. 3.32. Isopach map of the Abo-Mississippian interval, Bough intrashelf project area with boundaries of reservoirs productive from Upper Pennsylvanian and Lower Permian strata superimposed.

6. ***Reservoir analysis:*** Petrographic analysis of thin sections and acetate peels from two cores in the project area (see Fig. 3.33 for core locations) indicate three petro-facies are present in the Bough: a biomicrite facies, a biosparite facies, and a dolomitized facies.

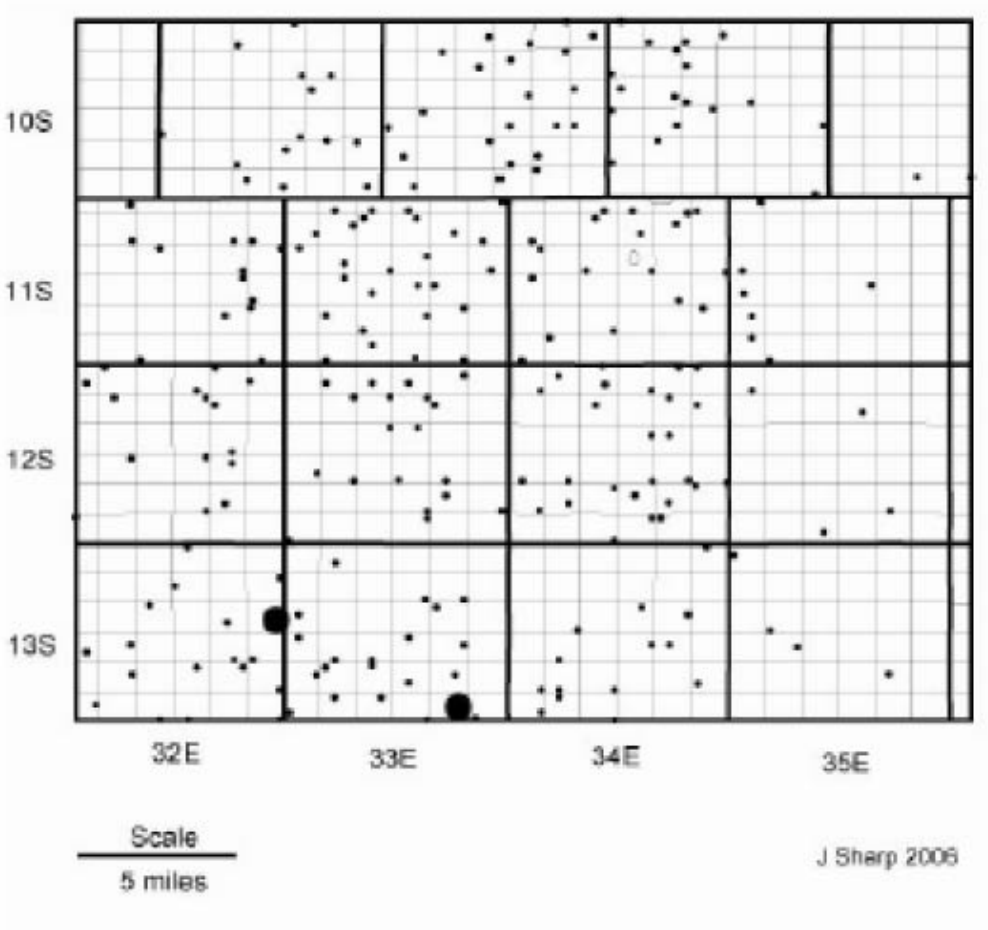


Fig. 3.33. Wells used in mapping and analyzing the detailed study area within the Bough intrashelf project area. Wells with cores are shown as larger symbols.

The *biomicrite facies* (Fig. 3.34) is the main reservoir facies and consists of phylloid algae, echinoderms, brachiopods, fusulinids, bivalves and bryozoans in a micrite matrix. According to the Dunham classification of carbonate rocks, most of the specimens obtained from the biomicrite facies are wackestones; a lesser number of samples are packstones. Porosity is formed mostly by the dissolution of the more mineralogically unstable fossils, especially phylloid algae. Pores are dominantly moldic with a subsidiary amount of solution-enlarged vugs as well as a small amount of fracture and channel porosity. This secondary porosity is preserved where it has not been infilled with sparry calcite cement.

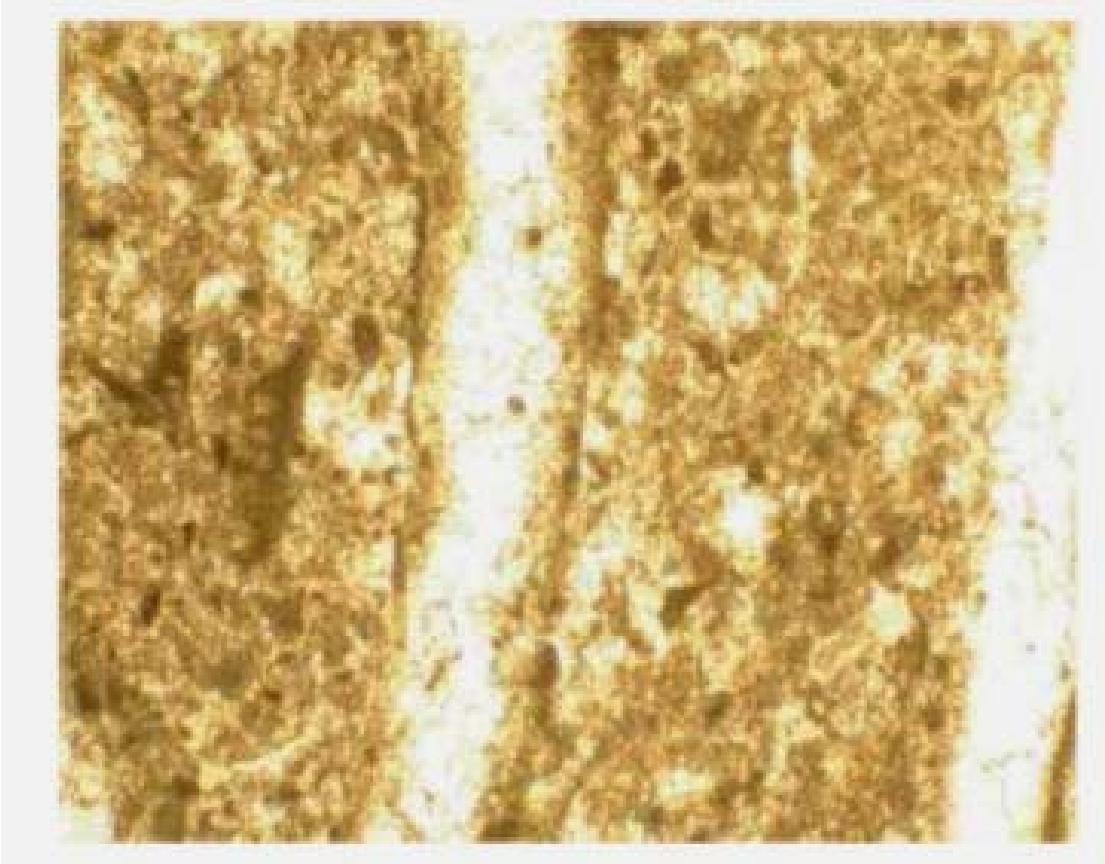


Fig. 3.34. Photomicrograph of biomicrite facies, transmitted light. Darker areas are micrite and lighter areas are calcite cement. Slide is vertically bisected by recrystallized phylloid algal blade with preserved wall structure. Field of view is 1.8 mm. From Sharp (2006).

The *biosparite facies* (Fig. 3.35) consists of phylloid algae, echinoderms, brachiopods, fusulinids, bivalves and bryozoans, but no micrite matrix. The bioclasts float in a sparry matrix. Although some porosity is present within this petrofacies, it is generally much lower than in the biomicrite facies because of occlusion by calcite spar. Porosity, where present, is dominantly moldic. Porosity within this facies attains a maximum value of approximately 30%.

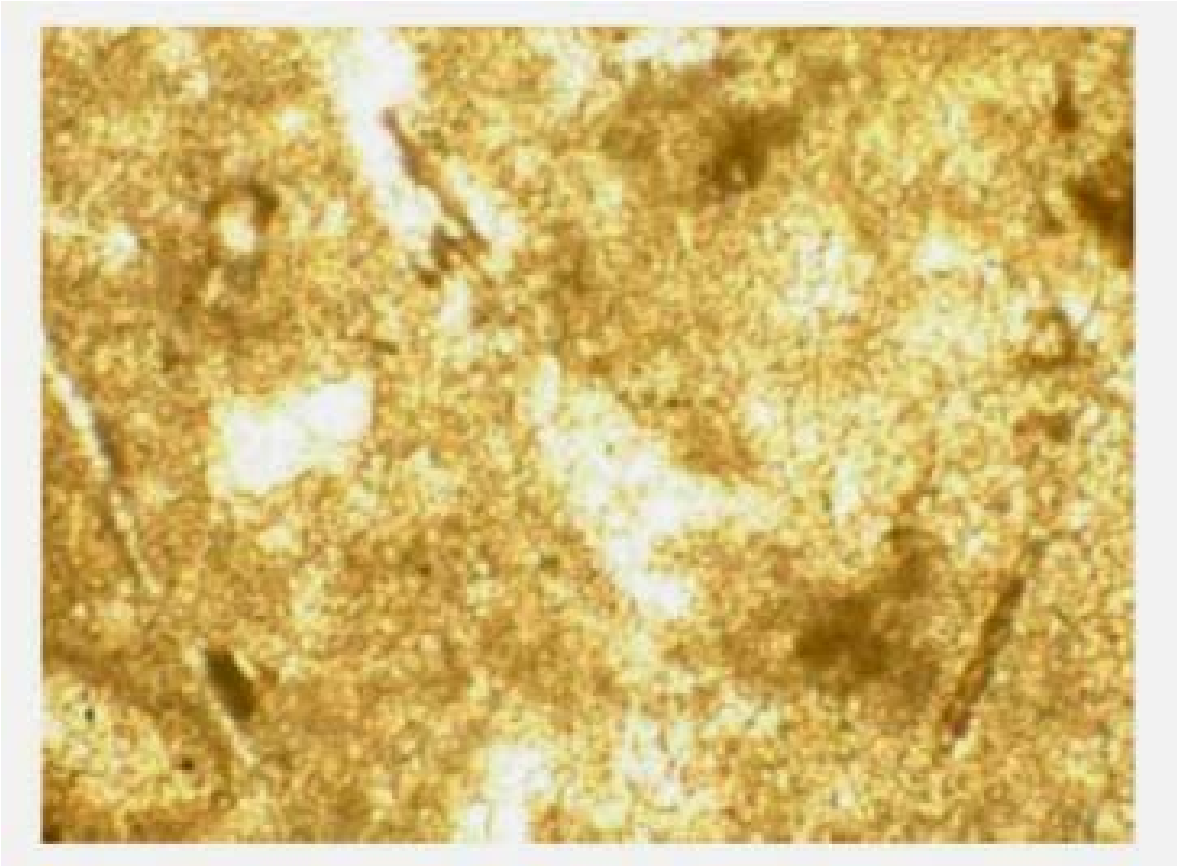


Fig. 3.35. Photomicrograph of biosparite facies, transmitted light. Darker areas are micrite and lighter areas are sparry calcite cement. Recrystallized phylloid algal blade is in center of slide. Field of view is 0.9 mm. From Sharp (2006).

