Multi-spectral inverse problems in satellite image processing

S. A. Starks and V. Kreinovich

NASA Pan-American Center for Earth and Environment Studies (PACES),
University of Texas at El Paso, El Paso, TX 79968, USA

ABSTRACT

Satellite imaging is nowadays one of the main sources of geophysical and environmental information. It is, therefore, extremely important to be able to solve the corresponding inverse problem: reconstruct the actual geophysics- or environment-related image from the observed noisy data.

Traditional image reconstruction techniques have been developed for the case when we have a single observed image. This case corresponds to a single satellite photo. Existing satellites (e.g., Landsat) take photos in several (up to 7) wavelengths. To process this multiple-spectral information, we can use known reasonable multi-image modifications of the existing single-image reconstructing techniques. These modifications, basically, handle each image separately, and try to merge the resulting information.

Currently, a new generation of imaging satellites (Lewis) is being launched, that will enable us to collect visual images for about 500 different wavelengths. This two order of magnitude increase in data amount should lead to a similar increase in the processing time, but surprisingly, it does not. An analysis and explanation of this paradoxical simplicity is given in the paper.

Keywords: Inverse problem, multi-spectral inverse problem, satellite imaging

1. SATELLITE IMAGING: AN IMPORTANT EXAMPLE OF INVERSE PROBLEMS

1.1. Satellite imaging

Nowadays, satellite imaging is one of the most important sources of geographical, geophysical, and environmental information. Satellite images can determine the amount and type of vegetation, the geological type of the underlying soils (and often, of the minerals below), etc.

However, with the current satellite images, it is sometimes difficult to decide what exactly we observe, because the existing Earth-sensing satellites, such as Landsat, only take the images at a few (≤ 7) frequencies.

1.2. An example of a problem in which a satellite image is currently not sufficient: kaolinite vs. dickite

Based on the (inevitably imprecise) measurements on the few frequencies, it is difficult, e.g., to distinguish between kaolinite and its rare amorphous but chemically similar forms such as dickite.

Kaolinite and dickite are the principle ingredients of *kaolin*, a soft white-clay mineral that is an essential ingredient in the manufacture of china and porcelain and is also widely used in the making of paper, rubber, paint, and many other products (see, e.g., Ref. 13). It is also used in medicine: e.g., in the treatment of diarrhea, kaolin powder is the most widely used absorbent powder. Due to kaolin's importance, it is desirable to determine not only its *presence*, but its *type* as well.

Since crystal-based kaolinite and amorphous dickite are chemically similar, their spectra are very similar. Therefore, currently, in order to distinguish between these two minerals, we have to complement satellite images with geophysical and radar data (see, e.g., Ref. 9,10).

E-mails {sstarks,vladik}@utep.edu

1.3. Hyper-spectral satellite imaging

To produce more data, NASA is planning to launch imaging satellites of the new generation, satellites that will have the ability to map the Earth on up to 500 optical frequencies. These coming satellites are nicknamed *Lewis* after the famous 19 century US geographer.

From the resulting multi-spectral images, it is, in principle, possible to determine many characteristics of soil and vegetation without using additional data; see Refs. 12,11. For example, it is, in principle, possible to distinguish between kaolinite and dickite because from hyper-spectral images, we can extract spectra in each point, i.e., the dependence of its brightness $I(f)$ on the frequency f; the corresponding spectra, although similar, have different number of local maxima.

2. MULTI-SPECTRAL SATELLITE IMAGING IS OFTEN SURPRISINGLY TOO EASY: TWO PARADOXES

2.1. Processing multi-spectral images should be hard

For the new satellites, the number of wavelengths is so huge that it becomes comparable with the numbers of vertical and horizontal pixels. We can, therefore, view wavelength as a third dimension, and use known techniques for solving 3D inverse problems to reconstruct the actual image.

It is doable, but the experience of 3D imaging in geophysics, in engineering (e.g., non-destructive evaluation of aerospace structures), shows that this is an extremely complicated and time-consuming task. The two order of magnitude increase in data amount should lead to a similar increase in the processing time. Surprisingly, it often does not.

2.2. First paradox: often, linear methods are paradoxically applicable, even in clearly non-linear problems

In many geophysical applications, even when the underlying equations are clearly non-linear, surprisingly, we often find that a simple linear model fits the data very well (see examples below).

