

PAPER • OPEN ACCESS

On several ill-posed and ill-conditioned mathematical problems of soil physics

To cite this article: M V Glagolev and A F Sabrekov 2019 *IOP Conf. Ser.: Earth Environ. Sci.* **368** 012011

View the [article online](#) for updates and enhancements.

On several ill-posed and ill-conditioned mathematical problems of soil physics

M V Glagolev^{1,2,3,4*} and A F Sabrekov^{3,4}

¹ Faculty of Soil Science, Lomonosov Moscow State University, GSP-1, Leninskie Gory, Moscow 119991, Russian Federation

² Institute of Forest Science of the Russian Academy of Sciences, Sovetskaya street, 21, Moscow Region 143030, Uspenskoe, Russian Federation

³ Yugorsky State University, Chehova street, 16, Khanty-Mansiysk 628012, Russian Federation

⁴ A N Severtsov Institute of Ecology and Evolution of the Russian Academy of Sciences, Leninskij prosp., 33, Moscow 119071, Russian Federation

*E-mail: m_glagolev@mail.ru

Abstract. Several well-known mathematical models of concentration fields in the soil (both at the single aggregate and the profile scales) are considered. It is shown that the respective boundary value problems for steady-state profiles belong to the class of ill-posed problems, since their solution does not exist. It occurs because a certain set of processes (for example, diffusion transport + first-order kinetic of the consumption) restricts possible boundary conditions, which, therefore, can no longer be arbitrary. Ill-posed inverse problems are also briefly described as well as one ill-conditioned inverse problem of parameters identification for mathematical model of the soil organic matter concentration profile. Exact solution for this model is the sum of two exponents. For a certain input data it was shown that this problem belongs to the class of ill-conditioned, since a small bias in the input data causes a significantly larger error in the solution (i.e. in calculated parameters).

1. Introduction

Methods of mathematical processing of the experimental results and mathematical modeling in ecology and soil science are currently used very widely and continue to develop [1-4]. In addition, the basic mathematical tasks in soil science are similar (or even identical) to the tasks of mathematical physics.

It is well known that the problems of mathematical physics can be divided into forward and inverse. Tasks which start with the causes and then calculate the effects represent the forward tasks. The inverse problems start from observed effects and then calculate causes not yet known [5]. For example, the soil properties and the intensity of organic matter input are the causes for the formation of soil organic matter (hereinafter – SOM) concentration profile. Here calculation of such a profile based on known soil properties and litter input represents a forward problem, whereas a calculation of soil properties or litter input, based on the profile of organic matter represents an inverse problem (hereinafter – IP). Besides the concept of forward and IP, classification of problems as well-posed and ill-posed has been introduced.



If sets of valid input data F and possible solutions V are given, the computational problem of determining $v \in V$ according to $f \in F$ from the equation $\hat{A}v = f$ (where \hat{A} – a continuous operator acting from the solution space to the input data space) is considered as well-posed if: (i) its solution v exists; (ii) for each f there is a unique v ; (iii) solution continuously depends on f i.e. solution is stable (small changes in input data f correspond to small variations of the resulting v). In this case, instability means that the solution is not a continuous function of data, i.e. a small perturbation of the data corresponds to a large perturbation of the solution [5].

A problem is ill-posed (hereinafter – IPoP), if one of the above conditions (i)-(iii) is not satisfied [5]. However, IPoP should not be considered as something wrong, as a manifestation of mathematical unprofessionalism. This is just a term for a special class of problems. Moreover, the ill-posed formulation of the problem makes it often possible to realize some important properties of the object under study. IPoP can arise among both forward and inverse problems. In addition, there is a large class of ill-conditioned problems (hereinafter – ICoP), which, although theoretically are well-posed, in fact, do not differ from IPoP (from the point of view of practical calculations).

Let us assume that the problem is correct (its solution exists, the solution is unique and stable to the uncertainty in the input data). Theoretically, a problem has stability if a sufficiently small error in input data induces a small error in the solution. However, in practice, the errors of the input data cannot be made arbitrarily small: their accuracy is limited. The sensitivity of solution to small errors in the input data is called *the conditionality of a computational problem*. A problem is called *ill-conditioned* if small input errors (δf) cause strong changes in the solution (δv). The value α in the inequality $\delta v \leq \alpha \delta f$ is called *condition number* (here δv and δf are *relative errors*). For ICoP $\alpha \gg 1$. However, an exact answer to the question about the value α , at which a problem should be considered as ICoP, essentially depends on the solution accuracy requirements and on the level of the accuracy provided for the source data [6]. In addition, for non-linear models, we should expect $\alpha \neq const$, therefore, the problem can be well-conditioned with some values of δf , but become ICoP with others (usually high) δf . Note that the problem that is *ill-posed* due to instability, is an extreme case of ICoP (with $\alpha = \infty$). However, in practice, there is no difference between such ill-posed problem and an ICoP, which is formally well-posed, but having so high $\alpha \neq \infty$ that (for a given δf) δv turns out to be completely unsatisfactory. The minimum possible values of δf are limited both by the accuracy of measurements (available at a given level of technological development) and by the minimum error of the model (available at a given level of the scientific progress).

In this paper, we consider some examples of both ICoP and IPoP that arose in soil science practice.

2. Failed condition for the solution existence

Apparently, in literature more attention was paid to inverse IPoP – see, for example, [5-9], but in soil science forward IPoP also occurs. In this section, some examples of both types are considered.

