5

Intelligent Gamma-Ray Data Processing for Environmental Monitoring

5.1 Introduction: gamma-ray spectroscopy for environmental monitoring and security

An important part of monitoring for environment and security is gamma-ray radioactivity monitoring. It extensively uses gamma-ray spectroscopy techniques. Gamma-ray spectroscopy is the science (or art) of identification and/or quantification of radionuclides by analysis of the gamma-ray energy spectrum produced in a gamma-ray spectrometer [Ortec, 2010]. It is a widely used technique with example applications in security and nuclear materials safeguards, geology and mineralogy, materials testing and reactor corrosion monitoring, industrial process monitoring, nuclear medicine and radiopharmaceuticals, health physics personnel monitoring, forensics, etc. Here we are especially interested in environmental and security-related radioactivity monitoring.

This areas use both remote and close-range gamma-ray spectroscopy techniques. Close-range gamma-ray spectroscopy analyses the spectra of individual objects or samples. Remote gamma-ray spectroscopy is used when there is no possibility of approaching or close contact with the monitored object, or when monitoring of vast territories or multiple sites is required.

Gamma-ray spectrometry is done ex-situ, i.e., by bringing the samples to laboratory, and in-situ, i.e. from aircraft, field vehicles, on foot, in boreholes, on the sea bottom, etc. Classical methods of gamma-ray spectroscopy involve taking samples into a counting room for analysis in a controlled environment and geometry. However, with laboratory analyses, there is much labor involved, and a long turnaround time for the analysis results. When the sample is spread over a wide area or is in a fixed location such as a pipeline, it is much easier and less costly to measure samples in-situ [Canberra]. These measurements, while simpler to carry out and available immediately, are prone to difficulty, usually related to calibration. Ex-situ and in-situ gamma-ray spectroscopy often complement each other.

An important part of environmental gamma-ray monitoring is gamma-ray surveys [IAEA, 2003]. Ground and airborne gamma-ray measurements may cover large areas of the earth's surface. They often result in maps of terrestrial radiation and radioelement concentrations. Radiometric surveys and maps of natural background levels are used for solving geological and environmental problems in a number of fields, including mineral exploration or prospecting, geochemical mapping, soil mapping for agriculture, environmental assessment, etc. In addition to geoscience, it can be also applied in emergent situations to monitor environmental contamination in the event of an accidental release or major nuclear event, including radioactive spill from nuclear reactors or during transportation and

storage of radioactive wastes, accidental loss of industrial and medical radiation sources, etc. as well as nuclear weapons incidents or tests.

Gamma-ray surveys provide a baseline against which man-made contamination can be estimated. This makes the radiological assessment of the environment possible, showing general regional trends in radionuclide distribution, etc. They may be used to estimate and assess the terrestrial radiation dose to the human population and to identify areas of potential natural radiation hazard, to determine the effectiveness of cleanup efforts after a spill or an industrial accident, to trace and to clean up old contamination, and to verify that new contamination is not occurring. For example, surveys are regularly carried out in the radiation risk zones around facilities where radioactive materials are used or produced, such as nuclear power plants and other facilities dealing with nuclear fuel cycle, radioactive waste and material repositories and their transportation routes, science, industrial, and medicine sites, etc. Similar surveys are used to assess the contamination in areas of former mining. Besides territories, of interest is gamma-ray monitoring of objects that use radioactive materials or fuel, contaminated buildings, containerized waste, industrial processes, laboratories, etc., as well as moving objects (vehicles, movable facilities, cargo, people, etc.).

Gamma-ray monitoring is also closely related to security, which in this context is mainly associated with the dangers of nuclear proliferation and, more recently, radiological and nuclear terrorism. Nuclear terrorism denoting the use, or threat of the use, of nuclear or radiological weapons ("dirty bomb") in acts of terrorism is nowadays becoming an immediate challenge for the entire world [Ruff, 2006], [Markman et al, 2003]. The challenges of coping with the consequences of possible acts of nuclear terrorism or of the detection of radiological or nuclear materials in a planted device are closely related to the issues of environmental gamma-ray monitoring and spectroscopy mentioned above.

This also concerns the challenges of pre-empting nuclear and radiological terrorism that requires detection, interdiction, and identification of smuggled nuclear materials, as well as preventing or intercepting the illicit trafficking, theft and loss of nuclear and other radioactive materials and weapons. Monitoring is required at the borders and ports of entry, including land, rail, marine, or air border crossings, as well as the facilities where radioactive materials and weapons are produced or stored. To detect illicit transport of radioactive material in vehicles, people, luggage, mail, cargo, containers, etc., radiation portal monitors are commonly used.

All those mentioned pressing challenges and applications require ever-enhancing accuracy, precision, stability, and reliability of gamma-ray monitoring. In its turn, this requires novel gamma-ray spectroscopic techniques to employ state-of-the-art approaches and methods of intelligent data processing and machine learning. In this chapter, we discuss gamma-ray spectrometry techniques for both close-range and remote monitoring. First, we consider gamma-ray spectroscopy with multichannel spectrometer both for reproducible and unknown geometry of measurements. Then, techniques for airborne gamma-ray spectrometry are discussed. Results of experimental comparison are provided to illustrate the performance of the employed machine learning techniques.

5.2 A machine learning approach to multi-channel gamma-ray spectroscopy

Gamma-rays are the most penetrating radiation from natural and manmade sources. Radionuclides emit gamma-rays of specific energies characteristic for elements and isotopes. Gamma-ray measurements are conducted in two modes. Total count measurements register gamma-rays of all

energies and are used to monitor the gross level of the gamma radiation field and to detect the presence of anomalous sources. Gamma-ray spectrometers, on the other hand, measure both the intensity and energy of radiation, thus enabling the source of the radiation to be diagnosed [IAEA, 2003].

In order to interpret spectrometer measurements in real-world conditions, a number of intelligent data processing techniques have been developed or adapted. In this section, we review available techniques, as well as machine learning approaches such as model selection, sparse approximation, and blind source separation that we develop and/or apply to gamma-ray spectroscopy. Performance of those techniques is illustrated using artificial and real gamma-spectrum data.

5.2.1 Gamma-ray spectroscopy for fixed and non-fixed geometry of measurements

Gamma-ray spectroscopy is the quantitative study of the energy spectra of gamma-ray sources. The result of the gamma-ray spectrum measurement by a multi-channel spectrometer (Figure 78) may be considered as L numbers z_i (i=1,...,L) each representing the count $f(z_i)$ of gamma-rays of particular energy z_i. This output signal of the spectrometer can be decomposed as

$$f(z_i) = \sum_{j=1,...,N} w_j \psi_j(z_i) + \varepsilon(z_i), \quad (1)$$

where z_i is the gamma-ray energy in the i-th channel, ψ_j is the detector response function, w_j is its weight, N is the number of detector response functions in the decomposition, $\varepsilon(z_i)$ is the noise realization.

Figure 78. Multi-channel spectrometer "Vector"

The detector response function $\boldsymbol{\psi}(z)$ (Figure 79) is the spectrometer output when the input is gamma-rays having the energy z. Since a multi-channel spectrometer should be able to deal with the gamma-ray sources of any energy from its working range, one needs to have detector response functions for all z_i (i=1,...,L). However, real monochromatic gamma-ray sources do not exist for much energy. Therefore, the detector response functions are usually generated using Monte-Carlo simulations of gamma-ray interactions with the detector crystal [Zerby, 1963], [Breismeister, 2000], [Hendriks et al, 2002], [Sempau et al, 2003].

The task is, using the results of spectrum measurement $f(z_i)$ and knowledge of the detector response function set corresponding to possible radionuclides, to estimate non-zero w_j (weights or activities) and thereby to identify radionuclides with their activities that produced the measured spectrum (Figure 80).

Gamma-ray spectrometry is considered in fixed and non-fixed geometries of measurements. For measurements in a **fixed geometry**, position of the detector relative to the radiation source and geometry of the source are known; if the radiation source is shaded by a substance that absorbs radiation, the geometry and nature of this screen are known as well. So, detector response functions can be generated for this known geometry. Since the maximum number N of detector response functions is equal to the number of channels L, and each function has its value in each of the L channels, the total set of detector response functions forms the L×L matrix.

Gamma-ray spectrum measurements in a fixed geometry are used to estimate the radionuclide content in the samples of food, soil, construction materials, etc., often using the Marinelli beaker.

The task is complicated by fluctuations in the natural radiation background, low levels of radionuclide activity, a complex spectral composition of radiation.

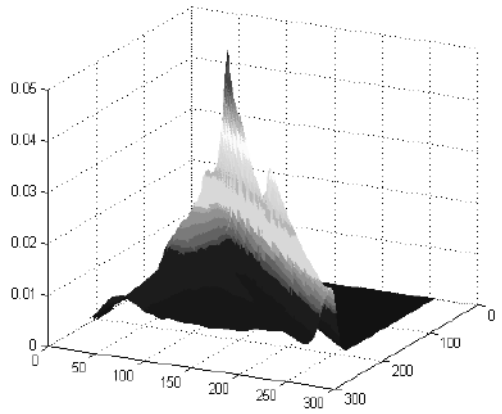

Figure 79. Detector response functions (200 functions of 256 channels each

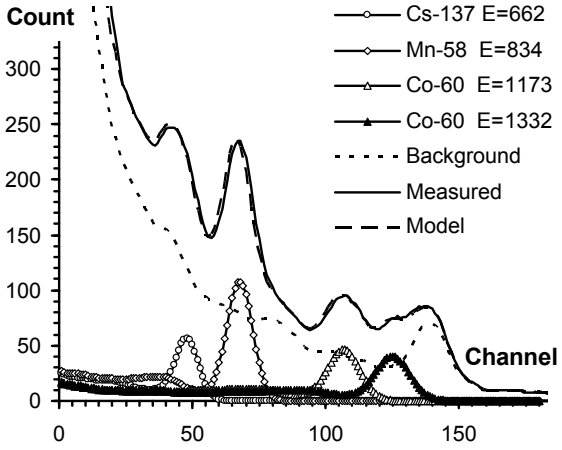

Figure 80. Spectrum decomposition using detector response functions

Measurements in a **non-fixed geometry** (unknown and complex) have the following peculiarities: position of the detector relative to the radiation source and the geometry of the source are unknown; the radiation source is shaded by a substance that absorbs radiation, geometry and nature of this screen are unknown.

