



## Inverse Problems and Data Fusion for crop production applications targeting optimal growth - Fertilization

**Kaur, Bipjeet; Owusu, Robert K. A.**

*Published in:*

Proceedings of the 1st DEXA Workshop on Integrated Processing, Control and Knowledge Systems for Sustainable Production in Farms and Forests

*Publication date:*  
2015

*Document Version*  
Peer reviewed version

[Link back to DTU Orbit](#)

*Citation (APA):*

Kaur, B., & Owusu, R. K. A. (2015). Inverse Problems and Data Fusion for crop production applications targeting optimal growth - Fertilization. In Proceedings of the 1st DEXA Workshop on Integrated Processing, Control and Knowledge Systems for Sustainable Production in Farms and Forests IEEE.

## DTU Library

Technical Information Center of Denmark

---

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

# Inverse Problems and Data Fusion for crop production applications targeting optimal growth - Fertilization

Bipjeet Kaur  
Center for Wireless Systems and Applications  
Section for Information Technology  
Technical University of Denmark  
DK-2750 Ballerup  
bika@dtu.dk

Robert K. A Owusu  
Novitek Solutions ApS  
DK-1660 Copenhagen V  
Denmark  
robert.owusu@novitek.dk

**Abstract**—This work in progress is a contribution to crop growth systems for planning and monitoring of farm activities and practices by farmers. The work outlines the initial findings related to modelling, simulation and visualization techniques for crop growth, specifically targeting the barley crop, such that the crop yield is optimized with respect to several parameters (e.g. high end user value and minimum environmental impact), thus obtaining a sustainable production. The growth process optimization is based on information, including sensor based measurements with sensor quality monitoring, from previous and the present growth season. Initially, references targeting the importance of site specific management for obtaining the objective of yield optimization under the constraint of minimizing the environmental load is pointed to. This is followed by key references on modelling, simulation and visualization of the crop growth process based on information on soil quality, field seeding, spraying/fertilization and environmental information in general. Finally, references to software tools, which could form the basis for an open source platform for a planning and monitoring system for optimal crop growth in multiple application areas are given. The contribution concludes with proposals of research questions to be pursued in the near future.

## I. INTRODUCTION

The agricultural sector is facing great challenges. The biggest challenge is to feed the increasing population. FAO predicts 9.6 billion people on earth planet in 2050, which demands 70 percent increase in food production [8]. Another challenge is soil sustainability. Crop production requires various nutrients from the soil where the most important are Nitrogen (N), Phosphorous (P) and Potassium (K). Extensive research has been done to develop the requirement specification with respect to nutrients of each crop. In this paper we will consider the growth of barley. It is used as animal fodder, as a source of fermentable material for beer and certain distilled beverages, and as a component of various health foods. It is widely grown among others in Finland for brewing and beverages.

This contribution is divided into the following parts. Initially the importance of the site specific growth process is reviewed, then follows a section covering the modelling, simulation and visualization of growth processes. It is

subdivided into the very important area of soil texture and property modelling and simulation, then follows crop growth modelling and simulation. The next section is targeting software for open source platform for crop growth simulation and finally is concluded with a selection of research questions to be pursued, with the final goal of obtaining a system for practical use, among others, in relation to farming.

## II. THE SITE SPECIFIC CROP GROWTH PROCESS

The constantly increasing constraints on the methods and resources used within agricultural crop production for food and fodder, originates in the societies' requirements for continuous higher production yields, such that the environmental impact is at least not increasing or preferably decreasing, thus leading to sustainable production.

Such a sustainable production with increasing yield throughout the years is very difficult. It requires detailed, site specific information of the variability of different soil attributes within a field [1], [13] and [9] for the decision making in the production process with respect to when to start seeding/fertilization, when to carry out spraying (and possibly simultaneously carry out a second fertilization) and finally when to harvest the barley. The approx. 85 days production cycle of barley is exemplified [2]. One of the most critical aspects of characterizing soil is actually to obtain, at a reasonable precision and cost, representative soil samples (i.e. collected with adequate spatial density at the proper depth and during the appropriate time) such that it can be used in the decision process.