2.1. Steady-state profile of SOM

2.1.1. Diffusion-biodestruction steady-state profile. First of all we give the example of IPoP when condition of the solution existence is not satisfied: a model of the SOM concentration profile. In [10], a following stationary model, based only on the assumption of the diffusion mechanism of organic matter distribution combined with first-order kinetics biodestruction, is presented among others:

$$D \frac{d^2 f}{dx^2} = kf, \quad (1)$$

where x is a vertical coordinate (m); f is SOM concentration (kg m^{-3}); D is the phenomenological coefficient of SOM diffusion ($\text{m}^2 \text{year}^{-1}$); k is a reaction rate constant (year^{-1}). Boundary conditions (hereinafter – BoC) were chosen as follows: firstly, a flux of SOM in the form of humified litter (L ,

kg m⁻² year⁻¹) on the soil surface; and, secondly, constant concentration semi-infinite condition (when $x \rightarrow \infty$):

$$-D \frac{df}{dx} \Big|_{x=0} = L; \quad f(\infty) = C_o \quad (2)$$

(another condition $-f(h) = C_o$ with a large but finite h – is also mentioned in [10]; however, since the original solution is given only for equations (2), we will consider the boundary value problem with these BoC). The “solution” to this problem is suggested in the form

$$f = C_{(x)} - C_o,$$

where

$$C_{(x)} = C_o + \frac{L}{\sqrt{(kD)}} \exp(-x\sqrt{k/D}),$$

So

$$f = \frac{L}{\sqrt{(kD)}} \exp(-x\sqrt{k/D})$$

However, it is obvious that such function is not a solution, because

$$f(\infty) = \frac{L}{\sqrt{(kD)}} \exp(-\infty) = 0,$$

whereas according to equations (2) $f(\infty) = C_o$.

2.1.2. Non-existence of problem (1)-(2) solution. One can assume that the mistake in the solution mentioned above could be easily fixed. However, the origin of the mistake is much deeper: the solution of problem (1)-(2) does not exist (so this problem belongs to the class IPoP). It could be shown as follows. One could write well established (see, for example, [11]) general solution of equation (1) in this form:

$$f(x) = C_1 (\exp(x\lambda) + \exp(-x\lambda)) + C_2 (\exp(x\lambda) - \exp(-x\lambda)), \quad (3)$$

where

$$\lambda = \sqrt{(k/D)},$$

C_1, C_2 – arbitrary constants. This solution can be rewritten as:

$$f(x) = (C_1 + C_2) \exp(x\lambda) + (C_1 - C_2) \exp(-x\lambda),$$

so a bounded solution for $x \rightarrow \infty$ exists only when $C_1 = -C_2$. Consequently,

$$f(x) = -2C_2 \exp(-x\lambda),$$

but if $x \rightarrow \infty$ $f(x)$ cannot have arbitrary value C_o , and inevitably must be zero:

$$f(x) = -2C_2 \exp(-\infty \cdot \lambda) = 0$$

2.1.3. Steady-state profile of SOM with a root-associated carbon source. Example described above is not unique. In soil science calculation of spatial concentration profiles is often an IPoP. Let us give another example, which will be considered in general form. Smagin [10] developed the model (1) assuming that root systems can serve as the source of SOM:

$$D \frac{d^2 f}{dx^2} = kf - R \exp(-bx), \quad (4)$$

where the exponential term describes the intensity of the organic matter input from roots (distributed vertically in accordance with the exponential law). BoC (2) were chosen again. This problem also belongs to the IPoP class, since its solution does not exist. Indeed, according to [12], a general solution of inhomogeneous linear equation (4) can be presented as the sum of its particular solution and its general solution of the corresponding homogeneous equation. In this case equation (1) will be a homogeneous equation needed and its general solution is equation (3). It is easy to check that function

$$u(x) = \left(R / (k - Db^2) \right) \exp(-bx),$$

is a particular solution for equation (4). Therefore, its general solution is:

$$f(x) = C_1 (\exp(x\lambda) + \exp(-x\lambda)) + C_2 (\exp(x\lambda) - \exp(-x\lambda)) + \left(R / (k - Db^2) \right) \exp(-bx)$$

This again shows that when $x \rightarrow \infty$ the only possible boundary condition for function $f(x)$ (i.e. solution) is zero and equations (2) could not be satisfied with arbitrary value C_o . It also means that the solution of the problem (2), (4) for $C_o \neq 0$ does not exist.

2.2. Oxygen concentration profile in a soil aggregate.

At the end of this section, we briefly mention the model of biogenic O_2 consumption inside a moist spherical soil aggregate considered in [13]:

$$\frac{D}{r^2} \frac{d}{dr} \left(r^2 \frac{dC}{dr} \right) = kC, \quad (5)$$

where C is O_2 concentration; r is distance from the center of the aggregate ($0 \leq r \leq r_m$; r_m corresponds to the surface of the aggregate); complete formulation of the problem in [13] includes equation (5) and the following BoC:

$$C(0) = 0; \quad C(r_m) = C_o, \quad (6)$$

where $C_o > 0$ is treated as the concentration of O_2 at the surface of the aggregate (and depends on the depth of its location in the soil). The "solution" of the boundary value problem (5)-(6) is given by the author in the form

$$C = \frac{C_o r_m}{r} (\exp(r\lambda) - \exp(-r\lambda)) / (\exp(r_m\lambda) - \exp(-r_m\lambda))$$