2.3. Second paradox: in linear problems, non-linear regression methods (such as neural networks) often give an (approximate) answer faster than linear regression

According to the standard practice of statistical data processing, if we want to know how a certain quantity y depends on some other quantities $x_1, \ldots, x_n$, we first try *linear regression*, i.e., we first try to fit the model by a linear dependence. Linear regression methods use standard linear algebra and matrix operations, and they are reasonably fast.

If it is not possible to fit the observations into a linear model, then we try non-linear regression methods; for example, we can try to fit the data by a higher-order polynomial, or by a function from a more complicated class. A practically useful class of highly non-linear regression models is provided by *neural networks.* As we go from linear to non-linear models, methods become more complicated and time-consuming. For example, the time that a neural network takes to find the fitting parameters for a non-linear model is often several orders of magnitude higher than the time that a linear regression algorithm would take to find the parameters of a linear model of the same dimension (with the same number of parameters).

In general, this is true, but in some inverse problem (see, e.g., Ref. 5), a *non-linear* neural network leads to a data-fitting model much faster than *linear* regression! Why?

2.4. What we are planning to do

In this paper, we will analyze and explain these two paradoxes.

2.5. A comment on pessimism and optimism

Before we go into technical detail, let us make a brief comment. A paradox of this type can have two types of answers:

- A *pessimistic* answer: we accidentally run into a few problem that are unusually simple, but in general, problems of this type are tough.
- An *optimistic* answer: in general, the original pessimistic estimates of computation time were indeed too pessimistic, and multi-spectral satellite imaging problems (as well as other inverse problems) are, in general, mush easier than we originally thought.

In this text, we will give arguments in favor of the optimistic viewpoint.

3. WHY LINEAR METHODS WORK IN SEEMINGLY NON-LINEAR SITUATIONS: ANALYSIS AND EXPLANATION OF THE FIRST PARADOX

In this section, we will explain why linear methods work well. The surprise efficiency of linear methods is a very general phenomenon that occurs not only in the new field of hyperspectral image processing, but also in such well-established areas as geophysics, economics, etc. Let us therefore describe a general background.

3.1. Indirect measurements

In many real-life situations, we must estimate the value of a physical quantity y that is difficult to measure directly. So, to avoid direct measurements, we measure y *indirectly:* namely, we measure whatever variables we can, and then based on the measured value $x_1, \ldots, x_n$, we try to estimate y.

These situations are very frequent in geophysics, when it is very costly to measure the properties of the deep layers, and much cheaper to measure the waves reflected from those layers.

3.2. In many cases, we do not know the exact relation between x_i and y

In some situations, we know the relationship between x_i and y, so we can use this known model to estimate y based on x_i. In geophysics, this relationship is usually highly non-linear.

In many situations, however, this dependency between x_i and y is not a priori known, and has to be determined experimentally.

3.3. There exist many methods and algorithms that extract the dependency from the experimental data

There exist many statistical methods that help us to discover such a dependency (see, e.g., Ref. 4). To apply them, we must have several situations in which we know both y and x_i. So, we have the values $x_{11}, \ldots, x_{n1}, y_1$ that correspond to the first situation, the values $x_{21}, \ldots, x_{2n}, y_2$ that were measured in the second situation, ..., and, finally, the values $x_{N1}, \ldots, x_{Nn}, y_N$ (here, by N we denoted the total number of such situations). From these data, we extract a function f such that $y_k = f(x_{k1}, \ldots, x_{kn})$ for $1 \leq k \leq N$.

When this function is known, we can use it to estimate y from x_i for the new measuring situations.

The simplest of these methods (called *linear regression*) uncovers linear dependencies, i.e., dependencies of the type $y = a_0 + a_1x_1 + \ldots + a_nx_n$ for some constant coefficients a_i. The corresponding statistical software packages have built-in checks on whether the data is really consistent with linear dependency (e.g., χ^2-method).

Non-linear curve-fitting methods also exist.

For the cases when we are not sure whether the relationship is linear or not, traditional statistical methodology requires that we first try simple linear regression methods, and then, if linear methods do not work, try more complicated non-linear methods.

3.4. In geophysics, we expect non-linearity, but surprisingly, linear regression methods work fine

The equations that describe the geophysical phenomena are *highly non-linear.* So, we would expect that in the majority of cases, linear regression will fail, and we will have to use non-linear methods.