So, the detector response functions for a non-fixed geometry of measurements cannot be generated in advance, and the decomposition of Eq(1) cannot be immediately applied to this case. To cope with this problem, we proposed [Zabulonov et al, 2004a] to split each detector response function into three separate functions $\boldsymbol{\psi}^A(z)$, $\boldsymbol{\psi}^C(z)$, $\boldsymbol{\psi}^R(z)$ corresponding to three characteristic spectrum regions: the total absorption peak, the Compton part and the reverse flight peak. So, each detector response function is represented as

$$\boldsymbol{\psi}(z) = w^A \boldsymbol{\psi}^A(z) + w^C \boldsymbol{\psi}^C(z) + w^R \boldsymbol{\psi}^R(z). \quad (2)$$

Therefore, for this case the maximum set of detector response functions becomes ψ_{js}, j = 1,…,L, s={A,C,R} and those functions form the L×3L matrix.

Gamma-spectroscopy in a non-fixed geometry of measurements often occurs in practice, since placing the assessed environmental objects into the fixed geometry conditions is often difficult or impossible. It concerns radiation assessment of cargo, vehicles, facilities, field sources, as well as gamma-ray surveys, etc.

5.2.2 Previous gamma-ray spectroscopic methods

Some traditional approaches to gamma-ray spectroscopy are shown in Figure 81.

The traditional Regions-of-interest (ROI) method and its modifications [Desbarats and Killeen, 1990] make spectrum analysis in certain regions of interest. Usually the number N of those regions is a small fraction of L. At the calibration stage, the "contribution matrix" is calculated. Its columns correspond to the expected radionuclides, rows correspond to ROIs, and elements are the gamma-ray counts registered in the appropriate ROI from the expected radionuclides.

The measured spectrum is processed as follows. All ROIs are checked, starting from the one having the maximum energy. If some ROI has count value above the threshold, the presence of the corresponding nuclide is reported and its contribution is subtracted from the total spectrum, weighted by the relevant elements of the contribution matrix from all ROIs. This is repeated for all ROIs, providing the radionuclides with their activities. The stripping method [Desbarats and Killeen, 1990] uses the detector response functions of expected radionuclides instead of the gamma-ray counts in the ROIs.

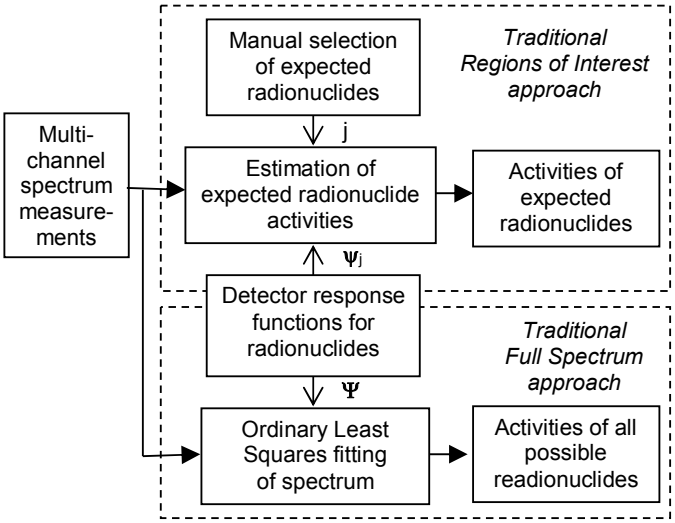

Figure 81. Regions-of-Interest and Full Spectrum approaches to gamma-ray spectroscopy

Thus, traditional methods require a-priori knowledge of radionuclide composition in order to estimate radionuclide activities and are not intended to assess samples of arbitrary composition. In order to overcome this drawback, full spectrum processing techniques were proposed [Minty, 1992], [Hendriks et al, 2001], [Guoa et al, 2004], [De Meijer, 2007], [Newman et al, 2008]. Here, the total matrix of detector response functions Ψ is used, and the activity vector $\mathbf{w}$ is obtained by the Ordinary Least Squares (OLS) solution:

$$\mathbf{w} = \Psi^{+}\mathbf{f}, \tag{3}$$

where Ψ^{+} is the pseudoinverse matrix.

However, due to the noise, this method assigns non-zero weights to nuclides actually not present in the spectrum. To increase noise resistance, maximum noise fraction (MNF) [Dickson and Taylor, 1998; 2000] and noise-adjusted singular value decomposition (NASVD) [Minty and McFadden, 1998], [Minty and Hovgaard, 2002], [Mauring and Smethurts, 2005] are sometimes used. Regretfully, those techniques require information about noise.

For a non-fixed geometry of measurements, performance of the traditional methods is even poorer than for a fixed geometry. If one uses Eq.(1), detector response functions should be known for all possible geometries and their proper set should be chosen in-situ, that cannot be implemented in practice. If one considers Eq.(2), the drawbacks of the traditional methods described above become more severe since the number of weights (activities) to be estimated grows.

Another approach to the non-fixed geometry spectroscopy consists in the following. Some hypotheses about the composition and activity of nuclides in the measured spectrum are set up. Then the Monte Carlo simulation of photon transport is done for those hypotheses, taking into account the particular geometry of measurements. If the measured spectrum is exactly reconstructed by the simulation, the desired composition and activity are found [Gal et al, 2001], [Toubon et al, 2006] This approach has a large computational complexity and so cannot provide real-time results, since a lot of hypotheses should be verified in case of poor a-priori information.

5.2.3 Intelligent data processing methods for spectrometry

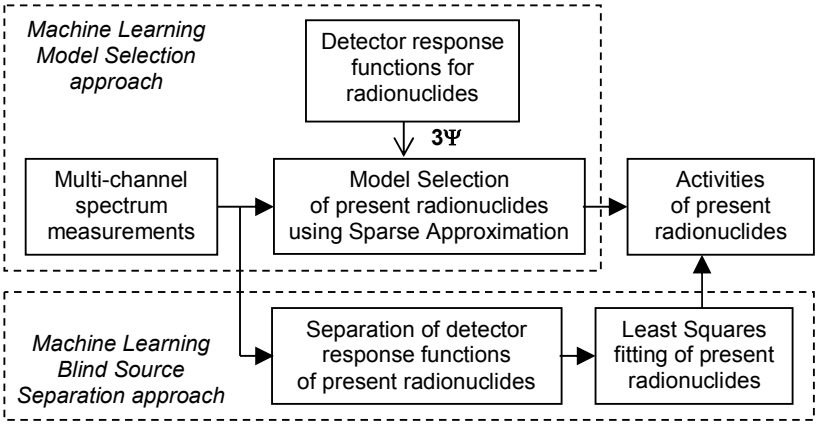

Figure 82. Gamma-ray spectroscopy by machine learning

To improve the accuracy and stability of radionuclide activity estimation, we develop an approach to full spectrum processing using machine learning methods of model selection, sparse approximation, and blind source separation (Figure 82).

✓ Model selection and sparse approximation

Model selection is a problem arising in the areas of machine learning and data mining. Given data usually consisting of input-output pairs, a model is built to relate the input and the output. A number of approaches have been proposed to build the model, including linear models, neural networks, classification and regression trees, kernel methods, etc. The task of model selection is to choose a close-to-optimal model in terms of the particular model usage and application.

A number of model selection methods use various model selection criteria [Akaike, 1974], [Mallows, 1973], [Hansen and Yu, 2001]. Model selection criteria are formulated in such a way that they balance the model complexity vs. the approximation error, and so automatically reduce the model complexity with increasing noise levels. At the optimal models, criteria reach minimum. Model selection criteria have been developed based on various assumptions, i.e. using predictive training error [Mallows, 1973]; generalization error [Sugiyama and Ogawa, 2001], [Niyogi and Girosi, 1996] and related criteria, such as based on information statistics [Akaike, 1974], description length [Rissanen, 1978; 2002], etc.

Let the data set D_L be represented by L pairs $D_L=\{(\mathbf{x}_i,y_i)\}_{i=1,\dots,L}$, $\mathbf{X}\in\Re^{L\times N}$, $y_i=y_{0i}+\varepsilon_i$, ε is Gaussian additive noise. The predictive training error is adopted as the error measure [Sugiyama and Ogawa, 2001] between estimated and true values at sample points contained in the training set:

$$I_{PTE} = E_\varepsilon||\mathbf{By}-\mathbf{y}_0||^2 = ||\mathbf{By}_0-\mathbf{y}_0||^2 + \mathrm{tr}(\mathbf{BQB}^T), \quad (4)$$

where $\mathbf{B}=\mathbf{X}_s(\mathbf{X}_s^T\mathbf{X}_s)^{-1}\mathbf{X}_s^T$ is the mapping from $\mathbf{y}$ to its estimation by the model, $\mathbf{Q}$ is the noise covariance matrix, $\mathbf{X}_s$ is the data for the model corresponding to $s\le N$ non-zero components of $\mathbf{w}$ corresponding to the model, E_ε denotes the ensemble average over the noise, $||.||$ denotes the norm, $\mathrm{tr}(\cdot)$ denotes the trace of an operator.