The *dolomitized facies* (Fig. 3.36) is not as common as either the biomicrite facies or the biosparite facies. The dolomitization in this facies has obliterated depositional textures. It consists mostly of microcrystalline planar-e dolomite matrix; minor spherical crystals of dolomite are contained in the matrix. It is thought that these larger spheres nay be either relict ooids or burrows. Some microcrystalline calcite is present within samples from this facies. Porosity within this facies is intercrystalline and has a maximum value of approximately 20%.

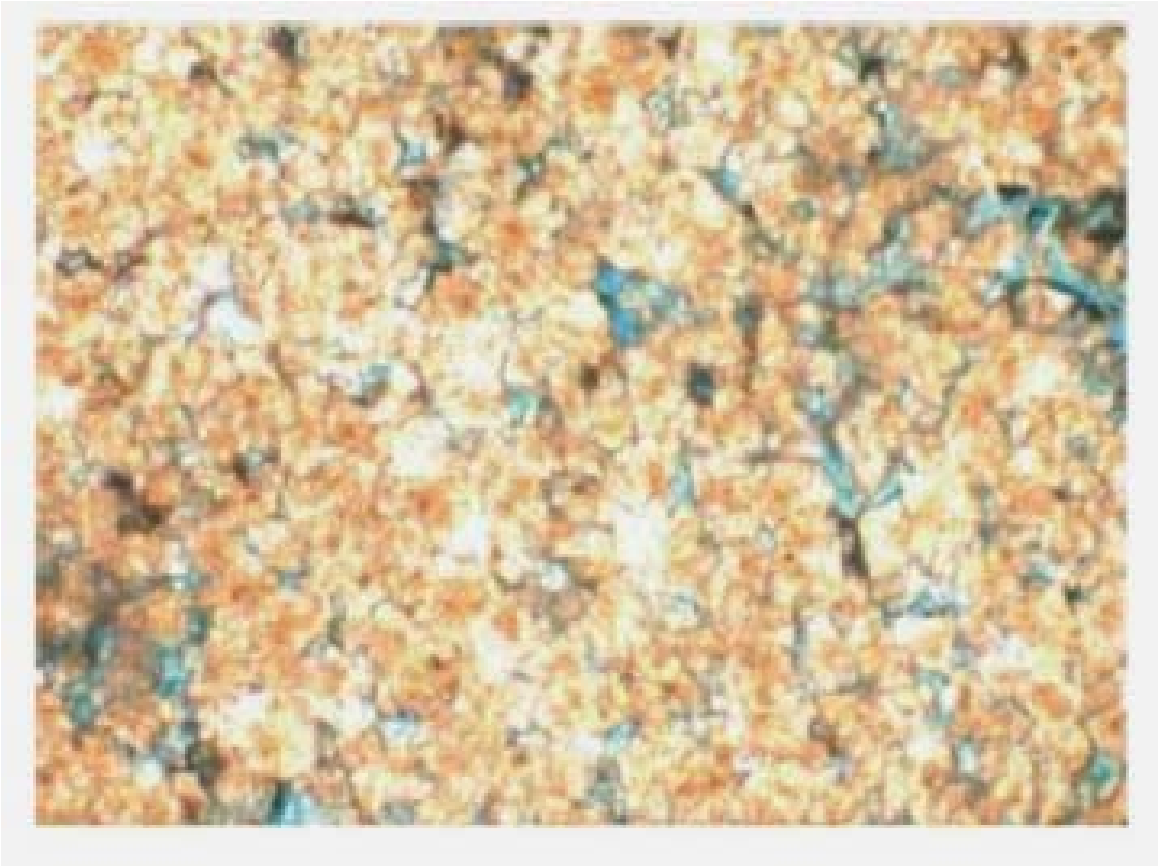


Fig. 3.36. Photomicrograph of dolomitized facies, transmitted light. Lighter areas are subhedral dolomite crystals and darker areas are crystal faces. The blue areas are inter-crystalline pores filled with blue epoxy. Field of view is 0.9 mm. From Sharp (2006).

Porosity within the Bough carbonate reservoirs is secondary. The main porosity type present is moldic, although solution-enlarged vugs, channel and fracture porosity are also present. Reservoir development is diagenetic and is related to dissolution of primary components. In some cases, a relatively late-stage cement has acted to occlude this secondary porosity.

Dagger Draw North Shelf-Margin Project Area

Introduction

Geologic investigation into a second project area was initiated during 2006. This area, referred to as the *shelf-margin project area* (Fig. 3.2), is located at the margin of the Late Pennsylvanian shelf where Upper Pennsylvanian phylloid algal mound reservoirs similar to those described above in the Bough area are prolifically productive of oil and associated natural gas. However, this project area was chosen because there are distinct differences between it and the Bough area. In the Bough area, the phylloid algal mound complexes grew on Late Pennsylvanian paleostructures present on the Northwest Shelf of the Permian Basin. In the shelf-margin project area, however, the phylloid algal mound complexes that form the oil reservoirs grew on a constructional shelf margin that separated the Northwest Shelf (on the northwest) from the deep Delaware Basin (to the southeast; see Cox et al., 1998; Speer, 1993). As will be discussed below, the difference in depositional setting of the phylloid algal mounds in each project area (growth on top of existing paleostructures in the Bough area vs. constructional shelf-margin buildup in the shelf-edge Dagger Draw area) mandates that different geological procedures and thought processes need to be used for exploration in each of the two types of areas, despite the similar biologic and genetic origins for the phylloid algal mound reservoirs in each area. Whereas it is relatively straightforward to map paleostructures in the Bough area, a lack of correlatable marker beds at the Abo level and perhaps a lack of paleostructures indicate that paleostructures cannot be mapped in the shelf-margin area. Instead, geologic procedures used to identify paleobathymetric relief, rather than paleostructural relief, must be employed to identify and delineate productive fairways,

During the 2006 project year, work began and was completed on collecting data, correlating relevant stratigraphic units, identifying productive trends, and producing appropriate computer-generated maps. As with the Bough intrashelf area, geologic contour maps were produced with *Surfer 8* (a product of Golden Software, Inc.). Geologic and well data obtained during the 2006 project year are presented in the attached Excel spreadsheet ***shelf margin area.xls***.

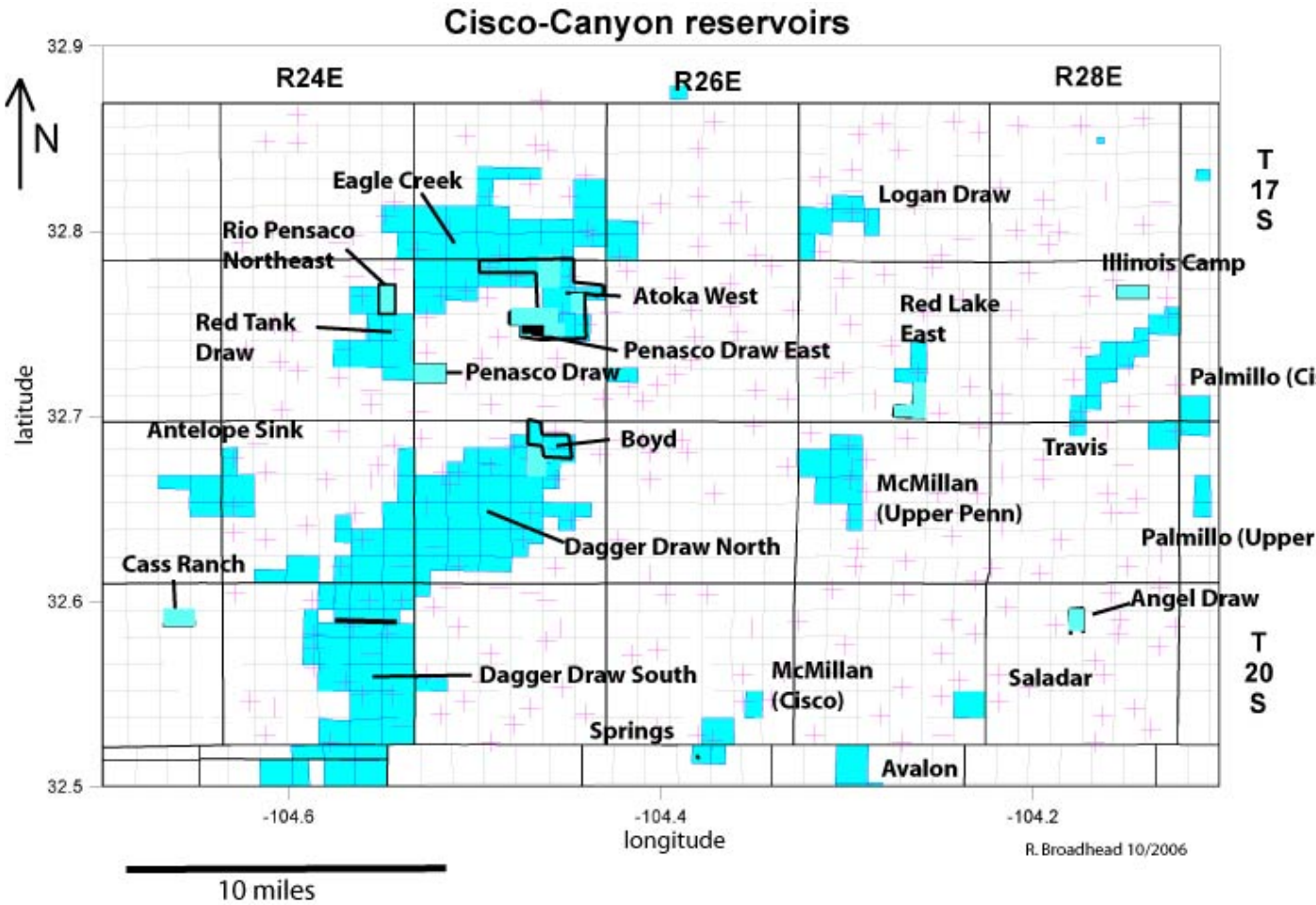


Fig. 3.37. Oil and gas reservoirs that are productive principally from Upper Pennsylvanian strata within the shelf-margin project area. See Table 3.3 for reservoir data.

Twenty-four oil and gas reservoirs are productive from Upper Pennsylvanian strata within the shelf-margin project area (Fig. 3.37, Table 3.3). These reservoirs had produced, cumulatively, 83 million bbls oil and 492 billion ft^3 gas at the end of 2003. During 2003, production from the 24 reservoirs was 667 thousand bbls oil and 12 billion ft^3 natural gas. The largest reservoirs are Dagger Draw North and Dagger Draw South, which have produced a combined cumulative of 78.5 million bbls oil, or 95% of the cumulative production for all 24 reservoirs. Dagger Draw North and Dagger Draw South have also produced a combined cumulative of 389 billion ft^3 gas, or 79% of the cumulative production for all 24 reservoirs. The 24 reservoirs contained a combined total of 504 oil and gas wells during 2003. Depth to production ranges from 6150 ft at Antelope Sink to 10020 ft at Angel Draw and Saladar.