Unexpectedly, in many cases, linear methods succeed! Namely, they generate reasonable linear dependencies that fit the experimental data perfectly (fits in the sense that built-in checks of linearity confirm that linearity is possible). This phenomenon occurs in all areas of geophysics: in seismology,[16,1] in electrical methods,[14,2] in general geophysics.[8,1]

The same strange phenomenon occurs in economics: when we, e.g., analyze the dependency of the workers-per-manager ratio on the parameters that characterize a business,[7,6] we also get a pretty good fit for linear regression in an evidently non-linear situation.

3.5. Why?

Why are linear methods working so well in non-linear situations? In this paper, we present an answer to this question. This answer will consist of two parts:

- First, we will show that in many real-life situations, the variables x_i are not independent, they are functionally dependent on each other.
- Second, we will show that these dependencies enables us to use linear regression.

3.6. If there is a model with m parameters, then, in principle, it is sufficient to measure m quantities x_i

Before we start the general argument, let us first consider the case when we know the model of the object that we are analyzing.

In many situations, there is a model that describes the analyzed phenomena. Usually, a model has several (unknown) parameters. For example, a geophysical area is often described in terms of 3 or more layers, with about 3 parameters to describe the properties of each layer.

If we know the model, then, of course, we know the number m of parameters $p_1, \ldots, p_m$ of that model. By saying that we have a model we mean that we know how the desired value y depends on these parameters, i.e., we know a function g that computes y from p_j: $y = g(p_1, \ldots, p_m)$. We also know how all other measurable quantities x_i depend on p_j: $x_i = h_i(p_1, \ldots, p_m)$. Suppose that we measured m quantities. Then, we have m equations $h_i(p_1, \ldots, p_m) = x_i$, $1 \le i \le m$, with the known right-hand sides and m unknowns $p_1, \ldots, p_m$.

In general, when the number of equations coincides with the number of unknowns, we have a unique solution. Therefore, from x_i, we can uniquely determine the parameters p_j. And as soon as we know the parameters, we can compute the value of $y = g(p_1, \ldots, p_m)$. So, from $x_1, \ldots, x_m$, we can (in principle) uniquely determine y.

3.7. What if we do not know the model, but we do know that the object is uniquely determined by m parameters?

In this case, we do not know how to compute y from $x_1, \ldots, x_m$, but we still know that in principle, we can uniquely determine y from $x_i, 1 \le i \le m$. This is the case when regression methods have to be applied. So, we make several measurements of x_i and y in different situations, and try to find the dependency $y = f(x_1, \ldots, x_m)$.

3.8. If we use the smallest possible number of measurements, then we cannot apply linear methods

If we use exactly m measurements, and the actual function f is non-linear, then, of course, we cannot use linear regression methods.

3.9. In real life, we usually perform more measurements to increase precision

Since measurements are usually not ideally precise, the estimates that we get from the smallest possible amount of measurements $x_1, \ldots, x_m$ are also not precise. To make the estimates more precise, we perform additional measurements.

As a result, the number n of variables x_i that we measure is much greater than the smallest possible number m.

3.10. As a result, the variables x_i are inter-dependent

The same logic that showed that y is uniquely determined by m parameters $x_1, x_2, \ldots, x_m$, shows that any other physical characteristic of our object is also uniquely determined by $x_1, \ldots, x_m$. In particular, it is true for the quantities $x_{m+1}, x_{m+2}, \ldots, x_n$.

Therefore, $x_{m+1} = f_{m+1}(x_1, \ldots, x_m)$ for some function f_{m+1}, $x_{m+2} = f_{m+2}(x_1, \ldots, x_m)$ for some other function f_{m+2}, etc.

In other words, the variables x_i are *not* independent: there is a functional dependency between them.

3.11. Example

As an example, let us consider the simplest case when one parameter p is sufficient to determine the values of all the physical quantities. In this case, $y = g(p)$, and $x_i = h_i(p)$ for some functions g and h_i.

Since $m = 1$, we need only one measurement to determine p uniquely. If x_1 is known, then we can determine p as the solution of the equation $h_1(p) = x_1$. So, $p = h_1^{-1}(x_1)$, where by h_1^{-1}, we denoted a function that is inverse to h_1. Therefore, $y = g(p) = g(f^{-1}(x_1))$. Similarly, $x_2 = h_2(p) = h_2(h_1^{-1}(x_1))$, i.e., x_1 and x_2 are functionally dependent on each other. Likewise, x_1 and x_3, x_1 and x_4, etc, are mutually dependent.