But this function could not be a solution, since after applying L'Hopital's rule it is easy to show that

$$C(0) = 2C_o r_m \lambda / (\exp(r_m\lambda) - \exp(-r_m\lambda)) > 0,$$

while according to equations (6) $C(0)$ should be 0. In general, it can be proved that the solution of the problem (5)-(6) does not exist: for equation (5) and BoC $C(r_m) = C_o$ the value of $C(0)$ is always positive. Therefore, it is impossible to satisfy BoC $C(0) = 0$ and the problem of solving equation (5) with conditions (6) is IPoP. But, as mentioned above, this fact should not be considered as negative. IPoP arises when the author relies on a set of assumptions and experimental facts: (a) inside the

aggregate only diffusion transport occurs; (b) consumption of O_2 satisfies the first-order kinetics; (c) in the center of aggregate the oxygen sensor reported zero concentration of O_2 ... Occurrence of IPoP (considering *forward problem*) indicates the need to carry out comprehensive analysis of all assumptions and “facts” to find those that most likely contain uncertainties (they are not real). For example, from the above items (a) - (c) the most reliable is (a). But the fact that the kinetics could differ from the first order seems to be quite probable, and here it is necessary to investigate how the change in the kinetics type (within acceptable limits) will affect the problem. Further, the fact that the concentration of O_2 in the center of the aggregate will be small (outside the sensitivity of the sensor), but, nevertheless, not zero, in general seems very likely. Fortunately, a similar model was built by Gontchar-Zaykin et al. [14]. They showed that a slight modification gives a well-posed problem: instead of zero boundary condition, it is enough to claim the *limitations*: $0 < C(0) < C_0$. Thus, the analysis of ill-posed problem leads to the possibility of a reasonable choice of further ways to study the object by experiment: using more accurate sensors could show whether the O_2 concentration in the center of the aggregate will differ from zero.

3. Failed condition of the solution uniqueness

Completely different reasons can lead to multiple (non-unique) solutions of a certain problem. In a short manuscript we are not able to cover this topic comprehensively and just give a few typical examples.

3.1. Cluster analysis

3.1.1. The multiple options of distance selection in classification of objects. In soil science in general and in soil physics specifically cluster analysis is widely used for the classification of any objects [15, 16, 17, 18]. The cluster analysis is designed to classify observations (objects) into more or less homogeneous groups. Its governing idea is to calculate a certain measure of similarity or dissimilarity (for example, some distance) between each pair of objects. Low value of this distance indicates that objects are similar or close to each other, while a large value indicates a lack of similarity. But similarity estimates could be different [19]. Moreover, other approaches are based on a non-distance concept. Specifically, an approach associated with fundamental concepts of the graph theory can be used [20]. For example, let us consider only MATLAB environment which is not designed namely for cluster analysis and operated only with the most common distances. Even in this case there are 10 different distance options (Chebychev distance, City Block metric, Hamming distance, Mahalanobis distance, Minkowski metric, Euclidean and Standardized Euclidean distances *etc.*). Moreover, for example, in the Minkowski metric parameter of the exponent (fixed number) must be set by user since there is an indefinite number of possible exponents. Therefore, there is an indefinite number of options for cluster analysis even if we use Minkowski metric only.

3.1.2. The multiple options of object grouping algorithms. The next problem is to obtain a hierarchical grouping of objects, where the objects with the highest similarity coefficients are placed together. Then the groups of objects are bound to form new groups (they are most closely connected too), and this process continues until a complete classification of the objects is obtained. Unfortunately, at this stage there are many methods for analyzing groups as well [19]. Moreover, along with the approach of *binding* single objects into clusters, it is possible to *divide* the entire set of objects into clusters. The difference in procedures may be caused by other reasons, for example, by choosing initial points of clustering. This can be a random selection of one or several points as well as not random “typical” point [20]. Considering only MATLAB environment again, there are 7 available algorithms (“nearest neighbor”, “furthest neighbor”, “average linkage” *etc.*). It may seem that this situation is not unusual because it is typical for almost any problems. For example, in the classic numerical analysis we have many algorithms for integrating of differential equations (see, for example, [6, 12, 21, 22]). However, this difference is fundamental. Various algorithms of the differential equations numerical integration

give solutions with error less than a small value ε comparing with the exact solution of a differential equation. Therefore, all the algorithms give the same solution with an error that can be considered arbitrarily small (within reasonable limits) while various algorithms of grouping in cluster analysis can give completely different solutions.

3.1.3. Multi-optional choice of the classes' number. Finally, one of the most important issues in solving a cluster problem is the choice of the required number of clusters [23]. In the literature hundreds (!) cluster analysis methods are described. Their diversity origins from the fact that the identification of clusters in many ways is "art". Many of researchers who use the techniques of this analysis create a new method themselves [20]. For example, in MATLAB there are two fundamentally different possibilities for this choice: "Finding the Natural Divisions in the Data Set" and "Specifying Arbitrary Clusters". Each of them generates a set of variants of the clusters number (in the latter case this number is just defined by user). Thus, at each stage of cluster analysis there is a variety of criteria and/or algorithms, that results in the non-uniqueness of the solution of the whole classification task.

3.2. Non-uniqueness of the model parameter identification problem.

When one identifies parameters of mathematical models, some distance between model prediction and experimental data is calculated [5, 19, 22, 24]. Since this distance is a function of the model parameters, it is possible to solve the minimization problem: to find parameters for which the distance reaches minimal value. With these values of parameters the model fits the experimental data in the best way. Therefore, these values are considered as a solution of the parameters identifying problem. Obviously, non-uniqueness arises here from one of the reasons that we have already mentioned discussing non-uniqueness of cluster analysis. Indeed, in practice various distances are used: sum of squared deviations between model predictions and experimental data [5] (it corresponds to the sum of the absolute error squares); sum of weighted squared deviations [6, 22, 25]; sum of relative error squares [26]; some other types of distances are also used [19, 27, 28]. However, in practice non-uniqueness of solutions does not always lead to catastrophic consequences, which are usually associated with IPoP. Although we get formally different solutions (different sets of numerical values of parameters), but the values of the same parameter, (i) will not greatly differ from each other and (ii) assuming the adequacy of the model, will be close to the true values for models supplied with comprehensive set of experimental information. Therefore, in our opinion, the ill-posedness associated with the solution instability is much more dangerous, as well as the poor conditionality of the formally well-posed problem. Both occur when the model of the chosen level of complexity is not provided with adequate experimental data (for example if the complexity of the model exceeds the experimental data resolution).