Table 3.3 Permo-Pennsylvanian Carbonate Reservoirs Productive within the Shelf Margin Project Area (See Fig. 3.37 for Locations of Reservoirs)

Reservoir name	Productive stratigraphic unit	Annual oil production 2003 (thousand bbls)	Cumulative oil production 2003 (million bbls)
Angel Draw	(Upper Pennsylvanian)	5	0.005
Antelope Sink	(Upper Pennsylvanian)	0.5	0.024
Atoka West	(Upper Pennsylvanian)	0.4	0.006
Avalon	(Upper Pennsylvanian)	0.1	0.015
Boyd	(Cisco)	0	0.001
Cass Ranch	(Upper Pennsylvanian)	0	0
Dagger Draw North	(Upper Pennsylvanian)	476	53.8
Dagger Draw South	(Upper Pennsylvanian)	124	24.7
Eagle Creek	(Permo-Pennsylvanian)	3	0.1
Illinois Camp	(Cisco)	0	0.0005
Logan Draw	(Cisco-Canyon)	0.01	0.03
McMillan	(Cisco)	0	0
McMillan	(Upper Pennsylvanian)	1	0.425
Palmillo	(Cisco)	0	0.001
Palmillo	(Upper Pennsylvanian)	0	0.03
Penasco Draw	(Upper Pennsylvanian)	10	0.5
Penasco Draw	(Permo-Pennsylvanian)	1	0.04
Penasco Draw East	(Upper Pennsylvanian)	9	0.03
Red Lake East	(Upper Pennsylvanian)	0	0.14816
Red Tank Draw	(Permo-Pennsylvanian)	0.03	0.0002
Rio Penasco Northeast	(Upper Pennsylvanian)	0	0
Saladar	(Upper Pennsylvanian)	0	0.001
Springs	(Upper Pennsylvanian)	3	0.7
Travis	(Upper Pennsylvanian)	34	2.1

The importance of the reservoirs within the Dagger Draw area cannot be overstated from either a production standpoint or from an exploration standpoint. The Dagger Draw reservoirs were discovered in 1964, but remained sparsely drilled and underdeveloped until 1990 as the reservoir and trapping mechanism were incompletely understood. In the late 1980's, however, additional exploration and development revealed that this seemingly small oil field was substantially larger than had originally been thought. As exploration and development revealed the true nature of the reservoir and trap, production from the field increased from a few thousand bbls oil per year to almost 10 million bbls oil per year (Fig. 3.38; Broadhead, 1999a). This phase of redevelopment led to exploitation of the full reservoir (Fig. 3.39) and what had been viewed as a small and relatively insignificant oil field became a major productive area that contributed slightly more than 13% of all oil production in New Mexico during 1996.

Reservoirs are mostly productive from Canyon strata, although some are productive from the Cisco (see Fig. 3.3). In Table 3, most reservoirs identified as being productive from the Upper Pennsylvanian have the Canyon as the primary productive interval. Reservoirs identified as Permo-Pennsylvanian are productive from uppermost Pennsylvanian strata and in some cases lowermost Permian strata as well. As discussed above, reservoirs are productive from complexes of phylloid algal mud mounds. The primary reservoirs in the project area, Dagger Draw North and Dagger Draw South, are formed by complexes of phylloid algal mounds that were located on the Late Pennsylvanian shelf margin (Fig. 3.40). The reservoirs strata have been dolomitized. Porosity is vugular (Speer, 1993; Cox et al., 1998). Updip seals are formed by impervious back-reef limestones (Fig. 3.41). Dark basinal shales that are adjacent to and interfinger with the dolostone reservoirs on the east (basinward) side are the apparent source rocks.

Exploration methodology for the Dagger Draw and Bough areas will differ because of the contrasting geologic setting of the phylloid algal mound reservoirs in each area. In the Bough area, the phylloid algal mounds grew on emerging Pennsylvanian-age structures on the interior of a shallow-marine shelf. The emerging structures localized growth of the phylloid algal mounds. Therefore, Bough reservoirs are directly related to paleostructures and off-structure areas generally are barren of reservoirs, as described

previously. Mapping of paleostructures will therefore be of primary importance in exploration.

In the Dagger Draw shelf-margin area, however, growth of phylloid algal mounds was primarily localized by paleobathymetric position on the shelf edge, which was constructional rather than structural in origin. Therefore, paleostructure mapping will not be helpful in the location of (exploring for) reservoirs, and other geologic techniques need to be employed.

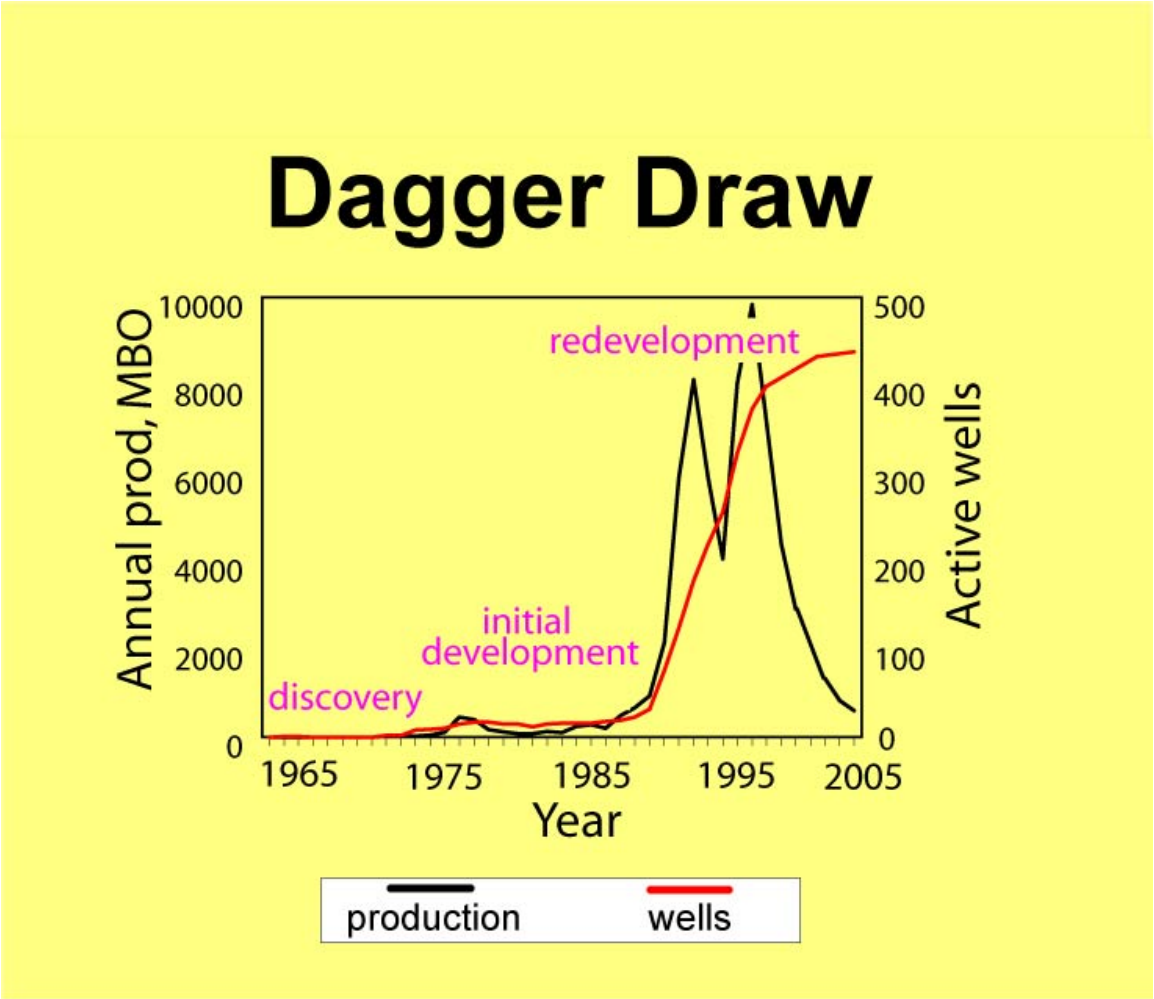


Fig. 3.38. Production history curve of the Dagger Draw reservoirs, showing periods of discovery and initial development of a seemingly small reservoir, and final full redevelopment that revealed the true size of this very large oil reservoir. After Broadhead (1999a).

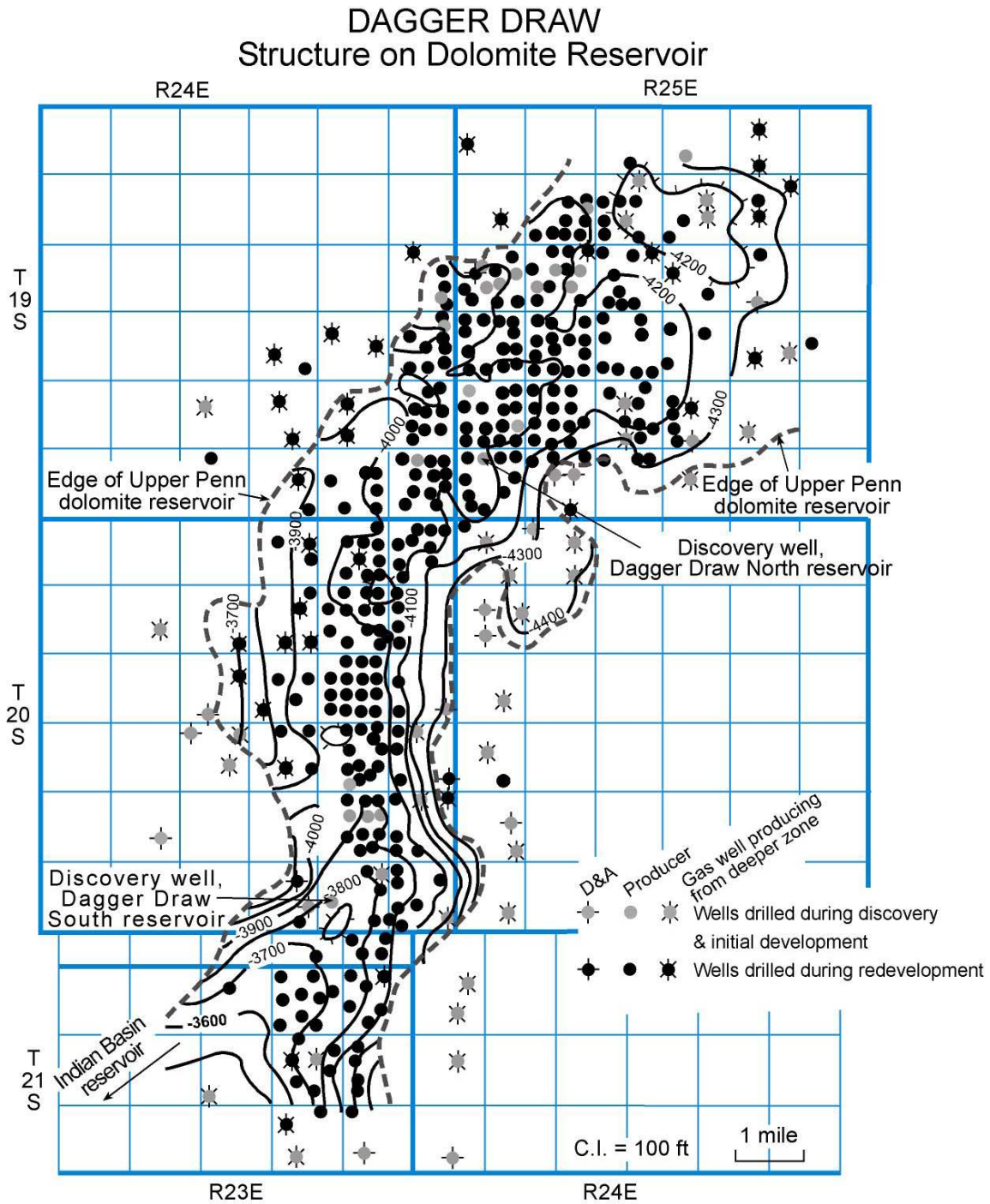


Fig. 3.39. Structure contour map on top of productive dolostone shelf-margin carbonate buildup facies, Dagger Draw and Dagger Draw South reservoirs indicating wells drilled during phases of (1) discovery and initial development, and (2) redevelopment. Structure contours from Reddy (1995).