The Mallows criterion C_P [Mallows, 1973] gives an unbiased estimate of the predictive training error:

$$C_P = RSS + 2\sigma^2 s. \quad (5)$$

where RSS is the residual sum of squares, σ^2 is the noise variance.

Generalization error is the difference between the expected risk of the obtained solution $f_{H,L}$ and the expected risk of the "optimal" (true) solution f_0 [Niyogi and Girosi, 1996]:

$$I_G = I(f_{H,L}) - I(f_0),\ f_{H,L} = \mathrm{argmin}_{f \in H} I_{emp}(f),\ I_{emp}(f) = 1/L \sum_{i=1,L} V(y_i, f(x_i)), \quad (6)$$

where $f_{H,L}$ is the solution obtained using the limited data set of length L and the model H with limited parametrization, I_{emp} is the empirical risk, $V(\cdot)$ is the loss function.

Subspace Information Criterion SIC gives an unbiased estimate of the generalization error [Sugiyama and Ogawa, 2001]. For least mean squares learning,

$$SIC = [\langle \mathbf{Ty}, \mathbf{y} \rangle - \sigma^2 \mathrm{tr}(\mathbf{T})]_+ + \sigma^2 \mathrm{tr}(\mathbf{T}_S^+),\ \mathbf{T} = \mathbf{T}_S^+ - \mathbf{T}_F^+ \mathbf{T}_S \mathbf{T}_S^+ - \mathbf{T}_S^+ \mathbf{T}_S \mathbf{T}_F^+ + \mathbf{T}_F^+, \quad (7)$$

where $(\cdot)^+$ denotes the matrix pseudoinverse, $\mathbf{T}_S$ and $\mathbf{T}_F$ are the L×L matrices with the elements obtained by calculation of kernel function values K(x,x') for the selected $\mathbf{X}_s$ and for the whole $\mathbf{X}$ respectively, $[\cdot]_+$ is defined as $[t]_+ = \max(0,t)$.

Akaike's information criterion [Akaike, 1974] evaluates the generalization error from the information statistics point of view. It is calculated as

$$AIC = 2s + L \ln RSS, \quad (8)$$

where L is the number of observations, s is the number of parameters.

Minimum description length criteria for regression can be written as

$$I_{MDL} = DL(\mathbf{y}|\mathbf{X}_s) + DL(s), \quad (9)$$

where $DL(\mathbf{y}|\mathbf{X}_s)$ is the description length of $\mathbf{y}$ by the model of complexity s, DL(s) is the description length of the model.

In [Hansen and Yu, 2001] expressions for LyMDL and gMDL criteria were proposed, whose minima correspond to the minimum description length criterion:

$$\begin{gathered} LyMDL = 0.5(\mathbf{y}^T\mathbf{y} - FSS)/\sigma^2 + 0.5\, s_1\, [1 + \log(FSS/(\sigma^2 s))] + 0.5\log(L) \text{ for } FSS > s\sigma^2, \\ L(\mathbf{y}) = 0.5(\mathbf{y}^T\mathbf{y})\,\sigma^2 \text{ otherwise}, \end{gathered} \quad (10)$$

$$\begin{gathered} gMDL = 0.5L \log(RSS/(L - s_1)) + 0.5s \log(F) + \log(L) \text{ for } R^2 \geq s/L, \\ gMDL = 0.5L \log(\mathbf{y}^T\mathbf{y}/L) + 0.5 \log(L) \text{ otherwise}, \end{gathered} \quad (11)$$

where $FSS = \mathbf{y}^T\mathbf{X}_S(\mathbf{X}_S^T\mathbf{X}_S)^{-1}\mathbf{X}_S^T\mathbf{y}$, $F = (L-s)(\mathbf{y}^T\mathbf{y} - RSS)/(s\ RSS)$, R is the usual multiple correlation coefficient.

For the spectroscopy task, the model selection approach is assumed to avoid the situation when the model includes elements actually not present. In order to use the approach, we must generate the models and choose the one with the minimal model selection criterion value. Since exhaustive search is impossible for practical settings, we use greedy techniques based on the ideas of sparse approximation [DeVore, 1998], [Donoho et al, 2004].

The field of sparse approximation is concerned with the problem of representing a target vector using a short linear combination of vectors drawn from a large, fixed collection called a dictionary This problem arises in applications as diverse as machine learning, signal and image processing,

imaging sciences, communications, numerical analysis and statistics. It is known that sparse approximation is computationally hard in the general case. Nevertheless, it has recently been established that there are tractable algorithms for solving these problems in a variety of interesting cases. It was also shown that, under certain conditions, sparse approximation and Support Vector Machine technique proposed by V. Vapnik give the same solution [Vapnik and Chervonenkis, 1970].

Sparse approximation methods are used to solve $\Psi\mathbf{w} = \mathbf{y}_0$ for the complete basis Ψ (i.e. the basis that exactly represents $\mathbf{y}_0$) with the maximum sparseness in terms of the l_0 norm ($\mathbf{w}^*=\min_\theta||\mathbf{w}||_0$). They include step-wise regression [Seber, 1977], k-term approximation [Temlyakov, 2003], matching pursuit [Mallat and Zhang, 1993], [Tropp and Gilbert, 2007].

Matching pursuit at the (k+1)-th step calculates the approximation of $\mathbf{y}$, $\mathbf{f}_{k+1}=\mathbf{f}_k+w_{k+1}\psi_{k+1}$ choosing ψ_{k+1} ($\psi_{k+1} \subset \Psi$ without Ψ_k) and w_{k+1} that minimize the residual sum squared $||\mathbf{R}_k-w\psi||^2$:

$$(w_{k+1}, \psi_{k+1}) = \text{argmin}_{w,\psi} ||\mathbf{R}_k-w\psi||^2, \mathbf{R}_k=\mathbf{y}-\mathbf{f}_k. \quad (12)$$

The drawback of matching pursuit is that its application to noisy $\mathbf{y}$ requires stopping criteria.

We proposed to use model selection criteria as stopping criteria [Revunova, 2007b]. It works as follows.

Initialize $\mathbf{f}_0=0$, $\mathbf{R}_k=\mathbf{y}$.

Find the basis function (column ψ) of $\mathbf{X}$ that maximizes the scalar product with the current residual $\mathbf{R}_k$:

$$\gamma_k = \text{argmax}_{i=1,\dots,N} |\langle \mathbf{X}(\cdot,i), \mathbf{R}_k \rangle|. \quad (13)$$

In order to get a more accurate solution proposed in the framework of orthogonal matching pursuit [Tropp and Gilbert, 2007], recalculate the solution for the already chosen basis functions Ψ_k by performing the OLS fitting:

$$\mathbf{w}_k=\Psi_k^+ \mathbf{y}. \quad (14)$$

Calculate a new residual:

$$\mathbf{R}_{k+1} = \mathbf{y} - \mathbf{w}_k \Psi_k. \quad (15)$$

Calculate the value of the used model selection criterion CR(k). If CR(k)≥CR(k–1), select the model with s=k–1 as optimal. Otherwise, go to Eq.(13) and repeat the cycle.

Let us compare traditional methods and our model selection approach for spectroscopy in a fixed and non-fixed geometry of measurements [Revunova, 2008], [Zabulonov et al, 2004b], [Zabulonov et al, 2009a]. In our approach, Ψ contains detector response functions, and $\mathbf{y}$ is the result of spectrum measurement.

For a fixed geometry of measurements, we compared the relative error of the traditional region of interest (ROI) and stripping (STR) methods vs. orthogonal matching pursuit with the stopping by the model selection criterion (MPCR). Fixed geometry of measurements was ensured by placing the test samples in a standard 0.5l Marinelli beaker and using a detector with a collimator. Traditional ROI and STR methods were set up for processing of ^{137}Cs (Cesium) and ^{40}K (Potassium). Without changing this setup, an additional nuclide ^{230}Th (Thorium) or two additional nuclides ^{230}Th, ^{226}Ra (Radium) were then added to the sample.

The relative error was estimated for the radionuclide with the lowest energy ^{137}Cs, since its spectrum is "hidden" by the other radionuclides. Model selection criteria of Mallows, Akaike, Bin Yu were used with MPCR. Slightly better results were obtained for the Mallows criterion Cp. The experimental results are given in Table 3.

Table 3. Relative error of ^{137}Cs activity in the multi-nuclide sample obtained by ROI, STR, MPCR

Nuclides	Cs, K	Cs, K, Th	Cs, K, Th, Ra
ROI	0.1%	11%	19%
STR	0.5%	5%	12%
MPCR	0.35%	0.94%	3%

When processing the spectrum of known composition (^{137}Cs, ^{40}K), the ROI method provides the best relative error, as it was set up precisely for those nuclides and is influenced by noise only inside its regions of interest. MPCR provides a smaller error compared to STR. When processing the spectrum with additional radionuclides (Th or Th, Ra) unknown to the ROI and STR methods, MPCR provides the smallest relative error measuring of the ^{137}Cs activity.

For the first experiment in a non-fixed geometry of measurements, the samples including ^{60}Co(Cobalt) (2 spectrum lines), ^{137}Cs, ^{134}Cs, ^{58}Mn (Manganese) have been simulated by their 3-part detector response functions Eq.(2) (Figure 83). The weight of the reverse flight peak was 0 and the weight of the total absorption peak was 1 for all radionuclides. The weight of the Compton part was 2 for ^{137}Cs and ^{134}Cs and 1 for the remaining radionuclides. This simulates the situation when ^{137}Cs and ^{134}Cs which fell on the soil surface gradually seep into its lower levels and thus the gamma-rays from the total absorption peak redistribute to the Compton part, so that their coefficients in Eq.(2) become different. The reverse flight peak in this situation is practically not observed and so its coefficient is set to zero.