Let us see what happens when the values x_i are inter-dependent.

3.12. Explanation of the first paradox for the simplest case of only one parameter p

Let's first consider the case when everything depends on only one parameter p.

As long as the dependency of y and x_i on p is smooth (and it usually is), we can expand the functions $y = g(p)$ and $x_i = h_i(p)$ into the Taylor series:

$$y = a^{(0)} + a^{(1)}p + \ldots + a^{(l)}p^l + \ldots$$

$$x_i = a_i^{(0)} + a_i^{(1)}p + \ldots + a_i^{(l)}p^l + \ldots$$

We measure both x_i and y with a certain precision ε. Therefore, if we get the value $\tilde{x}_i$ as a result of the measurement, it means that an actual value of x_i belongs to an interval $[\tilde{x}_i - \varepsilon, \tilde{x}_i + \varepsilon]$. Because of this imprecision, we can use the approximate formula for $x_i(p)$ as long as the approximation error does not exceed ε. So, instead of taking all Taylor terms, we can retain only those that guarantee the precision ε. As a result, we get the following formulas:

$$y = a^{(0)} + a^{(1)}p + \ldots + a^{(l)}p^l,$$

$$x_i = a_i^{(0)} + a_i^{(1)}p + \ldots + a_i^{(l)}p^l.$$

If we denote by N the total number of situations that we can use to determine the dependency of y on x_i, by $p_{(k)}$ the (unknown) value of the parameter p in k-th experiment ($1 \le k \le N$), then we get the following formulas:

$$y_k = a^{(0)} + a^{(1)}p_{(k)} + \ldots + a^{(l)}p_{(k)}^l,$$

$$x_{1k} = a_1^{(0)} + a_1^{(1)}p_{(k)} + \ldots + a_1^{(l)}p_{(k)}^l,$$

$$x_{2k} = a_2^{(0)} + a_2^{(1)}p_{(k)} + \ldots + a_2^{(l)}p_{(k)}^l,$$

$$\ldots$$

$$x_{nk} = a_n^{(0)} + a_n^{(1)}p_{(k)} + \ldots + a_n^{(l)}p_{(k)}^l$$

for all $k = 1, 2, \ldots, N$.

We can form $N-$dimensional vectors $\vec{x}_i = (x_{i1}, \ldots, x_{iN})$ and $\vec{y} = (y_1, \ldots, y_N)$. The above equalities can be reformulated in terms of these vectors as follows:

$$\vec{x}_i = a_i^{(0)}\vec{1} + a_i^{(1)}\vec{p} + \ldots + a_i^{(l)}\vec{p}^{\,l},$$

$$\vec{y} = a^{(0)}\vec{1} + a^{(1)}\vec{p} + \ldots + a^{(l)}\vec{p}^{\,l},$$

where we denoted

$$\vec{1} = (1, 1, \ldots, 1),$$

$$\vec{p} = (p_{(1)}, \ldots, p_{(N)}),$$

$$\ldots$$

$$\vec{p}^{\,j} = (p_{(1)}^j, \ldots, p_{(N)}^j),$$

$$\ldots$$

So, $n+1$ vectors $\vec{x}_1, \ldots \vec{x}_n, \vec{y}$ belong to a
$(k+1)-$dimensional space L: namely, to the linear space generated by $k+1$ vectors $\vec{1}$, $\vec{p}$, $\vec{p}^{\,2}$, …, $\vec{p}^{\,l}$.

When $n > l$, these vectors cannot be all linearly independent. Therefore, they are linearly dependent, i.e., there exist values c_i such that

$$c_1\vec{x}_1 + c_2\vec{x}_2 + \ldots + c_n\vec{x}_n + c_{n+1}\vec{y} = 0.$$

In general, all these coefficients c_i are different from 0. If we exclude the degenerate case when $c_{n+1} = 0$, then we can divide this equality by c_{n+1} (so that the coefficient at $\vec{y}$ will become equal to 1), and move all the terms except $\vec{y}$ to the other side of the equation. As a result, we get the formula

$$\vec{y} = a_1\vec{x}_1 + a_2\vec{x}_2 + \ldots + a_n\vec{x}_n,$$

where $a_i = -c_i/c_{n+1}$. If we now recall that each of these vectors $\vec{x}_i$ and $\vec{y}$ is actually formed by the measured values of the corresponding quantities in N situations, then from the above vector equality we can conclude that for all measurements $k = 1, 2, \ldots, N$, we have