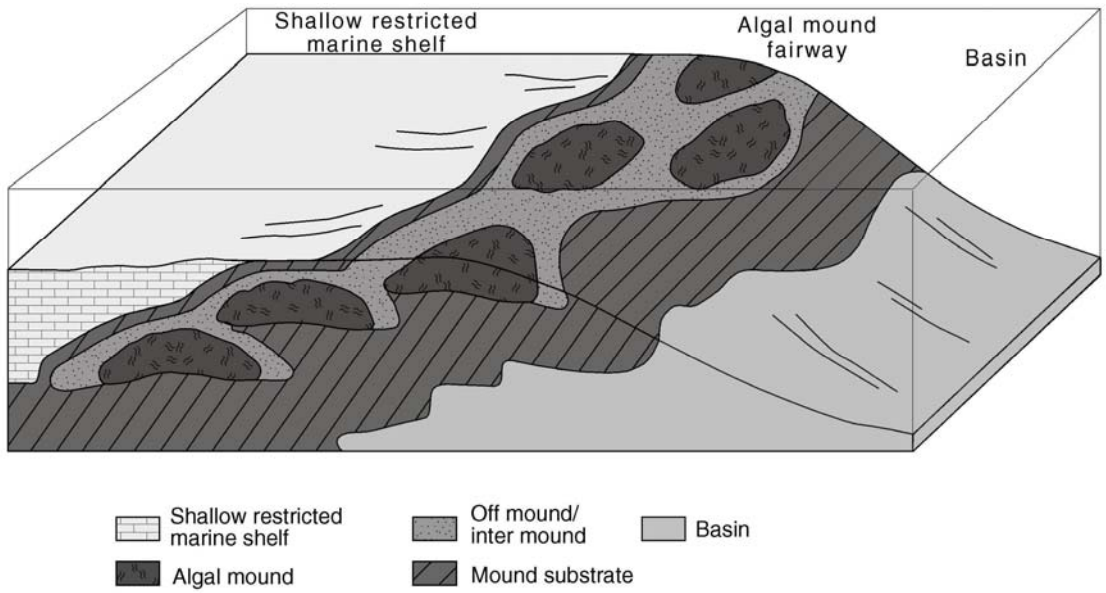


Fig. 3.40. Depositional model of shelf-margin phylloid algal mounds and intramound facies at shelf margin, Dagger Draw reservoirs. From Cox et al. (1998).

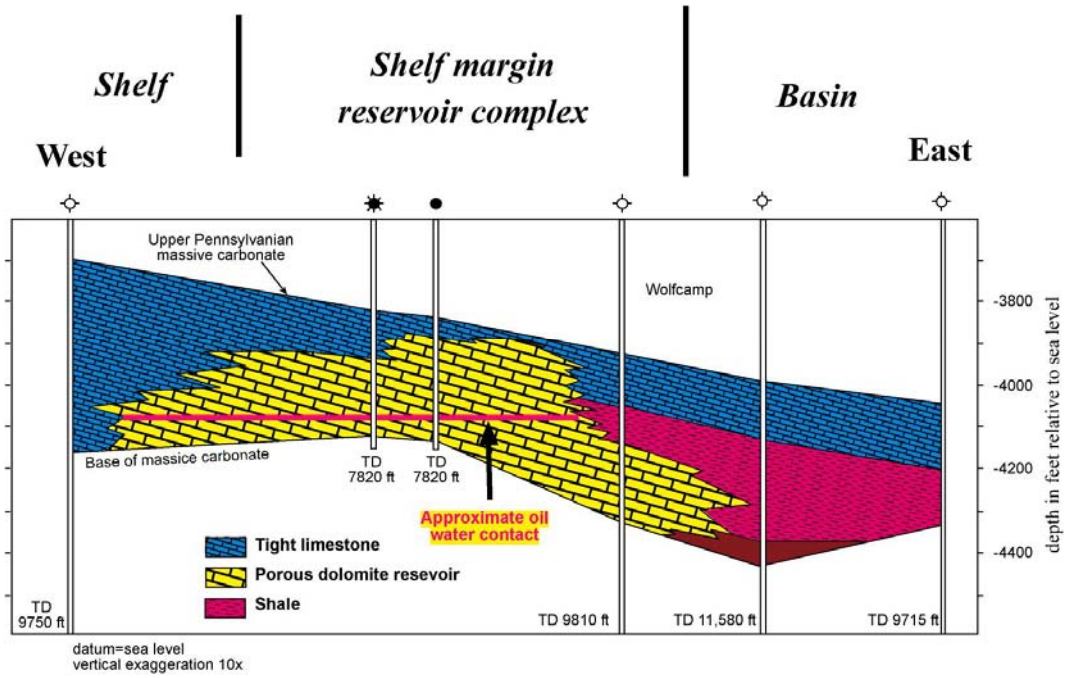


Fig. 3.41. West-east structural cross section though Upper Pennsylvanian strata in Dagger Draw South reservoir showing morphology of dolomitized shelf-margin reservoir facies, impermeable backreef limestones, and basinal shales. From Speer (1993).

Data Acquisition and Mapping

During 2006, 312 wells were mapped and correlated within the shelf margin project area (Fig. 3.42). Because the area straddles the boundary between the shallow Northwest Shelf and the deep Delaware Basin, few stratigraphic units are directly correlatable from shelf to basin. On the shelf (Fig. 3.43), the shelfal Abo carbonates directly overlie the Wolfcamp carbonates whereas in the basin (Fig. 3.44), the basinal Bone Spring Formation directly overlies the Wolfcamp section. Within the Dagger Draw reservoir along the shelf margin, the well logs reflect yet a different lithologic facies of the shelf-edge phylloid algal mound complexes (Fig. 3.45). Therefore, unlike in the Bough area, the Abo, which is not present in the basin, cannot be mapped across the entire project area and is therefore of limited use for either structure mapping or for stratigraphic mapping. Even the top of the Cisco Series and the Canyon Series cannot be reliably correlated across the entire project area (although they can be correlated reliably within the Dagger Draw North and Dagger Draw South reservoirs, see Reddy, 1995) and different operators make grossly different picks for the tops of these units. One lithologic unit that can apparently be correlated across the entire project area is a prominent shale (called the upper Wolfcamp shale in this report) in the upper part of the Wolfcamp. This unit, along with the top of the Mississippian Series, was correlated in all 312 wells.

A structure contour map on top of the Mississippian (Fig. 3.46) reveals no discrete break between the Northwest Shelf and the Delaware Basin. Instead, the transition from the shelf to the basin is marked by a uniform slope at the Mississippian level. Only perhaps a few subtle local structures are superimposed on the regional southeast slope.

A structure contour map on top of the upper Wolfcamp shale (Fig. 3.47) indicates a sharp break between the Northwest Shelf and the Delaware Basin at the Wolfcamp level. Because this transition from shelf to basin is evident at the upper Wolfcamp shale level but not at the deeper Mississippian level, it was concluded that the relief indicated on the upper Wolfcamp shale structure map (Fig. 3.47) is constructional rather than structural in origin. Furthermore, an isopach map of the stratigraphic interval between the top of the upper Wolfcamp shale and the top of the Mississippian (Fig. 3.48) does not reveal a thinning over the shelf margin as the Abo-Mississippian isopach map revealed a thin-

ning over paleostructures in the Bough area, but instead reveals a thickening at the shelf margin. This map apparently reflects drape of the Wolfcamp shale over the thickened Upper Pennsylvanian shelf-margin carbonate complex that grew at the constructional shelf edge. As such, this isopach map, as well as the structure map of the upper Wolfcamp shale (Fig. 3.47) reveals the location of principal areas of growth of the Upper Pennsylvanian phylloid algal mounds at the shelf margin.

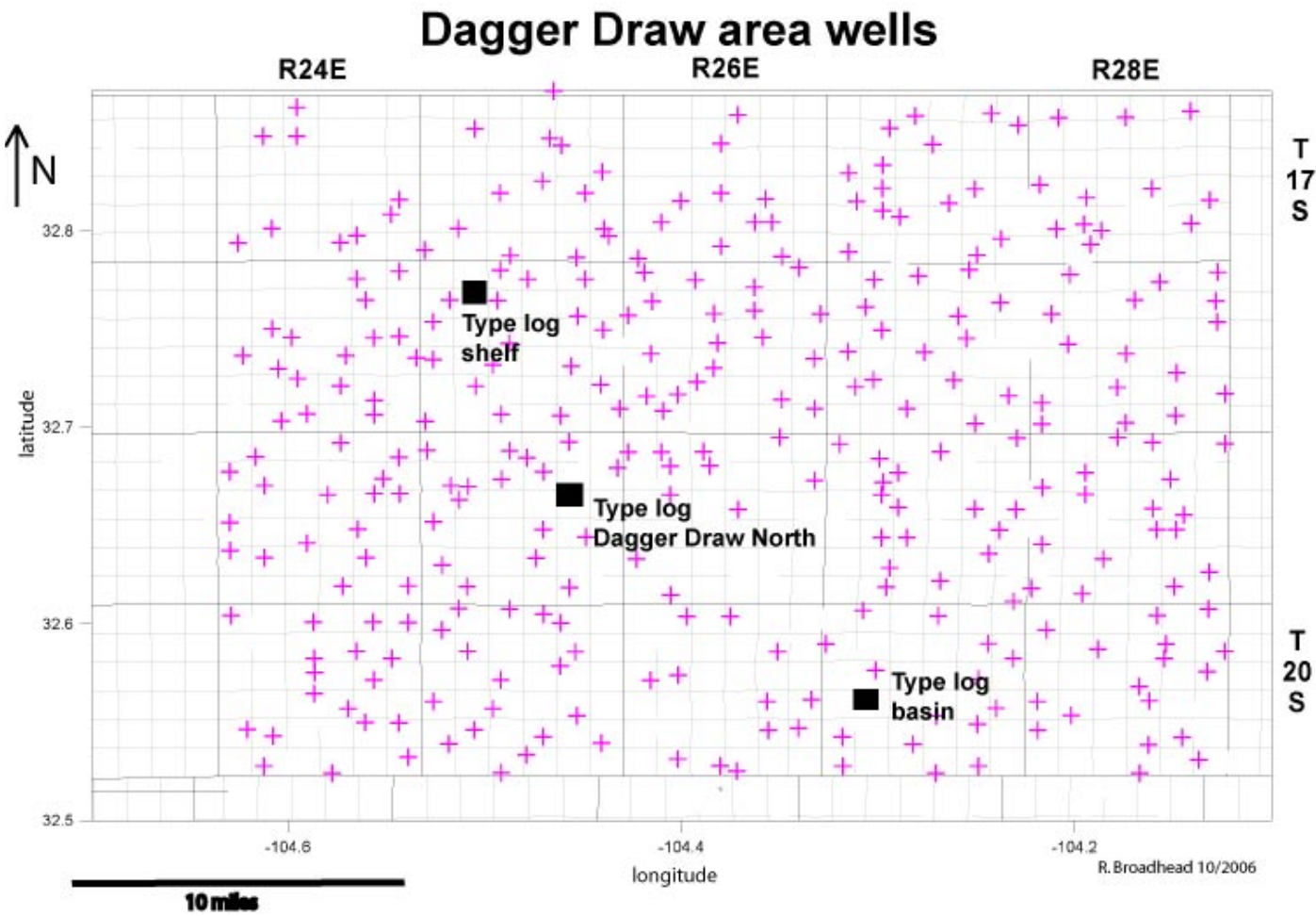


Fig. 3.42. Locations of wells analyzed for this project in the shelf-margin project area and locations of type logs for the shelf area, basin area and Dagger Draw North reservoir (Figs. 3.43–3.45).