## III. THE MODELLING, SIMULATION AND VISUALIZATION OF CROP GROWTH PROCESSES

### A. Soil Texture and Property Modelling and Simulation

#### *Inverse Modelling of Fertilization Process via a First Order Fredholm Integral*

The objective of inverse modelling technique of the fertilization process for plant growth is to estimate optimum values of model parameters from external controlled input and other stochastic measured data including environmental variables. Farm models are often complex non-linear dynamic models which can include numerous input factors e.g., soil and

vegetation parameters, agricultural practices, meteorological input data [5]. The variation in soil texture, surface moisture or vertical soil moisture gradient in larger scale atmospheric models may lead to significant variations in simulated surface fluxes of water and heat. Techniques to scale and aggregate the soil characteristics are presented in [10] and focuses on direct and indirect use in large scale meteorological models.

### Mathematical Model

Suppose that it has been determined that some mass of fertilizer has been deposited at a particular location and at a depth  $h$  beneath the surface of the earth in a farm. Our objective is to compute the mass distribution of the fertilizer deposit denoted  $f(t)$ , along  $t$ -axis.

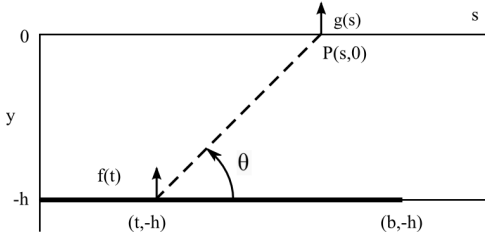


Fig. 1. Geometric illustration of the fertilization process problem in one dimension. The measured signal  $g(s)$  is the injected mass of fertilizer modelled as a one dimensional mass distribution  $f(t)$  at a depth  $h$ .

The description of the fertilization model from an inverse problem perspective can be formulated via a first order Fredholm Integral equation of the first kind [4];

$$g(s) = \int_{\Omega} Q(s, t) f(t) dt \quad (1)$$

where  $\Omega$  defines the limit of integration in  $n$ -dimensional space,  $Q$  is the kernel,  $f$  is the unknown function we seek to reconstruct and  $g$  is the output. The model (as shown in Fig. 1) assumes a 1-D mass distribution of fertilizer at a depth  $h$  below a given land surface during a fertilization operation in a farm. The system model assumes that the functions  $Q$ ,  $f$  and  $g$  are all identically zero outside the unit line. Thus, the limit of integration  $\Omega$  is confined to  $\Omega \subset [0, 1]$  for a one dimensional case<sup>1</sup>. The integral equation in Eqn. (1) can be viewed as a linear operator, operating on the function  $f(t)$  to produce  $g(s)$ . However, the nature of the operator does not often allow it to have a bounded inverse. This implies that an infinitesimal small change  $dg$  in  $g$  at  $P$  can cause a corresponding arbitrarily large change  $df$  in  $f$ . Hence, the ability to solve Eqn. (1) successfully depends largely on the accuracy of  $g(s)$  and the shape of  $Q$ .

### Derivation of Simulation Model

We consider a simplified one dimensional model. Suppose during the fertilization process, an amount of fertilizer is deposited for a given sampled plane stratum at the depth  $h$  beneath the surface of the earth. The geometry and location of the  $s$  and  $t$  axes are shown in Fig. 1. The  $s$ -axis represents the surface of the earth and the line  $y = -h$  the position of the injected mass of magnitude  $f(t)$ . The function  $f(t)$  is what we

wish to find. We are able to only measure  $g(s)$ , the vertical component the earth surface.

Based on the measurements at the surface of the vertical component  $g(s)$ , we want to compute the mass distribution  $f(t)$ , along the  $t$ -axis. Now at  $P$ , the value of  $g$  due the part  $dt$  is

$$dg = \frac{\sin(\theta)}{[\sqrt{h^2 + (s-t)^2}]^2} f(t) dt \quad (2)$$

where the angle  $\theta$  is as shown in Fig. 1.

The total value of  $g(s)$  for any  $s$  is

$$g(s) = \int_0^1 \frac{h}{[h^2 + (s-t)^2]^{3/2}} f(t) dt \quad (3)$$

By comparing Eqn. (3) with Eqn. (1), we have

$$Q(s, t) = \frac{h}{[h^2 + (s-t)^2]^{3/2}} \quad (4)$$

where the kernel  $Q$  defined as  $Q(s, t) = Q(s-t)$  is a function of the difference between  $s$  and  $t$ . In this paper, we have chosen the midpoint quadrature rule as our discretization scheme to be used in constructing an approximation to the definite integral for the interval  $(0, 1)$  of Eqn. (3) in the following manner;

$$t_j = \frac{j-0.5}{n}, \quad w_j = \frac{1}{n} \quad j = 1, 2, \dots, n \quad (5)$$

