ANNEX A OF APPENDIX D
IMPLEMENTING ARRANGEMENT #2
OF
THE U.S. DEPARTMENT OF ENERGY/RUSSIAN ACADEMY OF SCIENCES
MEMORANDUM OF UNDERSTANDING

# CHARACTERIZATION OF CONTAMINATED TERRITORIES, MONITORING NETWORK OPTIMIZATION, AND COST MINIMIZATION

***Background***

The purpose of this proposal is to further advances in the application of spatial prediction and simulation techniques to the characterization, assessment and remediation of contaminated soils at DOE sites. Estimates of DOE's soil contamination problem range from 73 to over 200 million cubic meters (K. Crowley, *Physics* ***Today,*** June 1997). The estimated cost to remediate this soil is 22 to over 60 billion dollars. Our experience indicates that at many sites spatial statistics and decision analysis techniques can defensibly show that the volume of contaminated soil requiring remediation has been overestimated.

***Objectives***

The work proposed herein will provide advanced spatial simulation techniques to be used for decision making in the remediation of contaminated soils at DOE sites. This work is aligned to contribute to the DOE vadose zone roadmapping by providing for the characterization, assessment and remediation of contaminated portions of the vadose zone. Milestone 2, as proposed below, is also configured to provide tools to DOE to understand the data needs and a defensible approach to monitoring network optimization of long-term monitoring of environmental stewardship sites.

This proposal will provide funding to the spatial statistics group headed by Dr. Mikhail Kanevski at the Institute for Nuclear Safety (IBRAE) in Moscow. Dr. Sean A. McKenna, Sandia National Laboratories will collaborate and oversee the work.

***Scope of Work***

This proposal will develop and test two different advanced techniques for the spatial predictions of both a time invariant concentration (e.g., Pu-239 in soil) and ***a*** spatially and temporally varying attribute (e.g., monitoring the migration and attenuation of a groundwater contaminant plume). The machine learning techniques will extend current geostatistical theory to better incorporate information on spatial and temporal trends observed in contaminant data sets. Neural network approaches will be developed, and a feasibility study on support vector machine/support vector regression (SVM/SVR) applications for spatial predictions will be carried out. For both the time invariant and spatial-temporal problems ANN application and validation will be pursued. This new approach will be tested against traditional geostatistical simulations on selected data sets including one provided by DOE.

IBRAE will use data sets of its own choosing to develop the NN technique. SNL/DOE will provide IBRAE with data set(s) on which to validate the new simulation techniques. Only a subset of the data will be provided to IBRAE with the remainder of the data held at SNL for independent validation of the IBRAE results. The time invariant data set (Milestone 1) will be taken from a well-characterized DOE-ER site (e.g., Rocky Flats, Pad 903; SNL ER Site 91) that is not familiar to the IBRAE staff. The spatially and temporally varying data set will also be provided by SNL. This data set will also be chosen from available DOE sites with the intent of approximating the type of data that will collected during future environmental stewardship monitoring (e.g., quarterly groundwater chemistry sampling data from Hanford). This will also be a data set that is unfamiliar to the IBRAE staff. This proposal depends on DOE permission to send these data sets to IBRAE.

The proposed project includes the development of innovative techniques with fundamental developments in modeling of spatial and spatial-temporal data using geostatistical and machine learning algorithms. Important real case studies and demonstrations of methodology are an integrated applied part of the project. As a result of the project several software prototype modules will be developed.

**Milestone 1: Development and testing of spatial simulation techniques for time-invariant contaminants**

Specific activities for this milestone are:

1) Literature review on current geostatistically-based techniques used to create stochastic simulations of the spatial distribution of contaminants in soils. This literature review is to be written into a report and delivered to SNL/DOE.

2) Obtain a two-dimensional data set from SNL and complete preliminary exploratory data analysis including calculation and modeling of experimental variograms. The data set obtained from SNL will provide coordinates and concentration/activity values for each data point. Additional information regarding the action (remediation) levels and the general mode of contaminant emplacement will also be provided. The coordinates of the verification samples, but not the concentrations/activities will also be provided to IBRAE.

3) IBRAE will use existing spatial simulation techniques to model the expected value of the concentration/activity across the site domain. These simulation results will also be processed to provide the probability of exceeding the specified remediation levels.

4) IBRAE will conduct basic research on artificial neural networks for spatial simulations. Development and application of neural network based approach to spatial predictions. Application of such algorithms to the creation of expectation maps and probability of exceedence maps using the data set provided by SNL.

5) Feasibility study on support vector machines -SVM and/or support vector regression - SVR methodology for spatial predictions.

6) Delivery of final report showing expectation and probability of exceedence maps for all data sets and all remediation levels as developed in Activities 3 and 4. For each of the verification sample locations, the full distribution of the simulated concentration/activity will be provided electronically in a tabular format. If these distributions are parametric, the parameters defining the distribution at each verification sample location will be provided. If this distributions are non-parametric, the distribution will be defined at each centile (0.0, 0.1, 0.2, . . . 1 .O)

7) Presentation of initial progress on Milestone 1 results by one representative of the IBRAE staff at the DOE sponsored research conference in Orlando, Florida, in June 2001.