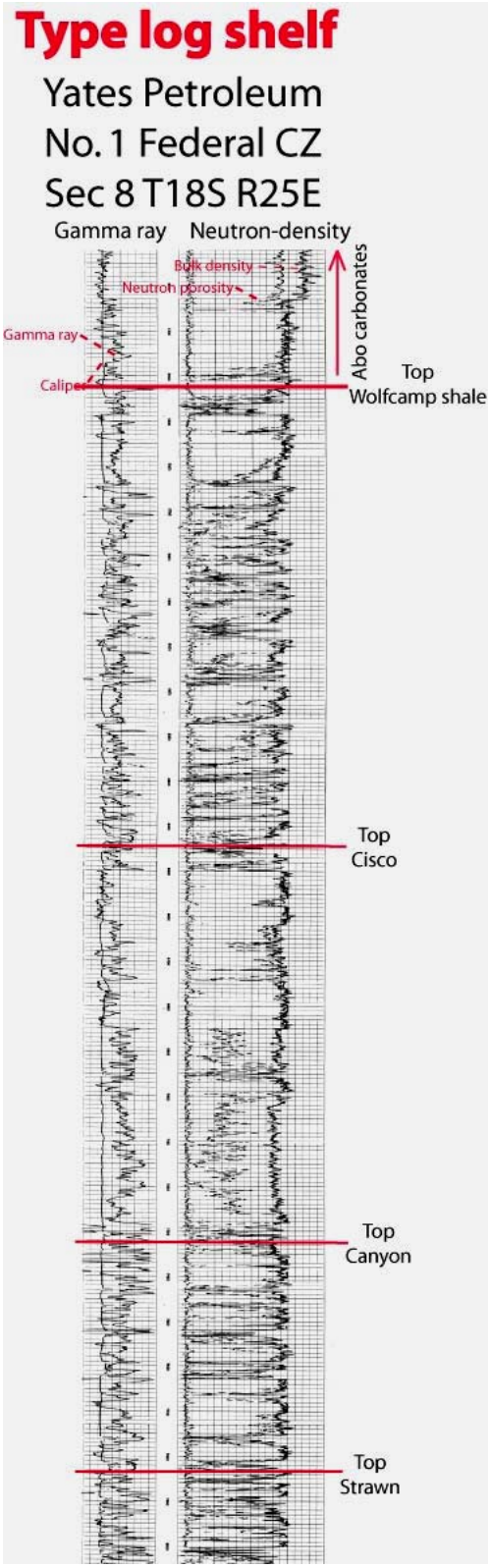


Fig. 3.43. Type log through Upper Pennsylvanian shelf facies. See Fig. 3.42 for location.

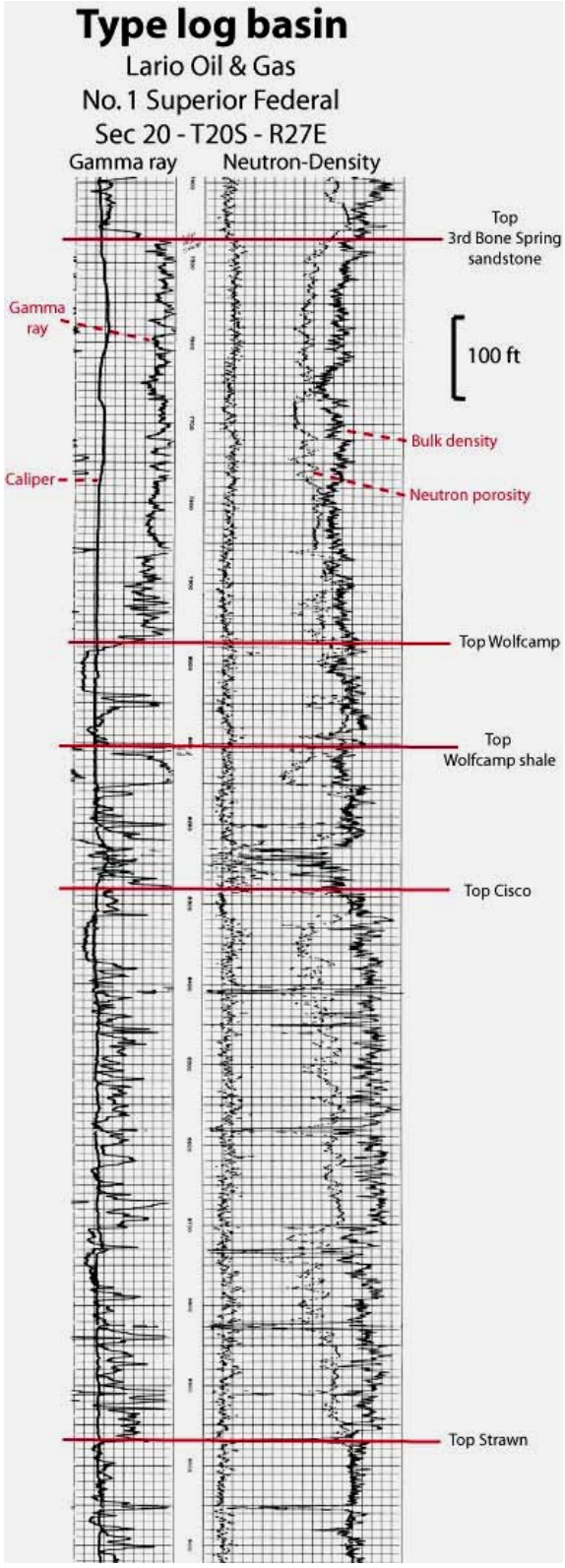


Fig. 3.44. Type log through Upper Pennsylvanian basinal facies. See Fig. 3.42 for location.

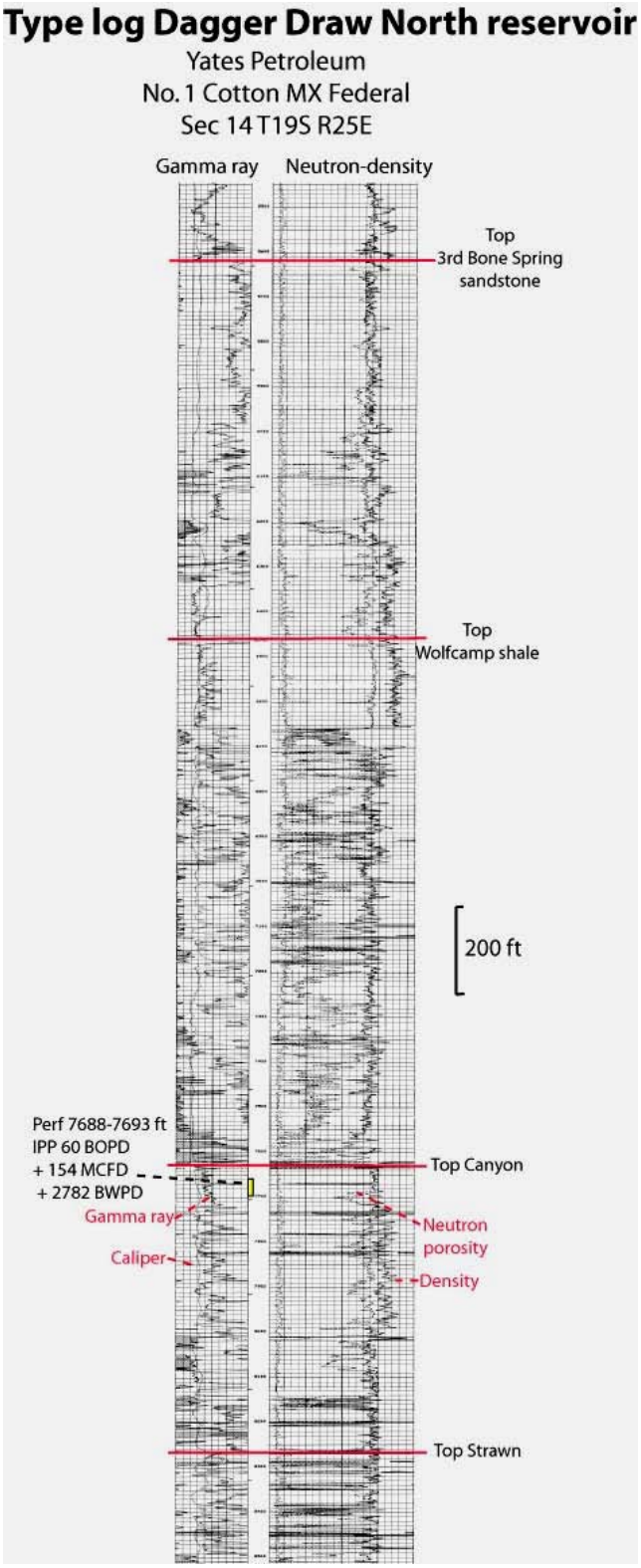


Fig. 3.45. Type log through Upper Pennsylvanian shelf-margin facies, Dagger Draw North reservoir, indicating productive zone. See Fig. 3.42 for location.