$$y_k = a_1 x_{1k} + a_2 x_{2k} + \ldots + a_n x_{nk}.$$

In other words, for every measurement, the following equality holds:

$$y = a_1 x_1 + a_2 x_2 + \ldots + a_n x_n.$$

But this means exactly that *linear regression works.*

3.13. Example

Let us consider the case when $y = p + p^2$, $x_1 = p$, and $x_2 = p - p^2$. In this case, the variable y can be explicitly expressed in terms of x_1: $y = x_1 + x_1^2$. This expression is clearly non-linear, so, if we will try to apply linear regression to determine the dependency of y on x_1, the linear regression method will indicate failure.

However, if we try to represent y in terms of both x_1 *and* x_2, then we have a linear expression: $y = 2x_1 - x_2$. Indeed, $2p - (p - p^2) = p + p^2$.

3.14. Explanation of the first paradox: general case

In the general case, we have m parameters $p_1, \ldots, p_m$. So, when we restrict the Taylor expansion to the terms $p_1^{d_1} p_2^{d_2} \ldots p_m^{d_m}$ of power $d_j \le l$, we get $(l+1)^m$ different terms (because each of d_j can take any value from 0 to l). Therefore, we need $(l+1)^m$ coefficients to describe an approximation.

In this case, if we similarly introduce the vectors, we will have $n+1$ vectors $\vec{x}_i$ and $\vec{y}$ that are linear combinations of $(l+1)^m$ vectors $\vec{p}^{\,d_1, d_2, \ldots, d_m}$ with components $p_{1j}^{d_1} p_{2j}^{d_2} \ldots p_{mj}^{d_m}$, $1 \le j \le N$.

So, if the number of variables n is sufficiently large (in this case, if $n+1 > (l+1)^m$), then we can likewise conclude that the vectors $\vec{x}_i$ and $\vec{y}$ are linearly dependent, and therefore, *linear regression methods can be applied.*

3.15. Conclusion

If linear regression works in a non-linear situation, one does not need to search for an error. Moreover, if this is the situation, then we can be sure that the variables x_i are functionally inter-dependent, so we can look for the dependencies between them.

4. WHY ARE NON-LINEAR METHODS SOMETIMES FASTER THAN LINEAR ONES? ANALYSIS AND EXPLANATION OF THE SECOND PARADOX

4.1. Paradoxical situation: a brief and formal reminder

This paradoxical situation concerns a similar regression problem. We know the data $(x_{1k}, \ldots, x_{mk}, y_k)$, $1 \leq k \leq N$, and we must find out a *model* that describes this data, i.e., a function $f(x_1, \ldots, x_n)$ such that for all the measurement results (i.e., for all k from 1 to N), $y_k = g(x_{1k}, \ldots, x_{nk})$ (or, to be more precise, that the difference between y_k and $f(x_{1k}, \ldots, x_{nk})$ is within the measurement inaccuracy).

In general, if we know that the data is consistent with a linear model, then, in principle, we have two choices:

- first, we can apply standard methods of linear regression;
- alternatively, we can apply some more general non-linear regression techniques.

Common sense tells us that methods of linear regression, methods which were specifically designed to handle exactly this case of linear models, should work faster than more general methods. And indeed, in most situations, they do work faster. This is so well known that the only reason why people may apply non-linear methods to linear problems is to test the correctness of these non-linear methods. Researchers do these tests all the time. If the tests are successful, then all the researchers *usually* get as a result (in addition to one confirmation of the new method's correctness) is a very time-consuming method of solving an already-solved and easy-to-solve problem.

Usually – but not always. After analyzing mechanical and structural data obtained in the process of non-destructive evaluation of aerospace structures, the authors of Ref.[5] found out that not only their non-linear method (a special type of radial-based neural network regression) was correct, but that it actually produced a fitting model much faster than linear regression.

How come? To answer this question, let us analyze this situation.

4.2. Statistical analysis of this paradoxical situation

Traditional statistical linear regression methods are based not only on the assumption that the variable y is approximately equal to a linear function of the variables x_i. Another important assumption behind this technique concerns the probability distribution of the measurement errors in measuring x_i and y. For example, least-square-type methods are based on the assumption that all these errors are independent and normally distributed.