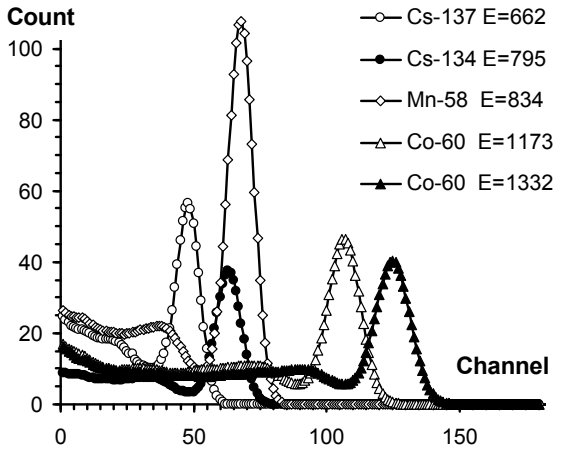

Figure 83. Detector response functions of the true model

So the number of non-zero detector response functions was 5×2=10. Then the full spectrum was simulated by the linear model Eq.(1), where the activities of radionuclides were set to w(^{137}Cs)=3, w(^{134}Cs)=4.5, w(^{58}Mn)=4, w(^{60}Co)=7. At last, a realization of an additive noise with the Gaussian distribution was added to the spectrum. To simulate unexpected radionuclides, 15×2=30 additional detector response functions ψ were added to Ψ, so the total number of basic functions in Ψ was N=40. The number of spectrometer channels L was 512. The tested methods used to estimate the radionuclide activities were: OLS with all basis functions (FULL), OLS with the detector response functions actually in the model (TRUE), MPCR using model selection criteria of Mallows C_P, Akaike AIC, BinYu L_Y (MDL). We compared the errors calculated as the l_2-norm of the difference between the

true activity vector and the estimated activity vector at different noise levels. The results in Figure 84 show that MPCR for all model selection criteria and all noise levels performs better than FULL, but worse than TRUE. The error for MPCR with MDL is somewhat less than that for C_P and AIC.

The next experiment was conducted using the really measured spectra of a point gamma radiation source on the background of a distributed source with a higher energy. ROI, STR, and MPCR with the MDL model selection criterion were compared. In the experiment, the point source was ^{137}Cs and the distributed source was ^{40}K; the ^{40}K activity was 51% of the ^{137}Cs activity. ROI and STR were set up for ^{137}Cs and ^{40}K. The relative error of the ^{137}Cs activity was 92% for ROI, 85% for STR, and 12% for MPCR.

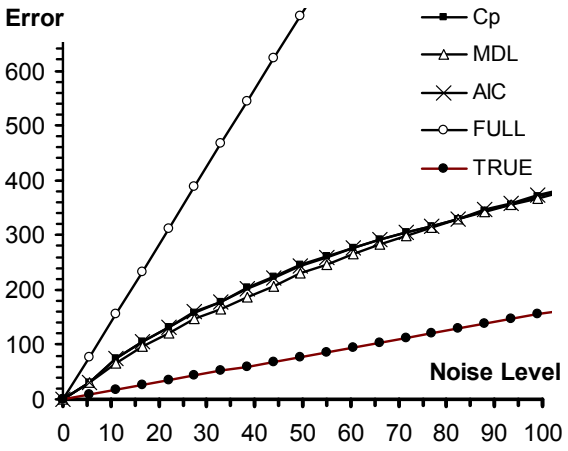

Figure 84. Radionuclide activity error vs noise level for various models and model selection criteria

✓ Model optimality tests

In order to determine why different model selection criteria vary in accuracy, we compared [Revunova, 2008], [Zabulonov et al, 2004a], [Zabulonov et al, 2009a] the dependence of the model dimensionality and the model error on the noise level. The comparison showed that the best accuracy is provided by the criterion that keeps the true dimension of the model at the higher noise levels than the other model selection criteria.

In [Gribonval et al, 2006] a test was proposed for checking if the particular set of basic functions in the model is true. It is based on the notion of l_0-optimal solution that provides both the minimum approximation error and the maximum model sparsity. Experiments showed [Revunova, 2008] that with the increasing noise levels the error using such a test is smaller than that of the traditional model selection criteria [Revunova, 2005a], [Revunova and Rachkovskij, 2005], [Zabulonov et al, 2005].

The drawback of this model optimality test is that it can only be used for some bases with valid "basis connectivity" condition [Gribonval et al, 2006]. So, research on how to solve this problem for certain types of detector response basis functions could lead to further improvement in the accuracy of radionuclide activity estimation under noisy conditions.

✓ Blind source separation

Blind source separation (BSS) is the separation of a set of signals from a set of mixed signals, without the aid of information (or with very little information) about the source signals or the mixing process. In machine learning, BSS is considered as one of the most powerful unsupervised learning techniques. Unsupervised learning is distinguished from supervised learning and reinforcement learning in that the learner is given only unlabeled examples. Theoretical foundations of the BSS methods are well developed [Chichocki and Amari, 2002], [Comon, 1994], [Hyvarinen, 1999].

The most straightforward case is linear BSS with the number of observed mixtures equal to the number of unknown sources. Here the data D_L are represented by the set of L mixture vectors $\mathbf{y}_i \in \Re^N$: $D_L=\{\mathbf{y}_i\}_{i=1,..,L}$. The mixture vector $\mathbf{y}$ depends on the source vector $\mathbf{x}$ as a linear model $\mathbf{y}=\mathbf{A}\mathbf{x}$, where the

full rank mixture matrix $\mathbf{A}\in\Re^{N\times N}$ is unknown. The task is to obtain L unobserved vectors **x** (matrix $\mathbf{X}\in\Re^{L\times N}$) using L known vectors **y** (matrix $\mathbf{Y}\in\Re^{L\times N}$) by estimating the un-mixture matrix $\mathbf{B}=(\mathbf{b}_1,\ldots,\mathbf{b}_N)^T\in\Re^{N\times N}$ and then calculating **X**=**YB**.

Well-known approaches to solution of the BSS problem use a priori information about the sources. The sources are supposed to be statistically independent and having nongaussian distribution (independent component analysis ICA [Amari and Cichocki, 1998], [Amari et al, 2002]), or uncorrelated and having the Gaussian distribution (principal component analysis PCA [Chichocki and Amari, 2002]), or sparse (sparse component analysis SCA [Li et al, 2004]), etc. Based on those assumptions, a cost function is constructed (e.g., mutual information of sources) and then minimized by the optimization algorithms, providing the solution.

Usually, BSS-related methods such as MNF [Dickson and Taylor, 1998; 2000] are used in spectroscopy for noise suppression. We propose using BSS in spectrometry to reconstruct detector response function as follows [Zabulonov et al, 2009b]. A system of several detectors located at various distances from the radiation sources simultaneously measures the gamma radiation spectra. Those multiple measured spectra are processed using the BSS method and at the output we get the separated detector response functions of individual radionuclides as reconstructed mixture components. This BSS approach to spectroscopy does not require measurement of the background radiation, knowledge of detector response functions, and the number of reconstructed hidden sources only depends on the number of detectors.

We carried out experiments [Zabulonov et al, 2009b] investigating BSS performance in spectrometric tasks. In one of the experiments, the radiation sources Cs^{137} and K^{40} were measured simultaneously by three detectors located at various distances from the sources.

The measured spectra (Figure 85) were input to the Independent Component Analysis algorithm FastICA [Hyvarinen, 1999] belonging to the BSS family. The separated hidden sources corresponding to ^{137}Cs, ^{40}K and background radiation are presented in Figure 86.

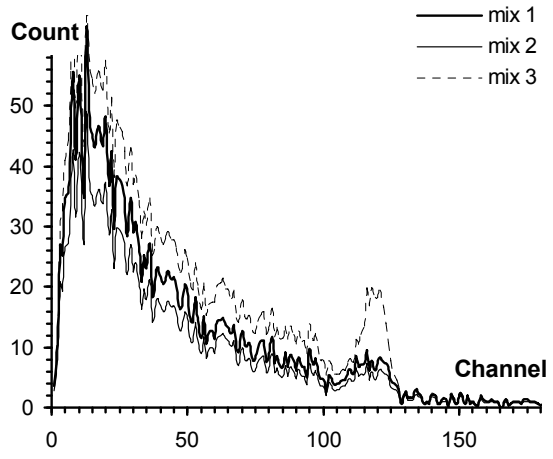

Figure 85. The results of gamma-ray spectra measurements by 3 detectors. ^{137}Cs, ^{40}K, and are mixed in the spectra

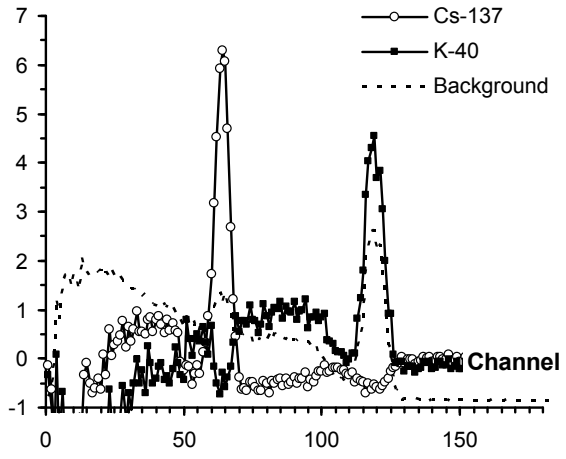

Figure 86. Independent components separated from the spectrum mixtures: ^{137}Cs, K^{40}, and background

The BSS methods are sensitive to noise and require large data sets, so that their usage in non-stationary applications is rather problematic [Bach and Jordan, 2002]. To improve BSS performance

in this situation, we have developed a cost function based on the algorithmic complexity approach, and algorithmic mutual information in particular [Grunvald and Vitanyi, 2003]:

$$I(x_1:x_2) = K(x_1) - K(x_1 | x_2), \tag{16}$$

where $K(x_1)$ is the algorithmic complexity of x_1, $K(x_1|x_2)$ is the complexity of x_1 conditioned upon x_2.