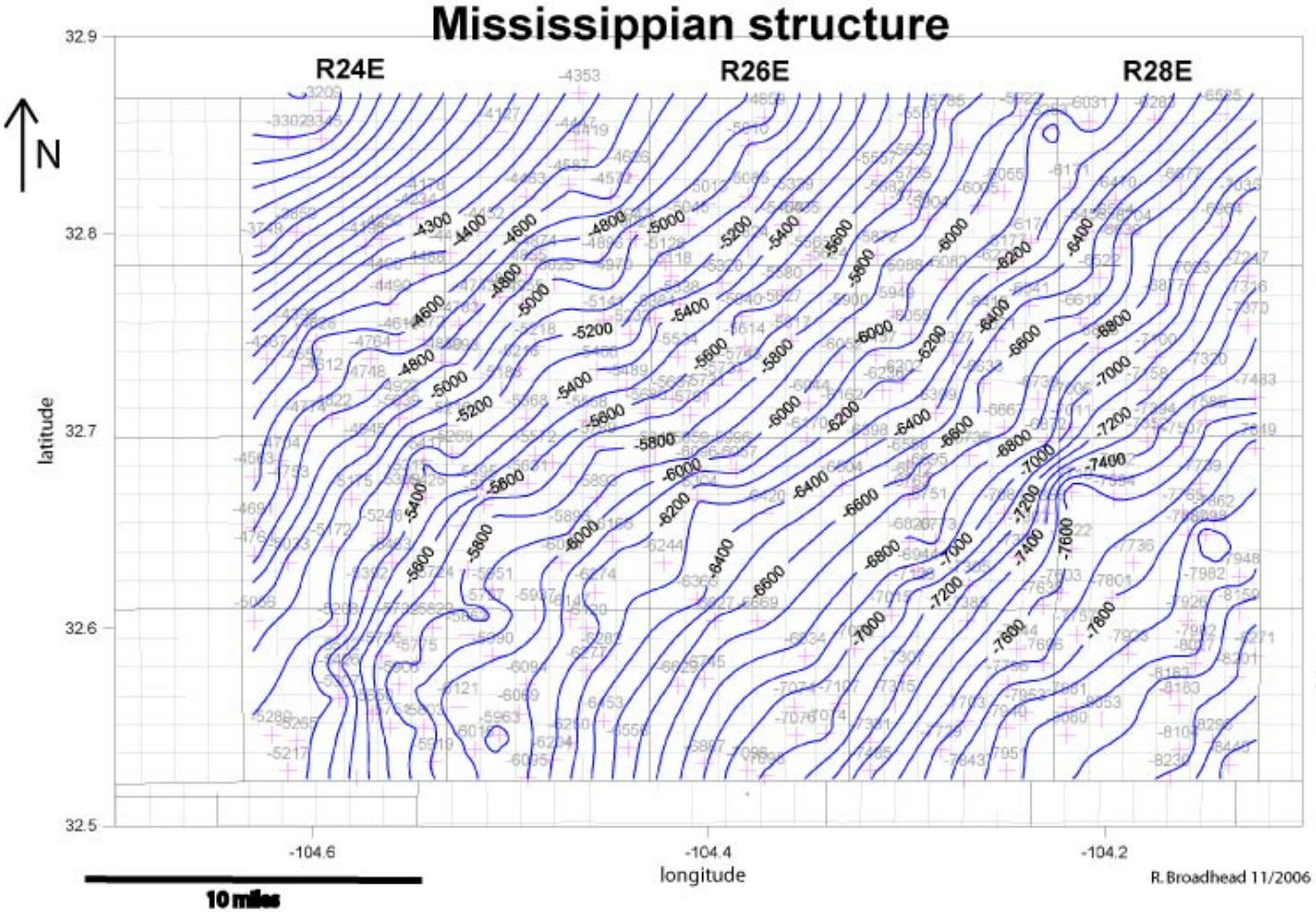


Fig. 3.46. Structure contours on top of Mississippian strata, shelf-margin project area. Contours in feet. Datum = sea level.

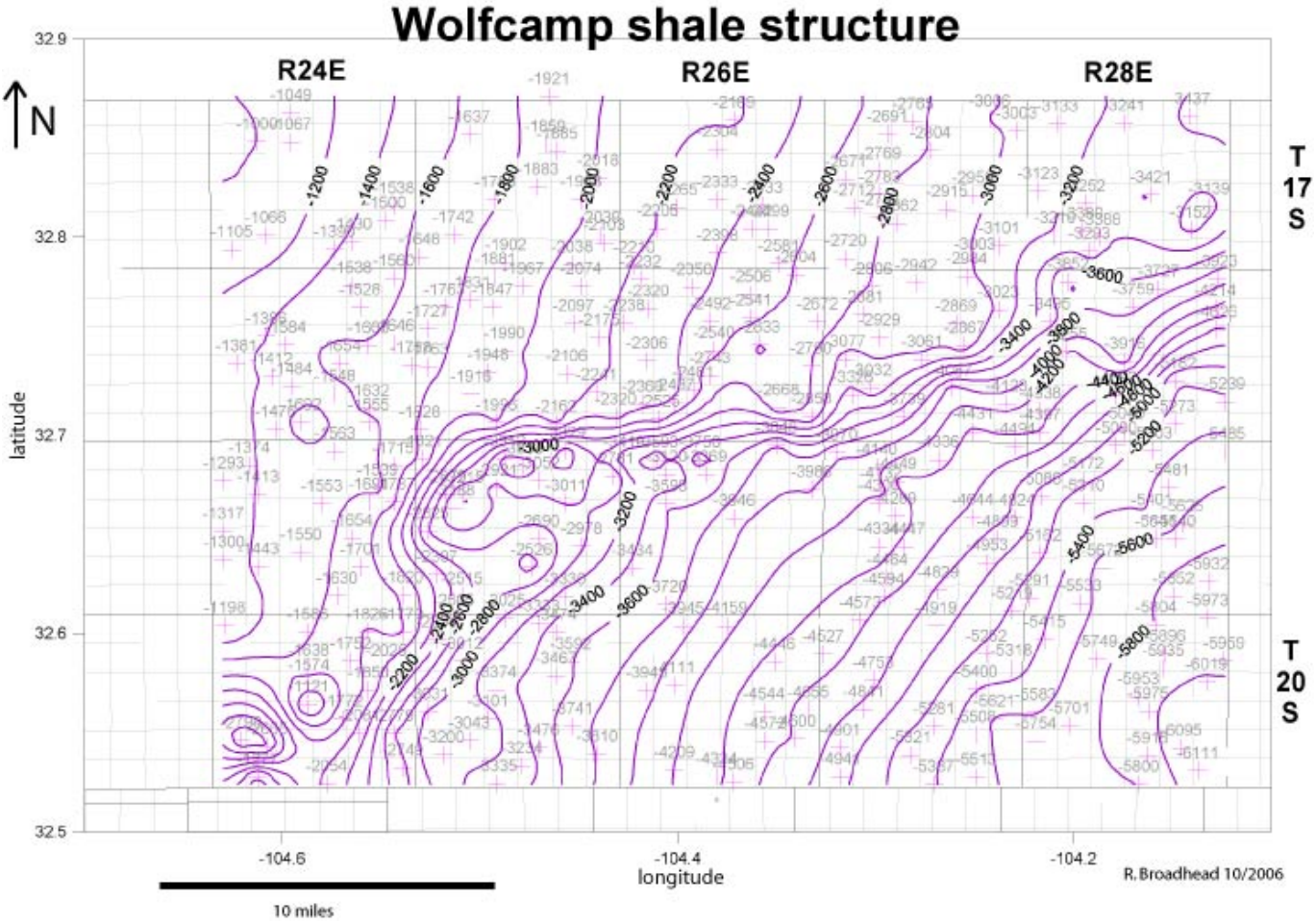


Fig. 3.47. Structure contours on top of upper Wolfcamp shale, shelf-margin project area. Contours in feet. Datum = sea level.

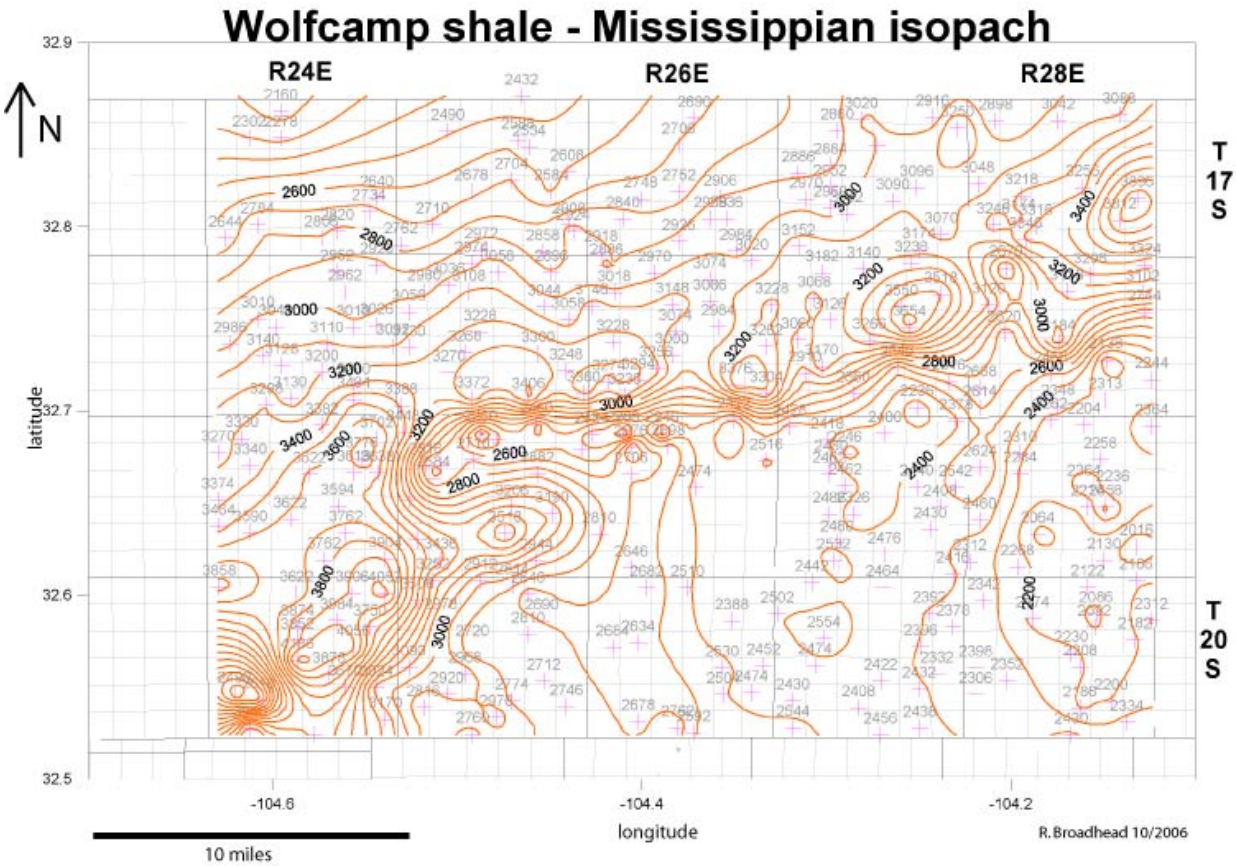


Fig. 3.48. Isopach map of stratigraphic interval between the top of the upper Wolfcamp shale marker and the top of the Mississippian System, shelf-margin project area.

Relationship of Geology to Oil and Gas Reservoirs

Superposition of the map of the Cisco and Canyon reservoirs on the Mississippian structure (Fig. 3.49) reveals no obvious correlation between the location of reservoirs and the structure. Although the larger reservoirs form elongate trends that parallel the Mississippian structure contours, there is no apparent distinction among contours that can be used as a predictive tool to determine which of the Mississippian contours might be parallel to the reservoir trends in the overlying Upper Pennsylvanian strata. However, when the map of Cisco and Canyon reservoirs are superimposed on the structure contour map of the upper Wolfcamp shale (Fig. 3.50) as well as the isopach map of the interval between the top of the Wolfcamp Shale and the top of the Mississippian System (Fig. 3.51),

an obvious correlation between the location of known productive reservoirs and structure emerges. Of course, as discussed above, the Wolfcamp shale structure is not tectonic, but instead reflects drape of the upper Wolfcamp shale over the shelf-margin carbonate complex. The primary Canyon reservoirs, Dagger Draw North and Dagger Draw South, are located along the shelf edge in the southwest part of the map. Other smaller reservoirs, Travis and Red Lake East, lie along this trend of closely spaced contours to the northeast of Dagger Draw North and Dagger Draw South.