4. Ill-conditionality of one inverse problem of the parameter identification

In soil science as well as in chemistry, biology, physics [20, 22, 29, 30] a problem of representing experimental data as a sum of exponential functions (where coefficients have an important physical meaning) [10, 31, 32] is common.

Many models of SOM decomposition use only two exponential functions (one for labile components and one for recalcitrant):

$$f(x) = C_1 \exp(-k_1 x) + C_2 \exp(-k_2 x), \quad (7)$$

where f is substrate (in % of the initial amount of substrate) at time x ; C_1 and C_2 are initial amounts of substrates (% , therefore $C_1 + C_2 = 100\%$) which are decomposed, respectively, with rate constants k_1 and k_2 [33]. If one needs to determine C_1 , C_2 , k_1 and k_2 from the residual substrate temporal dynamics, the solution of the problem can be presented (using the designation introduced in section "Introduction") as a vector

$$\vec{v} = (C_1, C_2, k_1, k_2)$$

The formula (7) is also a solution of the boundary value problem (2), (4) at $C_0 = 0$. In this case, the coefficients in equation (7) depend on physical parameters D, k, L, R, b :

$$C_1 = L/(k_1 D) - bC_2/k_1, \quad k_1 = \sqrt{k/D}, \quad C_2 = R/(k - Db^2), \quad k_2 = b \quad (8)$$

Shape of the SOM depth profile represents SOM accumulation, decomposition and mass transfer processes rates. Therefore, it seems that it is theoretically possible to assess the intensity of these processes from the data on the SOM profile distribution [10]. But can it be used in practice?

So let us consider the problem of the SOM profile distribution model parameters identifying, i.e. we need to find the parameters C_1, k_1, C_2 and k_2 based on known SOM concentrations at various depths – $f_i(x_i)$ – and then calculate the values D, k, b and R :

$$D = L/(k_1 C_1 + k_2 C_2), \quad k = Dk_1^2, \quad b = k_2, \quad R = (k - Dk_2^2)C_2 \quad (9)$$

As mentioned above, the problem of parameters identification is well studied and can be solved, for example, using the least squares method. However, we face the most difficult problem: how to find out that the *correct* results are obtained [34]. Apparently, the simplest solution is to test the properties of IP and its solution algorithm by following method: (i) set some values D, k, b, R and L ; (ii) then calculate values $f_i(x_i)$ by equation (7) for the depths x_i where data on measured SOM concentration are available; (iii) then slightly “bias” the calculated values by adding random noise Δf_i simulating the uncertainty of experimental data; and finally (iv) solve IP with “biased” values ($f_i(x_i) + \Delta f_i$). The solution is the set of parameters that we denote D^*, k^*, R^* and b^* (since their values are slightly different from the original). Because we know the output values D, k, b and R exactly, the described algorithm gives an estimate of the error which arises from a certain IP solution method at given level of noise (if we perform the described algorithm N times). Therefore, it is possible to estimate the influence of the uncertainty in the “experimental” data on the error of the result, i.e. assess the conditionality of the problem or, in other words, its sensitivity to the source data uncertainty. According to [35]

$$N = (N_p + 1)N_e,$$

where N_p – number of model parameters, N_e – number of experimental points.

The idea underlying this sensitivity test is very simple and has a general character: to solve a problem with several datasets and to check how sensitive the solution is to a variation in input data [34]. Let us give a particular example. For calculations, we use values given in [10] for typical chernozem: $D = 0.000251 \text{ m}^2 \text{ year}^{-1}$, $R = 0.227523 \text{ kg year}^{-1} \text{ m}^{-3}$, $k = 0.00278 \text{ year}^{-1}$, $b = 2.558 \text{ m}^{-1}$ and $L = 0.008 \text{ kg year}^{-1} \text{ m}^{-2}$. The values f_i (given in table 1) were calculated according to the formula (7) with coefficients calculated by equations (8).

Table 1. Dataset for equation (7) parameter’s identification.

SOM concentration	Depth (x_i m)																		
	0.04	0.07	0.11	0.18	0.22	0.33	0.40	0.51	0.62	0.84	1.02	1.21	1.43	1.61	1.79	2	2.5	3	
f_i	54.51	52.87	51.00	46.84	44.64	38	33.67	27.7	22.6	14.5	9.79	6.52	3.95	2.58	1.67	1.0	0.30	0.09	
1 st ($f_i + \Delta f_i$)	55.35	54.81	52.54	47.96	41.58	39.7	31.68	25.1	23.2	12.9	9.29	6.34	3.60	2.35	1.68	1.0	0.30	0.10	
2 nd ($f_i + \Delta f_i$)	56.97	52.55	52.50	50.58	44.41	40.7	31.08	27.3	23.0	14.3	9.99	5.80	3.90	2.71	1.48	1.0	0.31	0.09	
...
90 th ($f_i + \Delta f_i$)	56.88	51.94	53.55	45.45	40.83	37.8	33.75	27.6	21.1	14.6	9.37	6.96	3.96	2.58	1.71	1.0	0.31	0.09	

These values (but with Gaussian noise Δf_i) were used as “pseudo-experimental data” (further – PED). Several sets of PED for coefficient of variation $CV = 5\%$ are presented in table 1. For each

level of CV, $N = (4 + 1) \cdot 18 = 90$ sets of PED were generated; then 90 sets of C_1 , k_1 , C_2 and k_2 were calculated with help of MATLAB-function `lsqcurvefit`. Finally, 90 values of D , 90 values of k etc. were obtained using equations (9) and CVs for these parameters were evaluated (see MATLAB-code in Appendix).