When we derive the formulas, we also implicitly assume that all these variables x_i are independent – this independence is non-statistical, meaning, e.g., that each of these variables (factors) can be changed without changing the others. This non-statistical independence can be formalized in purely statistical terms: the prior distribution on the set of possible values of $(x_1, \ldots, x_n)$ corresponds to these variables x_i being independent.

At first, we have this prior distribution. Then, after each observation of the values $x_1, \ldots, x_n$, we can use Bayes formula to improve this distribution. After several observations, we get a *posterior* distribution which better describes the actual distribution of $x_1, \ldots, x_n$.

In the examples from,[5] the resulting distribution is not only not independent, it starts concentrating along a (highly non-linear) surface of dimension $d \ll n$. In other words, it looks like the variables $x_1, \ldots, x_n$ are functionally dependent.

Thus, this situation is very similar to the situation uncovered in our analysis of the first paradox. We are one step behind the solution, but before we go to the solution of the second paradox, let us analyze the same paradoxical situation from a different viewpoint.

4.3. Analysis of the resulting fitting models

Another possible way to analyze this paradoxical situation is to compare the fitting models produced by linear regression and by a non-linear (neural networks) technique.

Of course, when we use a non-linear technique, we get, in general, a non-linear model, so we cannot expect this non-linear model $f(x_1, \ldots, x_n)$ to be identical to the linear model $a_1 x_1 + \ldots + a_n x_n$ produced by linear regression. However, since both models fit the same data, their values must coincide (or at least be fairly close) for all observations $(x_{1k}, \ldots, x_{kn})$, $1 \leq k \leq N$.

An interesting question is: are they still close if we take the tuples $x_1, \ldots, x_n$ that do not come from measurements? The answer is: *No*. The numerical values produced by these models are *different*, even if we compare these models on the values x_i which are somewhat close to one of the observed points.

In other words, while a linear model uncovers a linear dependence of the variable y on the data x_i, a neural network describes a *different* non-linear dependence of y on x_i. Since for all observed values, these two models coincide, this means that we have a functional dependence between the variables $x_1, \ldots, x_n$: namely, a functional dependence of the type $f(x_1, \ldots, x_n) = a_1 x_1 + \ldots + a_n x_n$.

Now, after this two-step analysis, we are ready to explain the second paradox.

4.4. Explanation of the second paradox

According to our analysis, in the paradoxical situation from,[5] we have exactly the same situation as in the first paradox: we have several variable $x_1, \ldots, x_n$ which are functionally inter-dependent by some non-linear dependencies, and it is exactly these non-linear dependencies which lead (according to our analysis of the first paradox) to the linear formula $y = a_1 x_1 + \ldots + a_n x_n$. Let us explain why in such a situation, finding a non-linear fitting model may be faster.

The main computational complexity of methods of linear regression is related to the necessity to solve a system of n linear equations with n unknowns, or to invert an $n \times n$ matrix. If we use the standard Gaussian elimination technique, we need $O(n^3)$ computational steps; there exist asymptotically faster algorithms which take time $O(n^{2+\alpha})$ for some positive $\alpha < 1$. However, for all known algorithms, we need at least $O(n^2)$ computational steps (see, e.g., Ref. 3).

Let us now assume, for simplicity, that all the variables x_i are functionally dependent of each other, in particular, that each of these variables is equal to a function of a variable x_1. Since the quantity y is, in its turn, functionally dependent on $x_1, \ldots, x_n$, we can conclude that y is also equal to a function of x_1. So, to find a model that fits, it is no longer necessary to consider all other variables x_i, it is quite sufficient to only analyze how y depends on x_1. To find a fitting formula for this dependence, it is no longer necessary to do any complicated computations: it is sufficient to sort the values x_1, and to make a linear extrapolation. Sorting n values takes time $O(n \cdot \log(n)) \ll O(n^2)$, and linear extrapolation is a linear-time algorithm (i.e., takes time $O(n)$).

Similarly, if, e.g., two variables (say, x_1 and x_2) are functionally independent, and all the others are functionally dependent on them), then y can be represented as a function of only these two variables. Therefore, we can describe a fitting model by sorting all the triples (x_{1k}, x_{2k}, y_k), $1 \leq k \leq n$, and computing a 2D spline which passes through these triples. The required computation time is still much smaller than $O(n^2)$.