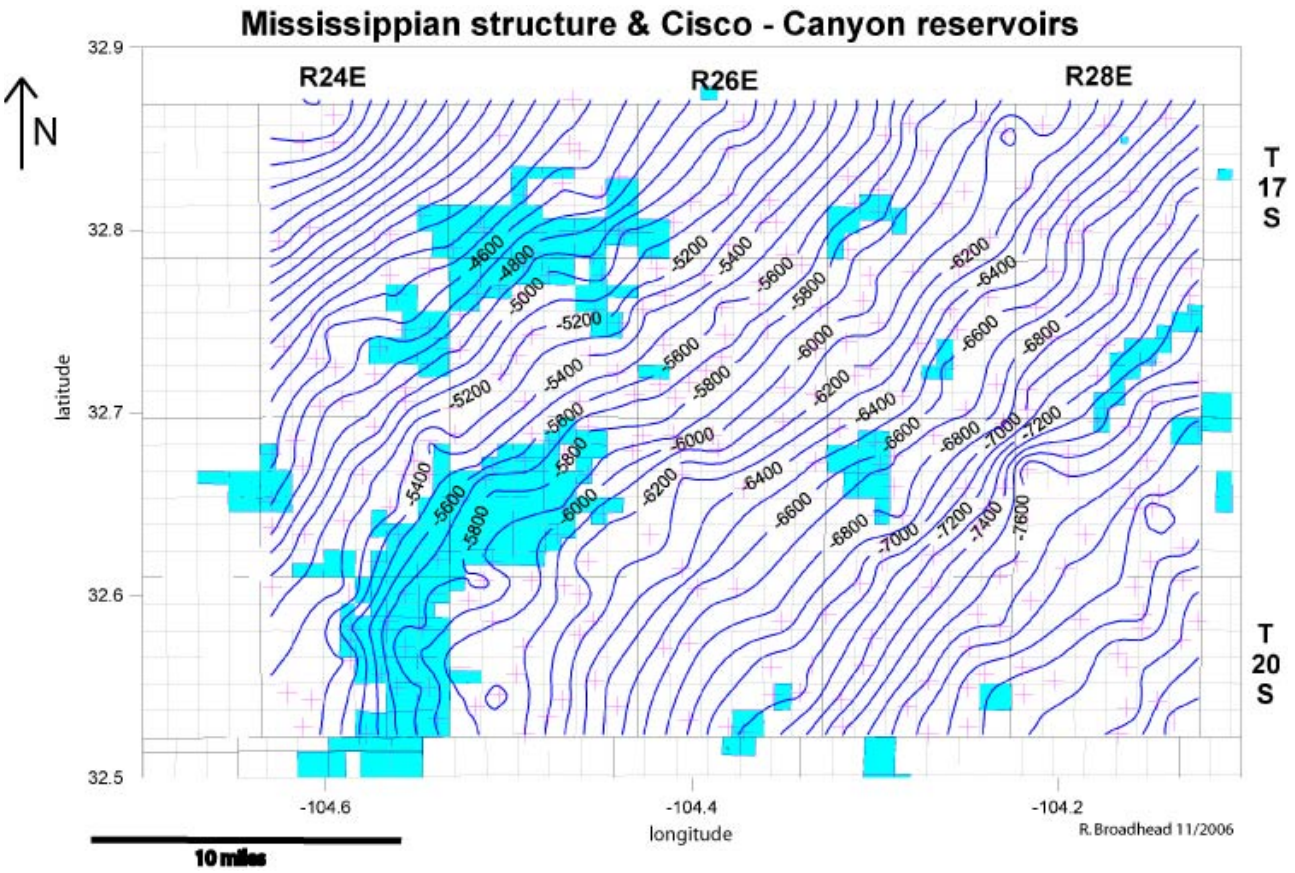


Fig. 3.49. Upper Pennsylvanian (Cisco and Canyon) reservoirs and Mississippian structure, shelf-margin project area.

The trend of the shelf-margin reservoirs is readily apparent on Figs. 3.50 and 3.51. The minor reservoirs south of the shelf margin may be formed by carbonate debris flows and their northern boundary may be identified by the shelf-margin trend apparent on Figs.

3.50 and 3.51. Reservoirs north of the shelf margin (e.g., Eagle Creek) appear to reside in somewhat younger strata than those of the shelf margin (e.g., Dagger Draw North) and are not resolved by maps prepared for this project.

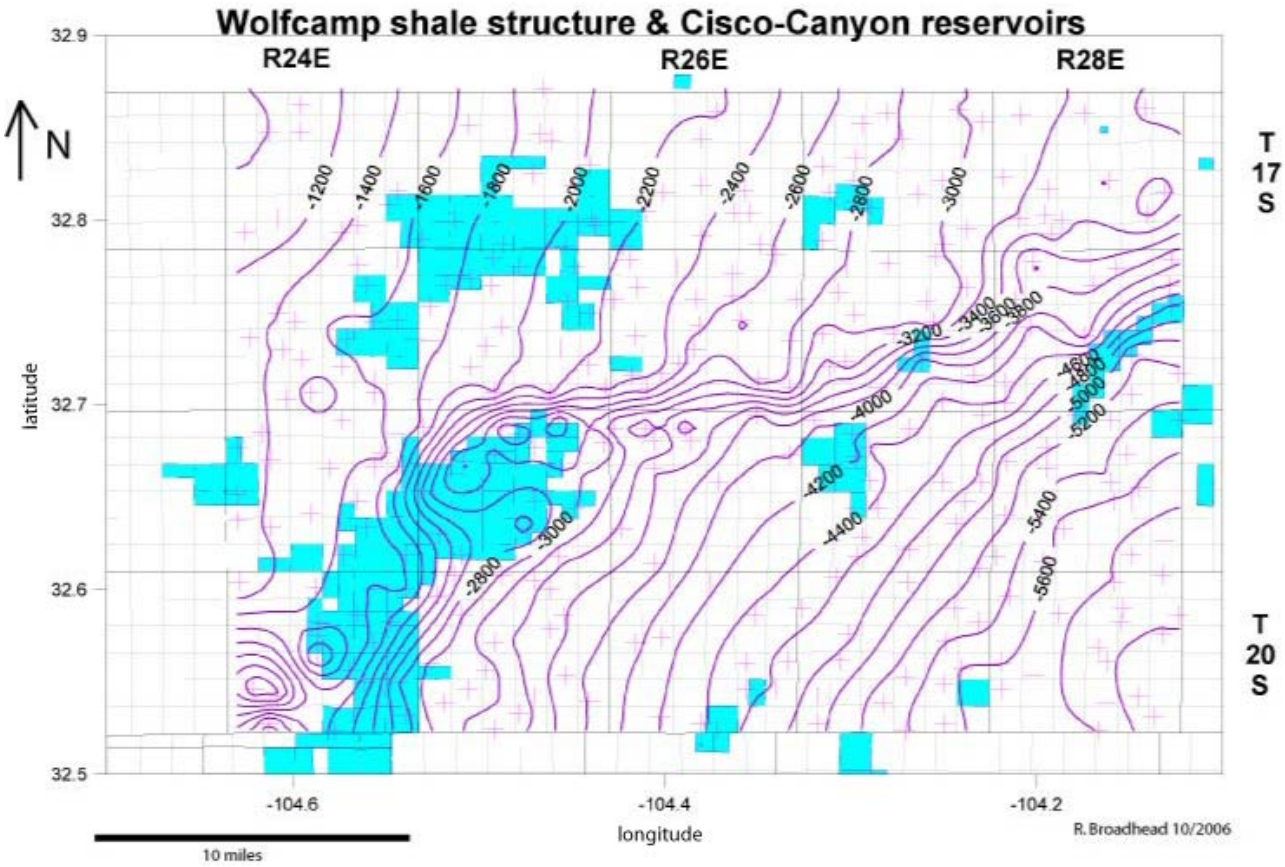


Fig. 3.50. Upper Pennsylvanian (Cisco and Canyon) reservoirs and upper Wolfcamp shale structure, shelf-margin project area.

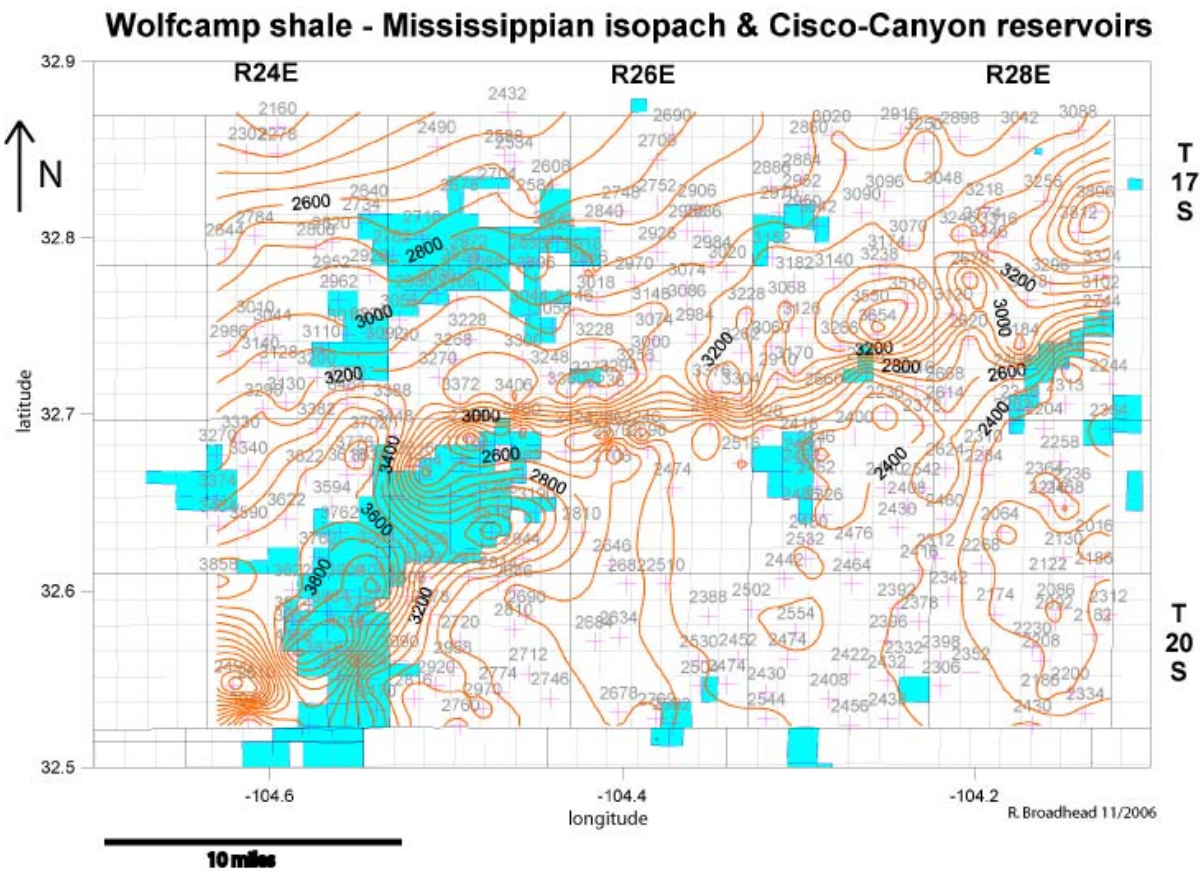


Fig. 3.51. Upper Pennsylvanian (Cisco and Canyon) reservoirs and isopach map of stratigraphic interval between the top of the upper Wolfcamp shale marker and the top of the Mississippian System, shelf-margin project area.

4. Experimental Methods

Laboratory experimentation is not a component of this project.