resulting in the the approximation

$$\int_0^1 Q(s, t) f dt \simeq \sum_{j=1}^n w_j Q(s, t_j) \tilde{f}(t_j) = \varpi(s) \quad (6)$$

where the  $t_j$ 's and  $w_j$ 's are the abscissas and respective weights of the midpoint quadrature rule chosen/used. Next, a collocation method is used to fulfil the requirement for/on  $g$  at the R.H.S for  $m$  ( $m \geq n$ ) given points satisfying;

$$\varpi(s_i) = g(s_i); \quad i = 1, 2, \dots, m$$

In this paper, we will assume that  $m = n$ . By substitution of  $\varpi$  into the collocation scheme, the system of linear equations at  $n$  given points  $s_1, s_2, \dots, s_n$  becomes;

$$\sum_{j=1}^n w_j Q(s_i, t_j) \tilde{f}(t_j) = \sum_{j=1}^n w_j Q(s_i - t_j) \tilde{f}(t_j) = g(s_i) \quad (7)$$

which can be written in the form

$$Kf = g \quad (8)$$

where  $K_{i,j} = w_j Q(s_i - t_j)$ ,  $f_j = \tilde{f}(t_j)$ ,  $g_i = g(s_i)$ ,  $\forall i, j = 1, 2, \dots, n$ , and the naive solution denoted  $\tilde{f}$  is given by;

$$\tilde{f} = K^{-1}g \quad (9)$$

Re-writing Eqn. (7) in a matrix-vector representation results in the following system equation;

$$\begin{pmatrix} k_{1,1} & k_{1,2} & \dots & k_{1,n} \\ k_{2,1} & k_{2,2} & \dots & k_{2,n} \\ k_{3,1} & k_{3,2} & \dots & k_{3,n} \\ \vdots & \vdots & \vdots & \vdots \\ k_{n,1} & k_{n,2} & \dots & k_{n,n} \end{pmatrix} \begin{pmatrix} f_1 \\ f_2 \\ f_3 \\ \vdots \\ f_n \end{pmatrix} = \begin{pmatrix} g_1 \\ g_2 \\ g_3 \\ \vdots \\ g_n \end{pmatrix} \quad (10)$$

<sup>1</sup>For a unit square, we have  $\Omega \subset [0, 1] \times [0, 1]$  and for a unit hypercube  $\Omega \subset [0, 1]_1 \times [0, 1]_2, \dots \times [0, 1]_n$

where  $\mathbf{K}_{i,j} = w_j \mathbf{Q}(s_i, t_j) = \mathbf{k}_{i,j}$ ;  $\mathbf{f}_j = \tilde{\mathbf{f}}(t_j)$ ;  $\mathbf{g}_i = \mathbf{g}(s_i)$  for  $i, j = 1, 2, \dots, n$ .

The system function  $\mathbf{K}$  is then derived from Eqns. (2-10) as

$$\mathbf{K}_{i,j} = w_j \mathbf{Q}(s_i - t_j) = \frac{n^2 h}{(n^2 h^2 + (i - j)^2)^{3/2}}, \quad i, j = 1, \dots, n \quad (11)$$

where  $h$  is the depth.

### Higher Dimensions and Ill-Posed Inverse Problems in Fertilizer Applications

In higher dimension, it can happen that the matrix  $\mathbf{K}$  has many singular values of different magnitudes close to the origin. This renders  $\mathbf{K}$  to have an ill-determined rank. The kernel  $\mathbf{K}$  smooths out high frequency components of the signal which results in loss of information at high frequency components of  $\mathbf{f}$ . Therefore, a least squares fit may not be able to capture the relevant information contained in  $\mathbf{g}$ .

Thus, at higher values of the discretization parameter  $n$ , the naive (or exact) solution denoted  $\tilde{\mathbf{f}}$  (or  $\mathbf{K}^{-1}\mathbf{g}$ ) often gives a poor representation of the true solution. This is when  $\epsilon = 0$ . This assumption is not realistic from a practical viewpoint since the elements of the computed naive vector  $\tilde{\mathbf{f}}$  are, in principle, mere approximations to the desired solution and they do not reflect on the true solution. Hence, Eqn.(8) is infact  $\mathbf{K}\mathbf{f} + \epsilon = \mathbf{g}$  with the additive noise vector  $\epsilon$  representing perturbation of the exact/true data. A good representation of the true solution is only attainable when  $\epsilon$  is non-zero. By introducing the matrix notations;  $\mathbf{K}_{i,j} = w_i \kappa_{i,j}$  and  $\mathbf{K}^{-1}_{i,j} = \Xi_{i,j}$ , it is easy to show that