4.5. Conclusion

If for some linear problem, a non-linear fitting technique work faster than linear regression, it does not necessarily mean that something is wring. It may as well be that in reality, in addition to the linear dependence $y = a_1 x_1 + \ldots + a_n x_n$ between x_i and y, there is a non-linear functional inter-dependence between the variables x_i; because of this dependence, there exists a simple and easy-to-find non-linear formula $y = g(x_{i_1}, \ldots, x_{i_k})$ which successfully predicts y based on the values of only some of the original variables x_i.

ACKNOWLEDGMENTS

This work was supported in part by NASA under cooperative agreement NCC5-209, by NSF grant No. DUE-9750858, by the United Space Alliance, grant No. NAS 9-20000 (P.O. 297A001153), and by the Future Aerospace Science and Technology Program (FAST) Center for Structural Integrity of Aerospace Systems, effort sponsored by the Air Force Office of Scientific Research, Air Force Materiel Command, USAF, under grant number F49620-95-1-0518.

The authors are thankful to Brian Penn and Wolfgang Sachse for valuable discussions.

REFERENCES

1. M. R. Baker, *Quantitative interpretation of geological and geophysical well data*, Ph.D. Dissertation, Department of Geological Sciences, University of Texas at El Paso, 1988.
2. D. E. Boerner and J. S. Holladay, "Approximate Frechet derivatives in inductive electromagnetic soundings", *Geophysics* **55**, pp. 1589–1595, 1990.
3. T. H. Cormen, C. E. Leiserson, and R. L. Rivest, *Introduction to Algorithms*, MIT Press, Cambridge, MA, 1994.
4. W. E. Deming, *Statistical adjustment of data*, Dover, N.Y., 1964.
5. I. Grabec and W. Sachse, *Synergetics of measurement, prediction, and control*, Springer Verlag, Berlin-Heidelberg, 1997.
6. V. Kreinovich, *How to explain the efficiency of linear regression for non-linear economic problems*, Technical Report, Leningrad Technological Center Informatika, Leningrad, 1989 (in Russian).
7. B. A. Litov, *Analysis of the personnel structure of the design institutions. Theory and Methodology*, Ph.D. Dissertation. Leningrad, 1981 (in Russian).
8. W. Menke, *Geophysical Data Analysis: Discrete Inverse Theory,* Academic Press, N.Y., 1984.
9. E. Merényi, B. Csathó, M. Bodrogi, and Á. Gulyás, "Utilization of Landsat images for mapping natural resources and for enviromnetal protection in Hungary", *Proc. Tenth Thematic Conference on Geologic Remote Sensing,* San Antonio, TX, May 9–12, **II**, pp. 491–502, 1994..
10. E. Merényi, B. Csathó, M. Bodrogi, and Á. Gulyás, "Integration of Landsat images, geophysical and radar data for mapping soil composition in temperate climate environment, Hungary", Submitted to *Remote Sens. Environ.*
11. E. Merényi, J. V. Taranik, T. B. Minor, and W. H. Farrand, "Quantitative comparison of neural networks and conventional classifiers for hyperspectral imagiery", In: R. O. Green (ed.), *Summaries of the Sixth Annual JPL Airborne Earth Science Workshop*, Pasadena, CA, March 4–8, **1**, 1996.
12. T. Moon and E. Merényi, "Classification of hyperspectral images using wavelet transforms and neural networks", *Proceedings of the Annual SPIE Conference*, **2569**, 1995.
13. S. H. Patterson, *Kaolin, refractory clay, ball clay, and halloysite in North America, Hawaii, and the Caribbean region*, Alexandria, VA: U.S. Department of the Interior, Geological Survey, 1984.
14. J. Pous, A. Marcuello, and P. Queralt, "Resistivity inversion with "a priori" information", *Geophysical Prospecting* **35**, pp. 590–603, 1987.
15. S. Starks and V. Kreinovich, "Soft Computing: Frontiers? A Case Study of Hyper-Spectral Satellite Imaging", *Working Notes of the AAAI Symposium on Frontiers in Soft Computing and Decision Systems*, Boston, MA, November 8–10, pp. 66–71, 1997.
16. C. H. Thurber, "Earthquake locations and three-dimensional structure in the Coyote Lake area, central California", *J. Geophys. Res.* **88**, pp. 8226–8236, 1983.