5. Technology Transfer

An integral component of this project is an active and vigorous technology transfer program. The Southwest section of the Petroleum Technology Transfer Council (PTTC) is

housed in the same building as the PRRC and tech transfer is a fundamental component of all research performed at the Center. For the CFS project, a concerted effort has been made to constantly keep industry updated on developments through talks, presentations, and expansion of the FEE Tool consortium. The project actively transfers technology on both regional and national levels; in the first two years, this included the papers and presentations listed below:

1. Balch, R. S., Ruan, T. and Schrader, S.M.: "Fuzzy Expert Systems in Oil Exploration," SIAM Conference on Computational Science and Engineering, Orlando, FL, Feb. 12-15, 2005.

2. Schrader, S.M., Balch, R.S., Ruan, T.: "Using Neural Networks to Estimate Monthly Production: A Case Study for the Devonian Carbonates, Southeast New Mexico," paper SPE 94089, presented at the 2005 SPE Production and Operations Symposium, Oklahoma City, April 17-19.

3. Ruan, T., Balch, R.S., Schrader, S.M. and Hart, D.M.: "A Web-based Fuzzy Ranking System and Application," presented at the 9th World Multiconference on Systemics, Cybernetics and Informatics, Orlando, July 10-13, 2005
4. Ruan, T., Balch, R.S., Schrader, S.M.: "A Fuzzy Expert System for Oil Prospecting in the Lower Brushy Canyon of Southeast New Mexico," presented at the 2005 IEEE International Conference on Information Reuse and Integration, Las Vegas, August 15.
5. Schrader, S.M., Balch, R.S., and Ruan, T.: "Knowledge Management, Collection and Storage in Expert System Development," *Upstream CIO* (September 2005) 22-24.
6. Schrader, S.M., and Balch, R.S.: "Automated Analysis of Gridded Geologic Map Data," presented at the 2006 AAPG Annual Convention, Houston, April 9 -12.
7. Balch, R.S, Schrader, S.M, and Ruan, T.: "Knowledge Engineering: Collec-

tion, Storage and Applications of Human Knowledge in Expert System Development," submitted to the Expert Systems Journal, March, 2006, revised and resubmitted, February 2007.

8. Sharp, J.: "Upper Pennsylvanian and Lower Permian phylloid algal mound reservoirs, southeastern New Mexico," in Raines, M.A., ed., Resource plays in the Permian Basin: resource to reserves: West Texas Geological Society, Publication 06-117, p. 107-113, 2006.

6. Conclusions

The project has enjoyed a successful second year, with all project goals for the second budget period met or exceeded, and with no significant obstacles identified. The project has reached the end of the primary software development stage and a functional prototype of the software has been implemented. The software is now ready for beta testing with outside users, which is scheduled for the third year of the project. The project is on task to produce expected results including deliverable software to aid and speed prospect generation using state-of-the-art computational intelligence technology. The onset of the third project year is currently delayed.

The system comprises four major components: a graphical user interface (GUI) for compiling and organizing expert knowledge, a GUI for organizing project data, a GUI for analyzing results, and a fuzzy inference engine for processing rules. The knowledge input into the system is to be expressed generally as "if-then" rules. Fuzzy rules will be employed when inputs are imprecise and factual data are insufficient or scarce. Local databases, such as the New Mexico Production Data System (ONGARD) located at the PRRC, and remote databases, such as "An Oil and Gas Development System," supported by DOE, are to be linked to the customizable system through the Internet and will allow users to search for additional data. Statistical tools and methods of interpretation have been designed to allow users to input raw data and calculate appropriate rules and analyses quickly and efficiently. This software is fully integrated and allows seamless integra-

tion of user data in the expert system design process. A new customizable fuzzy inference has been designed to provide the relational computations required to relate scarce local data with sparse regional data.

Collection and processing of test data, the Pennsylvanian Carbonates of SE New Mexico, has been completed. This test data set will aid in finalizing the work flow operations and wizards necessary for accurate and efficient expert system design. The dataset by itself represents a valuable resource for anyone interested in the play.

During the first project year geologic data acquisition and analysis was completed on the Upper Pennsylvanian and Lower Permian carbonate reservoirs within the project areas in the Permian Basin of southeastern New Mexico. During 2006, work on the Bough intrashelf project area was completed. Geologic data was acquired, mapped and analyzed on a second project area, the shelf margin project area that includes the Upper Pennsylvanian Dagger Draw oil reservoirs. The second project area was added to diversify the expert system's approach to prospecting for oil in carbonate stratigraphic traps. Although reservoirs in both areas are formed primarily by phylloid algal mounds that have seen multiple episodes of diagenesis, the reservoirs within the Bough intrashelf area grew on bathymetrically high paleostructures and the reservoirs within the shelf margin area grew primarily as bioherms on a constructional shelf margin and their geographic location does not appear to be related to paleostructures. Both of the project areas have significant production. The 58 Permo-Pennsylvanian reservoirs that are at least partially present within the Bough intrashelf project area have produced a combined total of 329 million bbls oil. The 24 Upper Pennsylvanian reservoirs within the shelf margin project area have produced a combined total of 83 million bbls oil. Data generated includes numerous maps shown in section 3 of this report.

Successful development of this customizable fuzzy expert system will bring great computational power for regional play analysis to companies of all sizes—benefiting large companies interested in recompletion opportunities but having reduced domestic exploration budgets and small companies without the resources to maintain full-time multidiscipli-

nary exploration staff. Ultimately, widespread use of this technology can increase efficiency and production from domestic oil and gas sources. Analysis of the Pennsylvanian carbonates, a significant by-passed pay play, should showcase the abilities of the software and provide immediately useful information to companies interested in developing production from that play.

REFERENCES

Abramowicz, M. and Stegun, I. (1964) Handbook of Functions, National Bureau of Standards

Apache Jakarta Project, http://jakarta.apache.org/poi

Broadhead, R.F., 1999a, "Underdeveloped Oil Fields – Upper Pennsylvanian and Lower Wolfcampian Carbonate Reservoirs of Southeast New Mexico," *Carbonates and Evaporites*, v. 14, no. 1, p. 84-105.

Broadhead, R.F., 1999b, "Underdeveloped Oil Fields in Upper Pennsylvanian and Lower Permian Carbonates of Southeast New Mexico: Initial Development Missed Major Reserves," The Leading Edge, September, p. 1012-1017.

Broadhead, R.F., 2001a, "New Mexico Elevator Basins 1 - Petroleum Systems Studied in Southern Ancestral Rocky Mountains," Oil and Gas Journal, v. 99, no. 2, pp. 32-38.

Broadhead, R.F., 2001b, "New Mexico Elevator Basins 2 - Petroleum Systems Described in Estancia, Carrizozo, Vaughn Basins," Oil and Gas Journal, v. 99, no. 3, pp. 31-35.

Broadhead, R.F., 2001c, "New Mexico Elevator Basins 3 - Elevator Basin Models - Implications for Exploration in Central New Mexico," Oil and Gas Journal, v. 99, no. 4, pp. 30-36.

Broadhead, R.F., Jianhua, Z., and Raatz, W.D., 2004, Play analysis of major oil reservoirs *in* the New Mexico part of the Permian basin: Enhanced production through advanced technologies: New Mexico Bureau of Geology and Mineral Resources, Open-file report 479, CD-ROM.

Cox, D.M., Brinton, L., and Tinker, S.W., 1998, Depositional facies and porosity development of an Upper Pennsylvanian algal mound reservoir, South Dagger Draw, Eddy County, New Mexico, *in* Winfree, K., ed., Cored reservoir examples from Upper Pennsylvanian & Lower Permian carbonate margins, slopes and basinal sandstones: West Texas Geological Society, Publication 98-103, pages not consecutively numbered.

Cys, J.M., 1986, “Lower Permian Grainstone Reservoirs, Southern Tatum Basin, Southeastern New Mexico, *in* Ahlen, J.L., and Hanson, M.E., Southwest Section of AAPG Transactions and Guidebook of 1986 Convention, Ruidoso, New Mexico: New Mexico Bureau of Mines and Mineral Resources, p. 115-120.

Cys, J.M., and Mazzullo, S.J., 1985, Depositional and diagenetic history of a Lower Permian (Wolfcamp) phylloid-algal reservoir, Hueco Formation, Morton field, southeastern New Mexico, in Roehl, P.O., and Choquette, P.W., eds., Carbonate petroleum reservoirs: Springer-Verlag, New York, p. 277-288.

Dutton, S.P., Kim, E.M., Broadhead, R.F., Breton, C.L., Raatz, W.D., Ruppel, S.C., and Kerans, C., 2005, Play analysis and digital portfolio of major oil reservoirs *in* the Permian Basin: Bureau of Economic geology, The University of Texas at Austin, Report of Investigations 271, CD-ROM.

Gonzalez, T., Sahni, S., and Franta, W. (1977) An efficient algorithm for the Kolmogorov- Smirnov and Lilliefors test, in ACM Transactions in Mathematical Software, Vol.3, 60–64.

Hart, A. (1986) Knowledge Acquisition for Expert Systems, McGraw-Hill.

Jankauskas, L. and McLafferty, S. (1995) Bestfit, Distribution Fitting Software by Palisade Corporation, Proceedings of the 1995 Winter Simulation Conference.

Kues, B.S., and Giles, K.A., 2004, The Late Paleozoic Ancestral Rocky Mountains system in New Mexico, *in* Mack, G.H., and Giles, K.A., eds., The geology of New Mexico, a geologic history: New Mexico Geological Society, Special Publication 11, p. 95-136.

Malek-Aslani, M., 1985, Permian patch-reef reservoir, North Anderson Ranch field, southeastern New Mexico, *in* Roehl, P.O., and Choquette, P.W., eds., Carbonate petroleum reservoirs: Springer-Verlag, New York, p. 265-276.

May, J. (2001) “Knowledge Base”, searchCRM.com definitions.
Reddy, G., 1995, Dagger Draw South, *in* A symposium of oil and gas fields of southeastern New Mexico, 1995 supplement: Roswell Geological Society, p. 209-213.

Press, W., Flannery, B, Teukolsky, S. and Vetterling, W. (n.d.) Numerical Recipes” The Art of Scientific Computing.

Schrader, S. (n.d.) Development, Testing and Application of an Expert System for Petroleum Exploration.

Schrader, S.M. Balch, R.S. Ruan ,T. and Goteti, R. (n.d.) Interactive components for a user-designed customizable fuzzy expert exploration tool. (Working paper)

Sharp, J., in press 2006, Upper Pennsylvanian and Lower Permian phylloid algal mound reservoirs, southeastern New Mexico, *in* Raines, M.A., ed., Resource plays in the Permian Basin; Resource to reserves: West Texas Geological Society, Publication 06-117, DVD.

Speer, S.W., 1993, PP-1. Upper Pennsylvanian, *in* Atlas of Major Rocky Mountain Gas Reservoirs: New Mexico Bureau of Geology and Mineral Resources, p. 156.

Sun Developer Network, http://www.java.sun.com

Wahlman, G.P., 2001, Pennsylvanian-Lower Permian mounds and reefs in the Permian Basin (west Texas-New Mexico): Composition, evolution, distribution, and reservoir characterization, *in* Viveiros, J.J., and Ingram, S.M., eds., The Permian Basin: Microns to satellites, looking for oil and gas at all scales: West Texas Geological Society, Publication 01-110, p. 57-64.

Xiong, H., Gaurav Pandey, Steinbach, M., and Vipin Kumar. (2006) Enhancing data analysis with noise removal. IEEE Transactions on Knowledge and Data Engineering. 18(3), pp. 304-319