The algorithmic information theory provides much of the foundation of machine learning and statistical and inductive inference. However, it is non-trivial to calculate it. We propose [Revunova, 2005b; 2005c] to calculate $K(\cdot)$ using minimum description length [Rissanen, 1978; 2002] based on a universal model nMDL proposed by Bin Yu [Hansen and Yu, 2001]:

$$\begin{gathered} nMDL = (L/2)\log(RSS/(L-k)) + 0.5\,k\log(F) + \log(L-k) - 3/2\log(k), \\ F = (\mathbf{x}^T\mathbf{x} - RSS)/(k\,RSS/(L-k)),\ RSS = ||\,\mathbf{x} - \Psi\mathbf{w}\,||^2. \end{gathered} \tag{17}$$

Here $\mathbf{x}$ is described by a linear model M as $\mathbf{x}=\sum_{i=1,k} w_i \boldsymbol{\psi}_i$, where $\mathbf{w}\in\Re^k$ are model parameters, $\boldsymbol{\psi}\in\Re^L$ are basis functions selected in process of constructing M and used to approximate $\mathbf{x}\in\Re^L$, L is the number of data samples used to construct the model, k is the number of parameters, ψ_{ij} compose matrix $\Psi\in\Re^{L\times k}$.

The developed cost function demonstrated robust performance in the BSS tasks in the presence of noise as well as a higher signal to noise ratio compared to FastICA at small level of noise and to PCA in the whole range of tested noise values [Revunova, 2007a, 2007b]. We believe that application of the BSS methods will increase the quality of nuclide identification.

Research of combined blind source separation approach and sparse approximation with model selection approach to gamma-ray spectroscopy looks promising (Figure 82).

5.3 Machine learning with regularization for airborne gamma-ray surveying

Airborne gamma-ray spectrometry is a remote sensing technique that reconstructs the surface density of radioactive sources. It is an important part of remote environmental radioactivity monitoring and is used for the direct detection of ore bodies and as a lithological mapping tool, assessing health risks associated with radon in houses, ground water discharge and salinity studies, soil mapping, the mapping of fallout from nuclear accidents, etc. [IAEA, 2003].

Airborne gamma-ray spectrometry measures gamma-ray emissions from radioactive isotopes within the earth's near surface (<0.5 m) or from the surface itself. Generally, 512-channel spectra are collected and analyzed. Each radiometric measurement represents a complex average of a relatively large region, with areas closer to the aircraft contributing more than those do farther away. Hence, individual features on the ground appear blurred in the aircraft image. This blurring of spatial detail can be removed partially by intelligent data processing, the quality of the result depending on the noise level within the data, the sampling density, and the mathematical model describing the smoothing process [Billings and Hovgaard, 1999], [Billings et al, 2003], [Dickson, 2004].

In this section, we consider two reconstruction techniques for. First, we observe best available convolution-based signal processing approach and its solution using the Fourier transform. Then, we consider an alternative machine learning method that we develop based on direct regularized solution to the ill-posed discrete problem in space domain. We get a stable solution based on

random projections, pseudoinverse, and show that its accuracy is comparable to the Tikhonov regularization, but is less expensive computationally.

5.3.1 Airborne survey technique

Airborne gamma-ray surveying is done as follows. The aircraft, a plane ("fixed wing") or helicopter that carries a gamma detector (Figure 87) moves over the target area with a given topographic relief (topography). Airborne geophysical surveys are normally flown on a regular grid along parallel lines ("flight lines"). The height and line spacing of a radiometric survey determine the spatial and spectral resolution that can be achieved. The line spacing (generally 100–400 m) controls the sampling density in the across-line direction, which can be significantly less than the sampling density along the line (generally 50–70 m). The footprint of a single measurement increases with flying height, while the count level decreases, thus reducing both spatial and spectral resolution. Ideally, to maximize both spatial and spectral resolution, a survey should be flown at the lowest possible safe altitude. Since gamma radiation is completely absorbed by the air layer of 200-300 m, surveys are typically flown at a constant height above the ground of between 40 m and 100 m, with helicopters able to fly considerably lower than most fixed-wing aircraft. There is obviously a trade-off in data acquisition between the observed count rates, and hence the accuracy of the measurements, and sampling time, aircraft speed, and spatial resolution of the data. Survey grids and traverse spacing should reflect the expected strength, size and distribution of sources [Billings & Hovgaard, 1999], [IAEA, 2003].

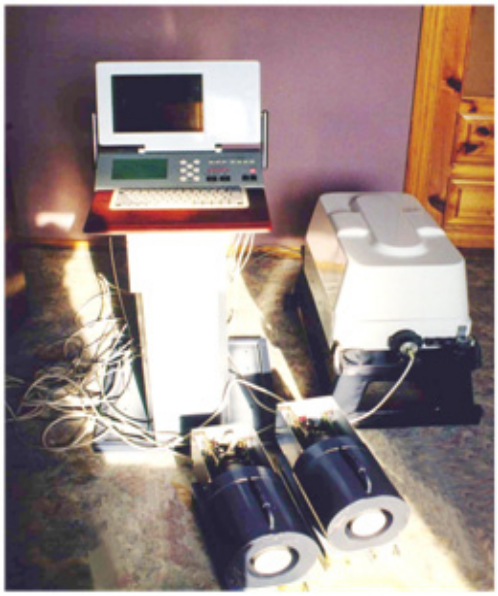

Figure 87. Onboard airborne gamma-ray system

Airborne gamma-ray spectrometric data can be severely distorted in rough terrain as follows: the radiation signal decays at an approximately exponential rate with increasing distance to the source; the rate of decay of radiation signal with distance depends on the source geometry; the resolution of discrete sources decreases with aircraft height; terrain clearance varies from line to line; the source geometry varies depending on the aircraft height and the severity of the topography; the geometry of the detector may have a directional bias.

Gamma-ray data are collected at height, and the detector cannot be focused like a camera. This results in a blurring of spatial detail, which gets worse as the height increases.

The height of the (x, y) point relative to some zero level is h (x, y). The time interval during which the gamma-ray count is accumulated before its recording is called the exposure time. Gamma-ray spectrometric data are usually acquired over a sample interval of one second. Thus, each flight line results in the array $F(t_i)$, $i=1,\ldots,N^d$ of detector readings. The carrier coordinates $(x^d(t),y^d(t))$, topography h(x,y), and height $h^d = h^d(x^d(t),y^d(t))$ are considered known.

Survey results are usually reported in units of gamma-ray dose rate (total-count surveys) or concentrations of the radioelements (spectrometer surveys). Measured count rates are dependent

not only on the ground radioelement concentrations, but also on the equipment used, and on the nominal height of the survey. This is undesirable, as measuring units should have a direct significance and be independent of the instrumentation and survey parameters. Count rates should therefore be converted to ground concentrations of the radioelements.

5.3.2 Problem formulation and intelligent data processing techniques

The task is to restore the radionuclide surface density function g(x,y) by the detector readings $F(x^d,y^d)$. Due to the large angular aperture of the collimator, the measurement results (detector readings, count rates) are smoothed and do not reflect in detail the surface density pattern (Figure 88).

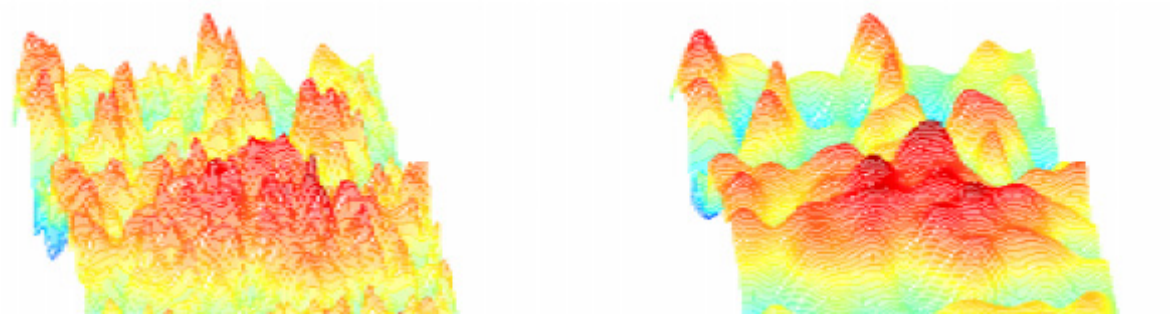

Figure 88. The surface density function (left) and its measurements (right)

Let us consider intelligent information processing techniques for reconstruction of the surface density. The detector count rates are related to the unknown surface density by a Fredholm integral equation of the first kind. For the case of stationary detector placed above the point (x^d,y^d) at the height h^d this equation is

$$\iint dx\, dy\, g(x,y)\, K(x^d,y^d,h^d,x,y,h) = f(x^d,y^d,h^d), \quad (18)$$

where $f(x^d,y^d,h^d)$ is the number of gamma-rays counted per time unit, g(x,y) is the surface density expressed in the number of decays per square meter per second, h(x,y) is the topography altitude at the point (x,y), $K(x^d,y^d,h^d,x,y,h)$ is the kernel function (also known as "point spread function" [Billings et al, 2003]) describing the smoothing process. The aircraft movement can be easily taken into account given solution to Eq.(18) [Billings et al, 2003], [Zabulonov et al, 2006].