**Milestone 1 Deliverable Summary**

| Activity | Deliverable | Due Date (time out from project start) |
|---|---|---|
| 1 | Report documenting literature review on spatial simulation techniques | 4 months |
| 2 | Report documenting exploratory data analysis of data set provided by SNL | 8 months |
| 3 | Report describing spatial predictions of contaminants using current geostatistical techniques | 10 months |
| 4 | Report describing spatial predictions of contaminants using different ANN algorithms. | 12 months |
| 5 | Report describing feasibility study on spatial predictions of contaminants using SVM/SVR. | 18 months |
| 6 | Preparation and delivery of final report comparing results of activities 3, 4 and 5. Delivery will include full distributions of the concentration/activity at the validation point locations | 24 months |
| 7 | Presentation of work at Orlando, 2001 | June, 2001 (fixed date) |

**Task 2: Development and testing of spatial-temporal simulation techniques for time varying contaminants**

Specific activities for this work are:

1) Literature review on current geostatistically-based techniques used to create stochastic simulations of the spatial-temporal distribution of properties. This literature review is to be written into a report and delivered to SNL/DOE.

2) Work with SNL to locate and develop spatial-temporal data set and complete preliminary exploratory data. The data set will provide coordinates and concentration/activity values for each data point. SNL will maintain the location and values across space and time of a subset of the data to be used for verification of the models developed by IBRAE. The coordinates of the verification samples, but not the concentrations/activities will also be provided to IBRAE.

3) Work with SNL using existing spatial prediction/simulation techniques to model the expected value of the concentration/activity across the site domain at pre-selected time steps. These simulation results will also be processed to provide the probability of exceeding the specified threshold levels.

4) Feasibility study on neural network based approach for spatial-temporal predictions. Demonstrations on selected data sets.

5) Feasibility study on ANN approach for local conditional probability density function estimations.

6) Delivery of final report for Task 2. This report will contain expectation and probability of exceedence maps for all data sets and all specified threshold levels. For each of the verification sample locations, the full distribution of the simulated concentration/activity will be provided electronically for each specified time step in a tabular format. If these distributions are parametric, the parameters defining the distribution at each verification sample location will be provided. If these distributions are non-parametric, the distribution will be defined at each centile (0.0, 0.1, 0.2, …1.0)

7) Presentation of results of the spatial-temporal modeling at a European environmental conference to be determined

**Milestone 2 Deliverable Summary**

| **Activity** | **Deliverable** | **Due Date (time out from project start)** |
|---|---|---|
| 1 | Report documenting literature review on spatial-temporal simulation techniques | 9 months |
| 2 | Report documenting exploratory data analysis of spatial-temporal data set provided by SNL | 15 months |
| 3 | Report describing spatial-temporal predictions of contaminants using current geostatistically-based techniques | 17 months |
| 4 | Report describing feasibility study on spatial-temporal predictions of contaminants using artificial neural networks. | 20 months |
| 5 | Preparation and delivery of final report comparing results of activities. Delivery will | 24 months |

| | | |
|---|---|---|
| | include full distributions of the concentration/activity at the validation point locations | |
| 6 | Presentation of spatial temporal modeling results at a conference to be determined (one IBRAE representative). | 24 months |

**Principal Investigators:**
Dr. S. McKenna, SANDIA
Dr. M. Kanevski, IBRAE

**Reports:**
All reports shall be prepared and submitted by the 1 5th of each relevant month to:

Dr. Sean A. McKenna
Sandia National Laboratories
P.O. Box 5800 MS0735
Albuquerque, NM 87 185-0735
EMAIL: samcken@sandia.gov

A copy of each report should be submitted by fax to:

Mr. Skip Chamberlain, Program Manager
U.S. Department of Energy, Office of Science and Technology, EM-53
19910 Germantown Rd.
Germantown, MD 20874, USA
FAX: 202-488-3 158

**Extension to future years:**
DOE is interested in the long-term monitoring of environmental sites. Toward this goal, upon the completion of the present project DOE will consider a proposed third-year extension to this proposal that may include the application of the neural network spatial - temporal modeling approach examined in Activity 4 of Milestone 2 to the problem of designing and optimizing long-term montioring networks.

## PROPOSED PERIOD OF PERFORMANCE AND PAYMENT SCHEDULE

This project will be conducted over the next 24 months. The beginning of the contract is October 1, 2000. The completion of the project is September 30, 2002. The total project cost is $ 200,000 USD. The payments will be made through standard DOE-EM procedures according to the following schedule:

| Schedule | Payment | Deliverable |
|---|---|---|
| December 2000 | $25,000 | Progress Report |
| March 200 1 | $25,000 | Report outlined in Milestone 1, Activity 1 |
| June 2001 | $25,000 | Reports outlined in Milestone 1, Activity 2 and Milestone 2, Activity 1 |
| September 200 1 | $25,000 | Reports outlined in Milestone 1, Activities 3, 4 |
| December 200 1 | $25,000 | Report outlined in Milestone 2, Activity 2 |
| March 2002 | $25,000 | Reports outlined in Milestone 1, Activity 5 and Milestone 2, Activity 3 |
| June 2002 | $25,000 | Report outlined in Milestone 2, Activity 4 |
| September 2002 | $25,000 | Reports outlined in Milestone 1, Activity 6 and Milestone 2, Activity 5 |

The payments will be made upon receipt of the specified deliverables. The final payment will be paid upon acceptance of the Final Report by DOE.

The responsibility of the parties for the violation of the terms of the Contract shall be determined by the Joint Coordinating Committee for Science and Technology.

### LEGAL ADDRESS

Russian Academy of Sciences, Nuclear Safety Institute
52, B. Tulskaya
1 13 191 Moscow, Russia

Tel: +(7-095) 952-242 1
Fax: +(7-095) 958-0040

### BANKING INFORMATION: XXXXXXXXXXXX

Signed: 5/31/00

Mr. S Chamberlain, PE
Program Manager
U.S. Department of Energy

Prof. Mikhail Kanevski
Principal Investigator
IBRAE

31.05.2000