The results are presented on a figure 1. Obviously, the identification problem is well-conditioned for parameter b because its value of CV approximately corresponds to CV of source data f . But for all other parameters we have typical ICoP. For example, if CV input data is 10% then CV for D will be $27.763 \cdot 10^{1.36} \approx 636\%$, for R and k – about 1000% (k and R have almost identical curves of “CV – noise level” dependence, so on a figure 1 they are overlapping). In other words, if the experimental measurements (f) were obtained with an uncertainty of $\approx 10\%$, then the coefficients R and k obtained by solution of IP for model (7)-(9) could differ from the real ones by a factor of 10.

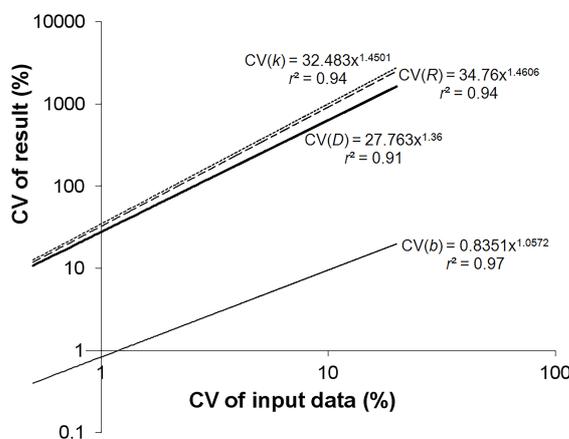


Figure 1. Conditionality for the parameter identification IP for the equation (7).

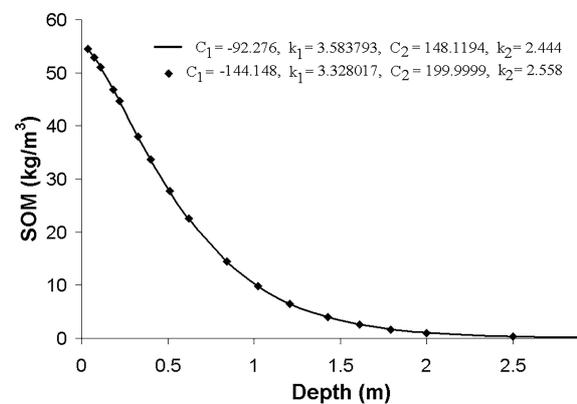


Figure 2. Ambiguity of data approximation by equation (7).

The reason of ill-conditional character of parameter identification is briefly illustrated on figure 2. As we see, the formula (7) with completely different values of the parameters (cf. $C_1 = -92.276$ and $C_1 = -144.148$) gives almost identical curves. Application of the method described here (conditional analysis through perturbation sensitivity) for other tasks in biology and soil science see, for example, in [36-38].

5. Basic principles of ill-posed problems solution

The question of the IP solution existence consists of two aspects. One is the physical existence of the set of parameters generating the observed data; the second is the existence of a mathematical solution of operator equation $\hat{A}v = f$ (see “Introduction”). A formal mathematical solution may not exist. To understand it better, we note that the measured data f_e always have some uncertainty (Δf): $f_e = f + \Delta f$. The question is whether it is possible to find a set of parameters v_e , that strictly generates a dataset f_e : $f_e = \hat{A}v_e$. The answer is that sometimes we cannot find such parameters, and it is easy to understand why. Indeed, f_e – are dataset f spoiled by noise. But this noise is not related to model parameters because it could be produced by phenomena that were described by concentration field equations given above. This is a reason why noise could not be predicted using the same operator equation, as for idealized dataset. It means that we should not expect that we always can find physically realistic model that exactly corresponds to observed dataset [5]. If a model fitted data perfectly, the following equality would have place: $\|\hat{A}v_e - f_e\| = 0$.

But does it make sense to find such a model? It does not give meaningful results for noisy data (data are always noisy). So we should think about certain approximate approach to inversion, based on the search for a model that would fit observations within the given accuracy. Therefore, we are

looking for the pseudosolution of inverse problem, i.e. about such a solution from the certain class of models that fits the observed data in the best way. Pseudosolution of IP exists if there is such v_e that

$$\left\| \hat{A} v_e - f_e \right\| \leq \Delta, \quad (10)$$

where $\|\cdot\|$ means a certain measurement of discrepancy between observed and predicted noisy data [5]. Different formulations of this measurement for variable types of v_e and f_e – vectors, functions, etc. – see, for example, in [5, 34, 39, 40]). Based on a using measure one can choose $\Delta = \Delta(\Delta f)$ but strictly speaking $\Delta \sim \Delta f$ and often simple rule of thumb $\Delta = \Delta f$ is used.

This simple idea (inequality (10)) – is a cornerstone of a regularization theory [5]. However, this inequality provides an existence of the solution but not its uniqueness. Indeed, a variety of different v_e can satisfy this inequality: for some of them left-hand side would be much less than Δ , for others – just slightly less, and for a certain value of a v_e right- and left-hand sides would be strictly equal. The correct solution of IPoP and ICoP can be obtained using a variety of regularization methods (deterministic, statistical or descriptive regularization) – see, for example, [5, 7, 39, 41] and references therein. The most commonly used method of regularization of ill-posed problems is Tikhonov regularization ("ridge regression", "weight decay", "Tikhonov–Miller method", "Tikhonov–Phillips regularization").