Recent progress in the processing of airborne gamma-ray data was connected with multichannel processing techniques. For example, NASVD [Hovgaard and Grasty, 1997], [Minty and McFadden, 1998], [Minty and Hovgaard, 2002], [Mauring and Smethurts, 2005] and MNF [Green et al, 1988], [Dickson and Taylor, 1998; 2000], see also [Dickson, 2004], [Ramos et al, 2007], use information contained in the whole 512-channel spectrum to reduce the noise and improvement the quality of measured data f and thereby in the sought-for g.

Assuming small variations in survey altitude and topography (h=const) the kernel (point spread function) of the integral equation Eq.(18) is presented as

$$K(d,c,h) = K(d-c,h), \quad (19)$$

where $d=(x^d,y^d)$, c=(x,y), so that Eq.(18) becomes the convolution equation.

In [Billings and Hovgaard, 1999] the kernel is expressed as

$$K(d-c) = D(d-c,h)\exp(-\mu r)\, C/r^3, \quad (20)$$

where C is a constant that can be determined by standard calibration procedures, $r = ((x^d-x)^2 + (y^d-y)+h^2_{const})^{1/2}$ is source-detector distance; μ is the linear attenuation coefficient of the air, and D incorporates the detector response.

The exponential term accounts for attenuation of gamma-rays in the air, a factor $1/r^2$ accounts for geometrical dispersion, and 1/r arises from attenuation within the earth (with its own attenuation coefficient). The detector model is derived from geometrical arguments. [Gunn, 1978], [Craig, 1993], [Craig et al, 1999] present a point spread function for deconvolving airborne radiometric data that neglects aircraft movement and assumes that the detector response is non-directional. The point spread function, which take into account aircraft movement and the detector characteristics, is considered in [Billings and Hovgaard 1999], [Billings et al, 2003].

In deconvolution approach, solution of Eq.(18-20) is obtained in the frequency domain using the Fourier transform (Figure 89). In the Fourier domain, Eq (18) has a very simple form

$$W(\mathbf{u}) = K(\mathbf{u})\, G(\mathbf{u}), \quad (21)$$

where and $\mathbf{u} = (u_x, u_y)$ is the spatial frequency, and upper case means Fourier transformation.

A naive way to estimate $G(\mathbf{u})$ is by spectral division

$$G(\mathbf{u}) = W(\mathbf{u}) / K(\mathbf{u}). \quad (22)$$

However, $W(\mathbf{u})$ is not known exactly and at high frequencies it is dominated by noise. Since $K(\mathbf{u})$ decreases rapidly with frequency the spectral division amplifies noise. So, the reconstruction usually uses the Wiener filter $V(\mathbf{u})$:

$$G(\mathbf{u}) = W(\mathbf{u})\, V(\mathbf{u}). \quad (23)$$

The quality of the resulting deblurring of spatial detail depends on the noise level within the data, the sampling density, and the kernel describing the smoothing process. The improvement achieved by the deconvolution is limited by the noise levels in the data and the degree of smoothing imposed by the detector height.

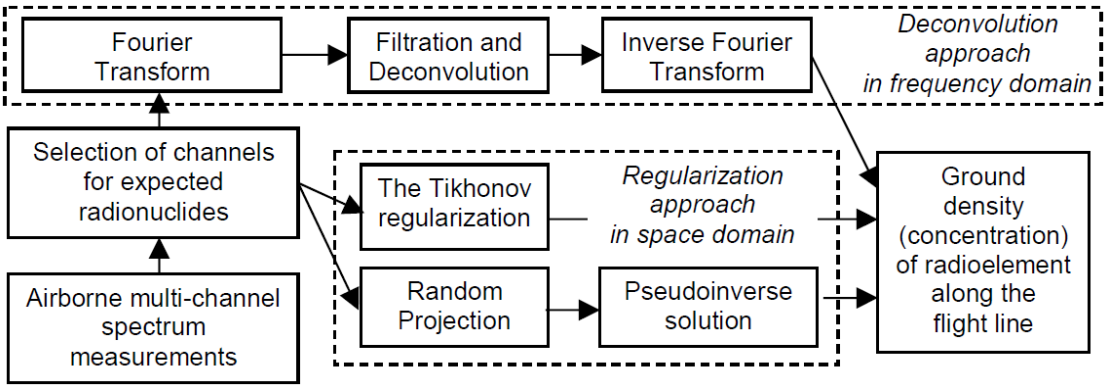

Figure 89. Deconvolution and regularization machine learning approaches to airborne gamma-ray surveying

The deconvolution approach has the following drawbacks. Derivation of the Wiener filter requires knowledge of spatial statistics of both the actual distribution density g(x,y) and the measurement noise $\varepsilon(x,y)$. Since both are unknown, some assumptions should be made, that may be not always

valid. For example, determining the noise levels after multichannel processing such as NASVD, and particularly after application of the adaptive filter, is difficult. So, in [Billings et al, 2003], after getting a ballpark estimate of the noise, manual adjustments are used to visually tune the deconvolution.

Also, the blurring in the data is assumed to be the same everywhere. This means that no accommodation can be made for changes in the detector height or for 3D terrain effects.

The noise is assumed additive and the same everywhere. Due to the Poisson nature of radioactive decay and gamma-ray detection, this is clearly not the case. To mitigate this effect, sometimes an adaptive 2D Lee filter [Ristau and Moon, 2001] is used to remove random noise from an image while maintaining image edges.

In situations where the terrain clearance (variations in survey altitude and topography) and noise vary significantly over the survey area, a partial solution would be to break the survey up into smaller segments and apply deconvolution separately to each sub-area.

An alternative approach [Zabulonov et al, 2006] is to use a full space domain formulation. For a particular example, let us consider a detector whose field of view is defined by a vertically oriented collimator of the horn type with a rectangular cross section and angular aperture (θ_x,θ_y), e.g. $0.5\theta_x = 0.5\theta_y \le 60°$. Therefore, the detector's field of view looks at a square piece of the surface topography, called "window" hereafter. The window size is determined by the angular aperture of the collimator and the height of detector. The window movement during the exposure time makes a rectangular exposure strip.

This setup is described by Eq.(18), where integration is over the area of the browsing window on the ground surface S^d, and

$$K(x,y, x^d,y^d) = s\, q(x^d,y^d,x,y) \exp(-\mu\rho)/4\pi\rho^2, \quad (24)$$

where $q(x^d,y^d,x,y) = (h^d-h(x,y))/\rho$, $\rho = ((h^d-h(x,y))^2 + (x^d-x)^2 + (y^d-y)^2)^{1/2}$ is the distance from the detector to the surface point, $q(x^d,y^d,x,y)$ is the detector sensitivity function. s is the effective area of the detector inlet, including the calibration constant, and taking into account, for particular radionuclide, the quantum yield, the detection efficiency, and the inlet area. Here, only surface radiation sources resulted from fallout are taken into account.

The major drawback of this full space domain formulation of the airborne gamma-ray survey problem is that it requires solution of discrete ill-posed inverse problem. Below, we consider an approach to a stable solution of this problem.

5.3.3 Regularized solution of the ill-posed inverse problem for a full spacedomain formulation

After discretization of the Fredholm integral equation Eq.(18), we get

$$\mathbf{Ax}=\mathbf{b}, \quad (25)$$

where the matrix $\mathbf{A}\in\Re^{N\times N}$ and the vector $\mathbf{b}\in\Re^N$ are obtained from discretization of the kernel K and $f(x^d)$, respectively; **x** represents the unknown g(x,y=const) along the flight line to be reconstructed. The vector $\mathbf{b}\in\Re^N$ is distorted by the additive noise $\boldsymbol{\varepsilon}\in\Re^N$: $\mathbf{b}=\mathbf{b}_0+\boldsymbol{\varepsilon}$.

In case when the singular values σ_i of **A** decay gradually to zero, the ratio between the largest and the smallest nonzero singular values is large, the problem is known as discrete ill-posed problem [Hansen, 1998]. Actually, these properties are present in the spectrometry task considered here.

Approximate solutions of discrete ill-posed problems as the least squares problem

$$\mathbf{x}' = \text{argmin}_{\mathbf{x}} ||\mathbf{b} - \mathbf{A}\mathbf{x}||_2, \quad (26)$$

using standard methods of numeric linear algebra such as LU, Cholesky, QR factorizations are unstable. It means that small perturbations in the input data lead to large perturbations in the solution.

To overcome the instability and to improve the accuracy of solutions to discrete ill-posed problems, regularization methods have been proposed [Hansen, 1998], [Tikhonov and Arsenin, 1977], [Morozov, 1984], [Engl et al, 2000]. In the fields of machine learning and inverse problems, regularization involves introducing additional information in order to solve an ill-posed problem or to prevent overfitting. This information is usually of the form of a penalty for complexity, such as restrictions for smoothness or bounds on the vector space norm. Regularization imposes some constraints on the desired solution that stabilizes the problem and leads to meaningful and stable solution. For example, Tikhonov regularization [Tikhonov and Arsenin, 1977], [Hansen, 1998] penalizes solutions with large l_2-norms.

The standard form of Tikhonov regularization problem is formulated as follows:

$$\mathbf{x}_{reg} = \text{argmin}_{\mathbf{x}} (||\mathbf{A}\mathbf{x} - \mathbf{b}||_2 + \lambda ||\mathbf{x}||_2), \quad (27)$$

where λ is the regularization parameter.

Regularization has a rich history, which dates back to the theory of inverse ill-posed and ill-conditioned problems inspiring many advances in machine learning, support vector machines and kernel based modeling techniques. Determination of the regularization parameter in the Tikhonov scheme is considered an important problem.

Solution of Eq.(27) can be obtained by the method of filtered SVD [Hansen, 1998]:

$$\mathbf{x}_{reg} = \mathbf{V}\ \text{diag}\ (f_i / \sigma_i)\ \mathbf{U}^T \mathbf{y}, \quad (28)$$

where $f_i = \sigma^2_i / (\sigma^2_i + \lambda^2)$ are filter factors.

One problem here is using the SVD of **A**, since it is a computationally expensive decomposition method. Another problem, that may appear even more important, is selecting the proper regularization parameter λ.

A number of methods for selecting the regularization parameter have been proposed [Engl et al, 2000]. The L-curve method [Hansen and O'Leary, 1993] makes a plot of $||\mathbf{x}_{reg}||_2$ vs $||\mathbf{A}\mathbf{x}_{reg} - \mathbf{b}||_2$ for all valid regularization parameters. For discrete ill-posed problems the L-curve, when plotted in log-log scale, often has a characteristic L-shape appearance, hence its name. A distinct corner separates the vertical and the horizontal parts of the curve. The regularization parameter not far from the corner is selected as optimal.

The discrepancy principle [Morozov, 1984] chooses the regularization parameters such that the residual norm for the regularized solution satisfies $||\mathbf{A}\mathbf{x}_{reg} - \mathbf{b}||_2 = ||\mathbf{e}||_2$, where **e** in the norm of perturbation of the right-hand side, and therefore requires an estimation of $||\mathbf{e}||_2$, i.e. noise.

The generalized cross-validation [Wahba, 1990] method is based on the idea that an arbitrary element b_i of the right-hand side **b** can be predicted by the corresponding regularized solution, and the choice of regularization parameter should be independent of an orthogonal transformation of **b**. This leads to choosing the regularization parameter that minimizes $||\mathbf{A}\mathbf{x}_{reg} - \mathbf{b}||^2 / D^2$, where D is a squared effective number of degrees of freedom (which is not necessarily an integer) that can be calculated as $D = m - \Sigma_i f_i$. Here the errors of **b** are considered as uncorrelated zero-mean random variables with a common variance, i.e., white noise.

✓ Solutions using random projections

The drawbacks inherent in the methods of solving discrete ill-posed problems based on Tikhonov regularization include their high computational complexity and the difficulty of selecting the proper regularization parameter (penalty weight) which influences the solution stability. At the wrong values of the regularization parameter, the error of solution may be substantial. Therefore, alternative approaches are required for solving discrete ill-posed problems that would have the accuracy comparable to Tikhonov regularization at lower computational costs.

We develop such an approach using the ideas of random projections [Johnson and Lindenstrauss, 1984], [Papadimitriou et al, 2000], [Achlioptas, 2003] and of our previous work on distributed representations (e.g., [Misuno et al, 2005]) inspired by the idea of information representation in neural networks of the brain. Random projections have recently appeared as a tool for dimensionality reduction and have been used to produce a number of interesting results, both theoretical and applied, including those in context of inductive supervised learning using machine learning methods [Halko et al, 2009], [Revunova and Rachkovskij, 2009s].

The technique plays a key role in several breakthrough developments in the field of algorithms. In other cases, it provides elegant alternative proofs. Recently, researchers working in the area of numeric linear algebra applied similar ideas to get fast randomized algorithms for the least squares problem, matrix factorization, principal component analysis, etc. [Sarlos, 2006], [Drineas et al, 2007], [Rokhlin and Tygert, 2008], [Tygert, 2009], [Halko et al, 2009]. It is therefore of interest to study those techniques and apply them to discrete ill-posed inverse problems.

Let us use the randomized algorithms not only to accelerate, but also to stabilize the solution **x'** of the ill-posed problem, as follows [Revunova and Rachkovskij, 2009s]. Multiply both sides of Eq.(25) by the matrix $\mathbf{\Omega} \in \Re^{K \times N}$, K≤N, whose elements are realizations of a normal random variable with zero mean and unit variance. The number of columns N of matrix $\mathbf{\Omega}$ is determined by the dimension of the matrix **A**, the number of rows K is a priori unknown since the numerical rank of **A** is ill-determined and the required numerical rank of approximation is unknown. We obtain

$$\mathbf{\Omega A}\, \mathbf{x} = \mathbf{\Omega b}, \text{ where } \mathbf{\Omega A} \in \Re^{K \times N}, \mathbf{\Omega b} \in \Re^{K}. \tag{29}$$

Then the least-squares problem is

$$\mathbf{x}_{Pr} = \operatorname{argmin}_{\mathbf{x}} \| \mathbf{\Omega A}\, \mathbf{x} - \mathbf{\Omega b} \|_2. \tag{30}$$

Signal reconstruction based on pseudo-inverse using random projection $\mathbf{\Omega}$ is obtained as

$$\mathbf{x}_{pinPr} = (\mathbf{\Omega A})^{+}\, \mathbf{\Omega b}. \tag{31}$$

Signal reconstruction based on the pseudo-inverse using the projection matrix **Q** obtained by the QR factorization of $\mathbf{\Omega A}$ is done as

$$\mathbf{x}_{pinQ} = (\mathbf{Q}^{T}\mathbf{A})^{+}\, \mathbf{Q}^{T}\mathbf{b}. \tag{32}$$

The pseudo-inverse $\mathbf{P}^{+}$ of a matrix **P** is actually computed based on SVD as:

$$\mathbf{P}^{+} = \mathbf{V}\, \mathrm{diag}\,(\phi_i / \sigma_i)\, \mathbf{U}^{T}, \text{ iff } \sigma_i > \mathrm{tresh}\ \phi_i = 1, \text{ otherwise } \phi_i = 0.$$
$$\mathrm{tresh} = \max(K,N)\, \mathrm{eps}(\max(\sigma_i)), \tag{33}$$

where **U**, **V, S** are obtained by the SVD of $\mathbf{P} = \mathbf{USV}^{T}$; $\sigma_i = \mathrm{diag}\ \mathbf{S}$ are singular values, the elements of a diagonal matrix **S**; floating-point relative accuracy eps(z) is the positive distance from abs(z) to the next larger in magnitude floating point number of the same precision as z.

In order to compare the quality of reconstruction of the surface density of radioactive sources, we conducted an experimental study of techniques for solving discrete ill-posed problems using the simulated data of radioactivity monitoring at h=100m.

The matrix $\mathbf{\Phi}$ (i.e., **A** in Eq.(25)) obtained by discretization of the kernel Eq.(24) has dimensionality of 200×200 (Figure 90), large condition number ($\sigma_{max}/\sigma_{min}>>1$), and singular values σ_i gradually decaying to zero. The right-hand side **y** (i.e., **b** in Eq.(25)) is distorted by an additive noise with the Gaussian distribution and various amplitudes. Figure 91 shows the measurement **y** produced by the doublet signal **x** to be restored.

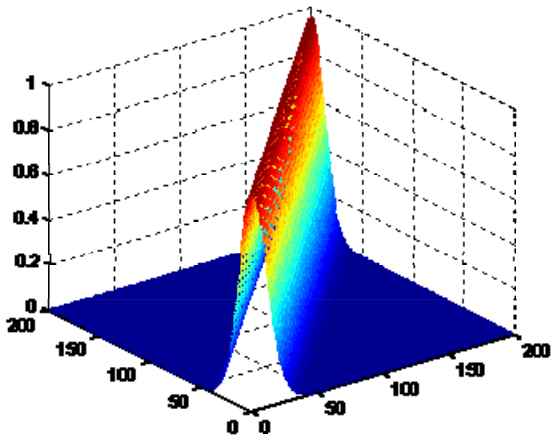

Figure 90. The kernel matrix Φ for gamma-ray surveying

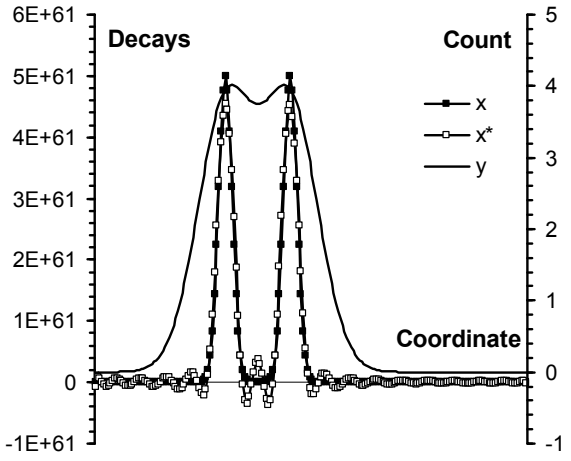

Figure 91. The modeled 1D surface density x, the measurement y, and the restored density x for gamma-ray surveying*

First, we have compared performance of three kinds of techniques: the pseudo-inverse solution Eq.(26), the Tikhonov regularization Eq.(28) with the selection of the regularization parameter λ using the L-curve, the generalized cross-validation, and the discrepancy principle [Hansen, 1998], [Wahba, 1990], [Morozov, 1984], and the pseudo-inverse with projection using the random matrix $\mathbf{\Omega}$ Eq.(31) and with the ortogonalized matrix **Q** Eq.(32). The random projection matrix $\mathbf{\Omega}\in\Re^{K\times N}$, N=200, K≤N is the Gaussian random matrix with the entries of zero mean and unit variance. Three noise levels are used $\{10^{-2},10^{-6},10^{-10}\}$. The results in terms of the signal reconstruction error vector norm are given in Table 4.