Tikhonov regularization is based on minimization of smoothing functional

$$J(\hat{A}, v_e, f_e, \xi) = \left\| \hat{A} v_e - f_e \right\| + \xi \cdot \Omega(v_e),$$

where $\xi > 0$ – regularization parameter, chosen based on equation (10); $\|\hat{A} v_e - f_e\|$ – functional of discrepancy between predicted and observed data; Ω – regularization term [40]. Both functionals should be constructed in a way to be nonnegative. Let us describe briefly one of possible options of Tikhonov regularization method (where Morozov principle for choosing the regularization parameter [5, 42, 43] was used, but there are also different principles of this choice – see, for example [42–45]). For an arbitrary ξ_0 corresponding value of $v_e(\xi_0)$ is calculated, providing a minimal value of J . If for calculated $v_e(\xi_0)$ it was obtained that $\|\hat{A} v_e(\xi_0) - f_e\| < \Delta$, next iteration (of functional minimization) is carried out using the value $\xi_1 > \xi_0$. If it was obtained that $\|\hat{A} v_e(\xi_0) - f_e\| > \Delta$, for the next iteration value $\xi_1 < \xi_0$ should be used. For the obtained new value $v_e(\xi_1)$ next check of discrepancy between functional and uncertainty level should be performed. If $\|\hat{A} v_e(\xi_1) - f_e\| \neq \Delta$, using the procedure described above new value of ξ_2 should be chosen. Then a new level of $v_e(\xi_2)$, minimizing functional, should be found, and again discrepancy should be compared with uncertainty level until it would be found a value of ξ_* providing $\|\hat{A} v_e(\xi_*) - f_e\| = \Delta$. Corresponding value of $v_e(\xi_*)$, providing a minimum of functional J , would be considered as a regularized solution.

It is obvious, that the scheme described above can be directly used in a case when solution of operator equation (v) is a function. But here we want to discuss a simpler example, when v is a vector (containing a certain number of components). A problem of parameter identification for equation (7) described in section 4 belongs to this class. In this case J is a function and not a functional. So instead of functional minimization relatively simple problem of minimization of multi-argument function should be considered. The term $\|\hat{A} v_e - f_e\|$ is a sum of squared deviations between model predictions and observed data. If there are any prior estimates of parameter values in equation (7), that could be considered as components of vector v_a , term $\Omega = \|v_e - v_a\|$ can be calculated as a sum of squared deviations between components of a resulting vector and components of a priory [46]. If components of the vector vary considerably, normalization is usually carried out in Ω . For example, these squared deviations are divided into squares of corresponding v_a vector components [41]. A priori estimates are usually based on data for systems similar in certain characteristics to systems of interest. For example

if we need to find parameters for chernozem and parameters for luvisol are known, one can use them for calculations.

6. Concluding remarks

Apparently Chudnovskii [47] was one of the first scientists who raised the question about ill-posed inverse problem in soil science. He considered IPoP for soil thermal conductivity identification and methods for solving of this problem. Some inverse IPoP of ecology and soil science were solved in [48-51] *etc.* However, the ideology of IPoP in soil science unfortunately has not been developed comprehensively, despite the fact that IPoP is common in this area. Some researchers treated IPoP as a usual (i.e. in this case – well-posed) problem of numerical analysis. Of course, the results obtained with this approach, are ambiguous, since their uncertainty is high.

In this brief article we were able to consider only a few issues related to ill-posed and ill-conditioned problems, as well as only a very limited number of examples of such problems in the soil science. Thus, we absolutely did not discuss the classical problems of experimental data processing. But the number of IPoP is very high among them. Specifically, the problem of function differentiation (if the function has an uncertainty), probability density estimation for empirical distributions, solving some types of integral equations, *etc.* lead to IPoP. On the other hand, in recent years, the processes of automatic collection and processing of experimental data in soil science and ecology have become increasingly widespread [52-55]. So solving different ill-posed and ill-conditioned problems could become much more important in the near future.

Finally, it should be noted that due to the limited volume of the paper we briefly considered the most important topic: how to solve IPoP and ICoP. Fortunately, these methods are well developed in the mathematical literature.

7. Appendix

```
function lsqcurvefit_test
% Experimental data
X = [0.04 0.07 .11 0.18 0.22 .33 0.4 0.51 0.62 0.84 1.02 1.21 1.43 1.61 1.79 2 2.5 3];
Y = [54.51 52.87 51 46.84 44.64 38 33.67 27.7 22.6 14.5 9.79 6.52 3.95 2.58 1.67 1 0.3 .09];
L = 0.008; % [kg/year/m2] Litter flux
N = 90; CV = 0.05; % N - the number of repetitions in statistical modeling; CV = std / mean
xdata = X; xlnY = xdata(end-4:end); options = optimset('MaxFunEvals', 8000, 'MaxIter', 2000);
for i = 1:N
    for j = 1:length(X) % We generate pseudo-experimental data with an error
        ydata(j) = Y(j) * (1 + CV * randn); if ydata(j) < 0 ydata(j) = eps; end % The concentration can not be < 0
    end % for j ...
lnY = log(ydata(end-4:end)); [P, S] = polyfit(xlnY, lnY, 1); % Starting guess
x0 = [-exp(P(2)), -P(1), exp(P(2)), -P(1)];
% The solution of the least squares problem:
[x, resnorm] = lsqcurvefit(@myfun, x0, xdata, ydata, [], [], options);
% The calculation of the coefficients by the formulas (9):
B(i) = x(4); d(i) = L / (x(1) * x(2) + x(3) * x(4)); K(i) = d(i) * x(2)^2; r(i) = x(3) * (K(i) - d(i) * x(4)^2);
end % for i ...
% Average values and CV according to the results of statistical modeling:
b = mean(B), CV_b = std(B) / b, D = mean(d), CV_D = std(d) / D,
k = mean(K), CV_k = std(K) / k, R = mean(r), CV_R = std(r) / R,
% The equation for approximation of experimental data:
function F = myfun(x, xdata)
F = x(1) * exp(-x(2) * xdata) + x(3) * exp(-x(4) * xdata);
```

Acknowledgment

Authors thank D. Il'yasov for help in figures design. This study is supported by The Russian Science Foundation, grant number 19-77-10074 (the study of inverse problems) and by RFBR according to the research project number 18-34-00695 (the study of ill-posed direct problems).

References

- [1] Balland V, Pollacco J A P and Arp P A 2008 Modeling soil hydraulic properties for a wide range of soil conditions *Ecol. Model.* **219** 900-16
- [2] Kurbatova J, Li C, Tatarinov F, Varlagin A, Shalukhina N and Olchev J 2009 Modeling of the carbon dioxide fluxes in European Russia peat bogs *Environ. Res. Lett.* **4** p 045022
- [3] Glagolev M V, Sabrekov A F, Kleptsova I E, Filippov I V, Lapshina E D, Machida T and Maksyutov Sh Sh 2012 Methane emission from bogs in the subtaiga of Western Siberia: The development of Standard Model *Eurasian Soil Sci.* **45** 947-57
- [4] Shein E V 2015 Physically based mathematical models in soil science: History, current state, problems, and outlook (analytical review) *Eurasian Soil Sci.* **48** 712-18
- [5] Zhdanov M S 2002 *Geophysical Inverse Theory and Regularization Problems* (Amsterdam: Elsevier) p 633
- [6] Engeln-Mullges G and Uhlig F 1996 *Numerical Algorithms with Fortran* (Berlin: Springer) p 602
- [7] Beck J V, Blackwell B and St Clair C 1985 *Inverse Heat Conduction. Ill-Posed Problems* (New York: A Willey-Interscience Publication) p 308
- [8] Hein R, Crutzen P J and Heimann M 1997 An inverse modeling approach to investigate the global atmospheric methane cycle *Global Biogeochem. Cy.* **11** 43-76
- [9] Kim H-S, Maksyutov S, Glagolev M V, Machida T, Patra P K, Sudo K and Inoue G 2011 Evaluation of methane emissions from West Siberian wetlands based on inverse modeling *Environ. Res. Lett.* **6** p 035201
- [10] Smagin A V 2001 Distributed kinetic models of soil organic matter profile formation *Mosc. Univ. Soil Sci. Bull.* **56** 3-7
- [11] von Kamke E 1942 *Differentialgleichungen, Lösungsmethoden und Lösungen* (Leipzig: Becker & Erler) p 642
- [12] Korn G A and Korn T M 1961 *Mathematical Handbook for Scientists and Engineers* (New York: McGraw-Hill Book Company) p 943
- [13] Smagin A V 2005 *Gas Phase of Soils* (Moscow: Moscow State University) p 301 (in Russian)
- [14] Gontchar-Zaykin P P, Dynkin L D, Dynkin S D and Zhuravlev O S 1981 Gas exchange model for "microbes-soil-atmosphere" *Modelling of Biogeosystem Processes* ed V V Galitsky (Moscow: Nauka) pp 142-8 (in Russian)
- [15] Kosaki T and Juo A S R 1989 Multivariate approach to grouping soils in small fields. II. Soil grouping technique by cluster analysis *Soil Sci. Plant Nutr.* **35** 517-25
- [16] Gradusov B P 2001 Cluster analysis of mineralogical and particle-size distribution data on soils developed from homogeneous and heterogeneous substrates *Eurasian Soil Sci.* **34** 1200-10
- [17] Dai Z, Li R, Muhammad N, Brookes P C, Wang H, Liu X and Xu J 2014 Principle component and hierarchical cluster analysis of soil properties following biochar incorporation *Soil Sci. Soc. Am. J.* **78** 205-13
- [18] Kholodov V A, Yaroslavtseva N V, Frid A S and Lazarev V I 2016 Interpretation of data on the aggregate composition of typical chernozems under different land use by cluster and principal component analyses *Eurasian Soil Sci.* **49** 1026-32
- [19] Davis J C 1986 *Statistics and Data Analysis in Geology* (New York etc.: John Willey & Sons)
- [20] Knoring L D and Dech V N 1989 *Mathematics for Geologists: Practical Recommendations* (Leningrad: Nedra) p 208 (in Russian)
- [21] Hamming R W 1987 *Numerical Methods for Scientists and Engineers* (New York: Dover Publications) p 721
- [22] Johnson K J 1980 *Numerical Methods in Chemistry* (New York, Basel: Marcel Dekker, INC)