Without projection, the standard pseudo-inverse provides an acceptable error e_{Pin} only at the lowest noise level and cannot be used at all at the larger noise levels due to very large error. The errors for the Tikhonov regularization with the selection of λ by the L-curve e_{RegT} Lcur, the generalized cross-validation e_{RegT} GCV, and the discrepancy principle e_{RegT} Dsc are generally comparable for all three noise levels. However, we observe an outlier for e_{RegT} GCV at the noise level 10^{-6}, giving the instance of unstable regularization parameter selection. The lowest signal reconstruction error is provided by the Tikhonov regularization with the L-curve.

Table 4. The signal reconstruction error for the discrete ill-posed problem of gamma-surveying obtained by the regularization methods without and with random projecting. K is the dimensionality of projection matrix.

Noise level	Error, solutions without projection				Min error, with projection		Error, with projection		Error, with projection	
	e_{Pin}	e_{RegT} Lcur	e_{RegT} GCV	e_{RegT} Dsc	e_{PinR} (K)	e_{PinQR} (K)	e_{PinR} gMDL(K)	e_{PinQR} gMDL(K)	e_{PinR} Lcur(K)	e_{PinQR} Lcur(K)
10^{-2}	$7.3\ 10^{8}$	48.9	47.3	64	53.8(26)	50.3(29)	53.8(26)	51.3(27)	53.8(26)	51.9(30)
10^{-6}	$7.3\ 10^{4}$	21.1	$1.6\ 10^{3}$	22.8	21.9(35)	21.7(39)	21.9(35)	21.8(35)	22.0(36)	21.7(39)
10^{-10}	11.1	9.4	11.1	15.5	9.01(77)	8.3(47)	21.9(35)	12.5(45)	N/A	N/A

With the random projection, the pseudo-inverse becomes stable and the reconstruction error values at the minimum (with the proper choice of K) are comparable and small for the projector matrix $\boldsymbol{\Omega}$ (Eq.(31)) e_{PinR} and **Q** (Eq.(32)) e_{PinQR}.

In order to see how to choose the optimal K for the projective methods, let us consider the dependence of the signal reconstruction error e_{PinQR} on the dimensionality K of the projector matrix **Q** (Figure 92, left Y-axis). The error minimum can be observed for all noise levels. With increasing noise level, position of the error minimum shifts to the lower values of K and the error value at the minimum increases. To choose the projection matrix dimensionality K at which the solution error is close to the minimum for real situations, i.e. where the exact solution is unknown, we propose using various types of model selection criteria:

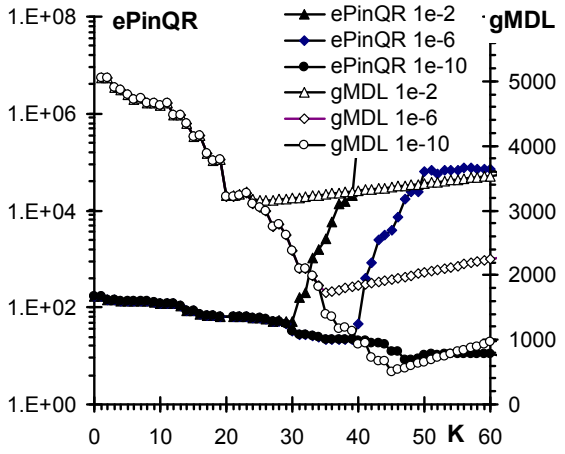

Figure 92. Dependence of the signal reconstruction error e_{PinQR} (left Y-axis) and gMDL values (right Y-axis) on the dimensionality K of the projection matrix at the noise levels 10^{-2}, 10^{-6}, 10^{-10}. Minima are at close K

(1) Criteria for model selection discussed above in context of sparse approximation and matching pursuit [Mallows, 1973], [Akaike, 1974], [Hansen and Yu, 2001];

(2) Criteria proposed to select the regularization parameters of Tikhonov-like regularization using the L-curve, the generalized cross-validation, and the discrepancy principle [Hansen, 1998], [Wahba, 1990], [Morozov, 1984].

Figure 92 (right axis Y) shows the values of Bin Yu model selection criterion gMDL Eq. (11) vs K. We observe that the minimum value of gMDL is near the true solution minimum. The results of K selection using gMDL, as well as those obtained with the L-curve, are shown in Table 4 and demonstrate the reconstruction error near or equal to the optimal one. An example of the reconstructed result **x*** is shown in Figure 91. So, the model selection criteria can be used for getting the required K providing the near-optimal solution.

Figure 93 shows the computation times of different techniques vs. K. As expected, the solution based on pseudo-inverse without random projection has a constant and rather large computation time compared to the solutions using random projection at the smaller values of K. For example, at K=30, the computation time with random projection (~0.003s for both considered projection methods) is 20 times less than that of ordinary pseudoinverse without random projection (0.06s). This time reduction is because the singular value decomposition is performed on the resulting (N×K) matrix after the projection, where K is a small fraction of N of the original (N×N) matrix Φ. Particularly large gains are achieved at the higher noise levels because their optimal K is small.

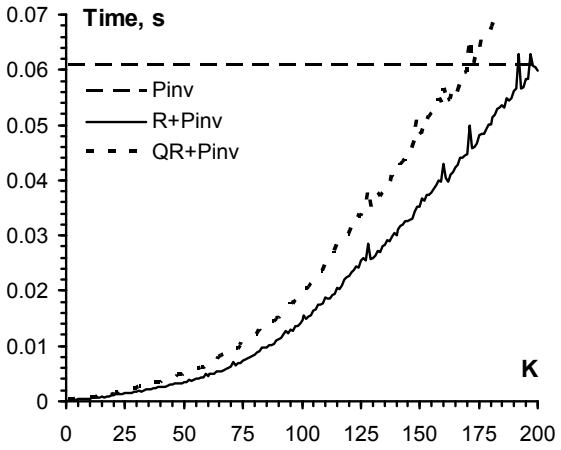

Figure 93. The computation times of pseudo-inverse solutions vs K: Pinv – ordinary, R+Pinv – with random projection, QR+Pinv – random projection orthogonalized

Thus, the study and application of intelligent regularization-based techniques for solving discrete ill-posed problems based on pseudo-inverse with a random projection is a promising direction due to the lower computational costs, and also because of their stability, manifested in the smooth change of the signal reconstruction error with the increasing noise and K.

Directions of further work in this area include:

- solution methods that take into account nonnegativity of solution, such as non-negative matrix factorization, non-negative least squares, etc.;
- techniques for a computationally efficient choice of the optimal K based on the knowledge about solution, such as its smoothness, nonnegativity, etc.;
- high-performance hardware implementation for real-time applications using efficient systolic architectures [Brent, 1988], [Brent and Luk, 1985], [Zabulonov and Revunova, 2006].

5.4 Discussion

In this chapter, we discussed machine-learning techniques for intelligent processing of gamma-ray spectroscopy data in environmental and security-related monitoring.

In particular, we developed and applied to intelligent processing of multichannel gamma-ray spectrometer data the supervised learning techniques of sparse approximation with model selection that resulted in a substantial increasing of data processing quality. Development of a new model selection test is suggested for further improving the quality of gamma-ray spectrometry. For this task, we also considered blind source separation approach of unsupervised learning, including a novel objective function based on algorithmic complexity. It is expected to become an important step in the development of the new generation of intelligent information technologies for multichannel and multidetector gamma-ray spectroscopy under a complex and unknown geometry of measurements.

We also described an advanced approach to processing of airborne gamma-ray spectrometry data in order to reconstruct the surface density of radionuclides. This approach requires solution of the discrete ill-posed inverse problem resulting from direct space domain problem formulation. In order to improve the speed and stability of the Tikhonov regularization traditionally used to solve such problems, we developed and employed a novel method based on random projection, which is conceptually close to distributed representations in neural networks. This resulted in substantial computational savings for noisy measurements, which is always the case. Again, model selection criteria were employed to choose the random projection parameter.

We would like to note that the developed and deployed intelligent data processing techniques are not limited to the case study of particular gamma-ray processing task considered. They can be applied to other kinds of environmental gamma-ray monitoring, as well as to other kinds of environmental monitoring in many areas beyond those considered in this chapter – those that require machine learning techniques for solving inverse problems that are ill-conditioned or ill-posed and call for model selection and regularization.

In addition to the machine learning methods considered in this chapter, another way to increase the level of intelligence in monitoring for environment and security is construction of and integration with knowledge bases and expert systems, particularly those employing analogical reasoning and knowledge mining [Gladun 1994], [Gladun and Vashchenko, 2000], [Gladun et al, 2008], [Markman et al, 2003], [Rachkovskij, 2001; 2004], [Rachkovskij and Kussul, 2001], [Slipchenko and Rachkovskij, 2009s]. This requires formation of data and knowledge bases and other intelligent memory structures accumulating the measurement results together with the results of their processing for varied monitoring situations, as well as with their expert estimations. Those bases could be used both for machine learning and for example-based reasoning in order to further raise the intelligent component of gamma-ray monitoring and of other kinds of environmental and security-related monitoring. An advanced access to available and emerging worldwide, national, and local databases, and their computation-intensive processing would benefit from the high-performance intelligent computations and multi-source data integration of geographically distributed Grid computing technologies.