p 503

- [23] Duran B and Odell 1974 *Cluster Analysis* (Berlin: Springer-Verlag) p 137
- [24] Stepanenko V M, Machul'skaya E E, Glagolev M V and Lykossov V N 2011 Numerical Modeling of Methane Emissions from Lakes in the Permafrost Zone *Izv. Atmosf. Ocean. Phys.* **47** 252-64
- [25] Sabrekov A F, Glagolev M V, Alekseychik P K, Smolentsev B A, Terentieva I E, Krivenok L A and Maksyutov S S 2016 A process-based model of methane consumption by upland soils *Environ. Res. Lett.* **11** p 075001
- [26] Himmelblau D M 1970 *Process Analysis by Statistical Methods* (New York: John Wiley & Sons) p 463
- [27] Claeskens G and Hjort N L 2008 *Model Selection and Model Averaging* (Cambridge: Cambridge University Press) p 312
- [28] Sabrekov A F, Runkle B R K, Glagolev M V, Terentieva I E., Stepanenko V M, Kotsyurbenko O R, Maksyutov S S and Pokrovsky O S 2017 Variability in methane emissions from West Siberia's shallow boreal lakes on a regional scale and its environmental controls *Biogeosci.* **14** 3715-42
- [29] Vasil'eva G K, Surovtseva E G, Semenyuk N N, Glagolev M V and Panikov N S 1995 A method for enumerating chloroaniline-degrading microorganisms in soil proceeding from the substrate half-degradation period *Microbiol.* **64** 480-8
- [30] Marple S L 1987 *Digital Spectral Analysis with Applications* (Englewood Cliffs: Prentice-Hall)
- [31] Sanoyan S V and Egiazyryan G M 1989 *Abstracts of the VIII Soviet Union Soil Scientists Congress (Novosibirsk)* vol 4 (Novosibirsk: Institute of Soil Science and Agrochemistry) p 270 (in Russian)
- [32] Kuzyakov Y 2011 How to link soil C pools with CO₂ fluxes? *Biogeosciences* **8** 1523-37
- [33] Tate R L 1987 *Soil Organic Matter: Biological and Ecological Effects* (New York: John Wiley & Sons) p 291
- [34] Rice J R 1981 *Matrix Computations and Mathematical Software* (New York: McGraw-Hill Book Company) p 248
- [35] Varfolomeev S D and Gurevich K G 1999 *Biokinetics: Practical Course* (Moscow: FAIR-PRESS) p 720 (in Russian)
- [36] Eremeyev N L, Karyakin A A and Kazanskaya N F 1989 Kinetics of dissolution of solid protein substrates by proteinases: Selection of the reaction mechanism *Biochemistry (USSR)* **54** 503-10
- [37] Panikov N S, Blagodatsky S A, Blagodatskaya J V and Glagolev M V 1992 Determination of microbial mineralization activity in soil by modified Wright and Hobbie method *Biol. Fert. Soils* **14** 280-7
- [38] Glagolev M V 2012 Sensitivity analysis of the model *Environ. Dynamics Glob. Climate Change* **3** 31-53 (in Russian)
- [39] Tikhonov A N and Arsenin V Y 1977 *Solutions of Ill-Posed Problems* (Washington: V H Winston & Sons) p 258
- [40] Matsevity Yu M and Lushpenko S F 1990 *Identification of Thermophysical Properties in Solids* (Kiev: Naukova dumka) p 216 (in Russian)
- [41] Mitsel A A, Pshenichnikov A M and Scherbakov A P 2004 Using the local regularization for fitting the line contours in molecular spectra *Comput. Technol.* **9** 63-9 (in Russian)
- [42] Engl H, Hanke M and Neubauer A 1996 *Regularization of Inverse Problems* (Dordrecht: Kluwer) p 328
- [43] Sizikov V 2017 *Direct and Inverse Problems of Image, Spectroscopy and Tomography Reconstruction in Matlab* (Saint Petersburg: Lan) p 412 (in Russian)
- [44] Hansen P C 2010 *Discrete Inverse Problems: Insight and Algorithms* (Philadelphia: SIAM) p 213
- [45] Sizikov V S and Stepanov A V 2015 Method of training examples in solving inverse ill-posed

- problems of spectroscopy *Sci. Techn. J. Inform. Technol., Mech. Optics* **15** 1147-54
- [46] Dmitriev V I 1990 Methods of solution of inverse problems in geophysics *Computational mathematics and techniques in exploration geophysics* ed V I Dmitriev (Moscow: Nedra) pp 171-209 (in Russian)
- [47] Chudnovskii A F 1976 *Thermal Physics of Soils* (Moscow: Nauka) p 352 (in Russian)
- [48] Galanin M P and Tikhonov N A 1985 *Sov. Soil Sci.* **17** 131-5
- [49] El Badia A, Ha-Duong T and Hamdi A 2005 Identification of a point source in a linear advection–dispersion–reaction equation: application to a pollution source problem *Inverse Problems* **21** doi:10.1088/0266-5611/21/0/000
- [50] Shein E V, Kokoreva A A, Gorbatov V S, Umarova A B, Kolupaeva V N and Perevertin K A 2009 Sensitivity assessment, adjustment, and comparison of mathematical models describing the migration of pesticides in soil using lysimetric data *Eurasian Soil Sci.* **42** 769-77
- [51] Terentieva I E, Sabrekov A F, Glagolev M V and Kotsyurbenko O R 2017 Methane emission from municipal solid waste landfills *Russ. Meteorol. Hydrol.* **42** 327-34
- [52] Arkhangel'skaya T A and Umarova A B 2008 Thermal Diffusivity and Temperature Regime of Soils in Large Lysimeters of the Experimental Soil Station of Moscow State University *Eurasian Soil Sci.* **41** 276-85
- [53] Egorov Yu V, Kirichenko A V, Bobkov A V and Galitskii V I 2010 Averaged assessment of the soil moisture by medium-frequency dielectricity *Eurasian Soil Sci.* **43** 187-93
- [54] Rezanezhad F, Couture R-M, Kovac R, O'Connell D and Van Cappellen P 2014 Water table fluctuations and soil biogeochemistry: An experimental approach using an automated soil column system *J. Hydrol.* **509** 245-56
- [55] Alekseychik P, Mammarella I, Karpov D, Dengel S, Terentieva I, Sabrekov A, Glagolev M and Lapshina E 2017 Net ecosystem exchange and energy fluxes measured with the eddy covariance technique in a western Siberian bog *Atmos. Chem. Phys.* **17** 9333-45