$$\mathbf{f} = \mathbf{K}^{-1}\mathbf{g} + \mathbf{K}^{-1}\epsilon \quad (12)$$

and it explains the reason why the behavior of the kernel  $\mathbf{K}$  must be taken into consideration since  $\mathbf{f}$  of Eqn. (12) is a linear function of  $\mathbf{g}$  and  $\epsilon$ . By taking the partial derivatives of  $\mathbf{f}_i$  with respect to either  $\mathbf{g}_j$  or  $\epsilon_j$  is the inverse kernel  $\Xi_{i,j}$ ;

$$\frac{\partial \mathbf{f}_i}{\partial \mathbf{g}_j} = \Xi_{i,j} = \frac{\partial \mathbf{f}_i}{\partial \epsilon_j} \quad ; \quad i, j = 1, 2, \dots, n$$

In the situation where the dimension is high, the matrix  $\mathbf{K}$  may be rank deficient, hence a stable inverse does not exist. The difficulties associated with such instability is often due to the fact that in practice the specification of  $\mathbf{g}_i$  is usually inexact because of the data at hand. This means that the actual (or true) data  $\mathbf{g}$  are often corrupted with some noisy samples at certain discrete abscissas  $1, 2, \dots, n$ . This can lead to an ill-conditioned inverse problem in contrast to a well-conditioned inverse problem. Thus, we have

$$\mathbf{g} = \mathbf{K}\mathbf{f} + \epsilon = \mu + \epsilon \quad (13)$$

where  $\mathbf{g}$  is the measurement data,  $\epsilon$  is an arbitrary function referred to as measurement noise which is measured based on some condition about the size and  $\mu = \mathbf{K}\mathbf{f}$  is an estimate of  $\mathbf{g}$ . The problem statement is often related to a functional inequality  $|\epsilon|$  bounded above such that

$$|\epsilon| \leq \mathbf{M} \quad \text{or} \quad \int_{\Sigma} w'(s)\epsilon^2(s)ds \leq \tilde{\mathbf{M}}$$

where  $w'$  are weights.

The condition on the magnitude of  $\epsilon$  is defined as

$$\sum_{i=0}^n \epsilon_i = \epsilon^2$$

where  $\epsilon^2$  is a constant.

### Fertilization Process Model

The fertilization process model of sampling each subpopulation (stratum) of fertilizer deposits independently is given by the linear model;

$$\mathbf{g} = \mathbf{K}\mathbf{f} + \epsilon = \mu + \epsilon \quad (14)$$

where  $\mathbf{g} \in \mathbf{R}^n$  is the measurement data,  $\mathbf{K} \in \mathbf{R}^{n \times n}$  is the discretized kernel of rank  $n$ ,  $\mu \in \mathbf{R}^n$  is an estimate of the measurement data  $\mathbf{g}$ ,  $\mathbf{f} \in \mathbf{R}^n$  is the input, and  $\epsilon \in \mathbf{R}^n$  is additive white Gaussian noise with mean zero and variance  $\sigma^2 I$ . Our emphasis is on the case where the least squares do not make sense when put into the context of the generator of the measurement data  $\mathbf{g}$ . Fig. (2) illustrates the sparse representation of the fertilization process that adequately represents the size of the population for ensuring credibility, reliability and validity of information collected for the whole monitoring framework and thus bring a big picture perspective to the problem.

From Fig. (2) we can write the stratified sampled data of the fertilization process as below,

$$\begin{aligned} \mathbf{g}_1 &= [\mathbf{g}_{(1,1)} \quad \mathbf{g}_{(2,1)} \quad \dots \quad \mathbf{g}_{(m,1)}]^T \\ \mathbf{g}_2 &= [\mathbf{g}_{(m+1,2)} \quad \mathbf{g}_{(m+2,2)} \quad \dots \quad \mathbf{g}_{(2m,2)}]^T \\ \mathbf{g}_3 &= [\mathbf{g}_{(2m+1,3)} \quad \mathbf{g}_{(2m+2,3)} \quad \dots \quad \mathbf{g}_{(3m,3)}]^T \\ &\vdots \end{aligned}$$

$$\mathbf{g}_n = [\mathbf{g}_{mn-n,n} \quad \mathbf{g}_{mn-n+1,n} \quad \mathbf{g}_{mn-n+2,3} \quad \dots \quad \mathbf{g}_{mn,n}]^T$$

At a depth  $h$  set by the machinery operator, the amount/volume

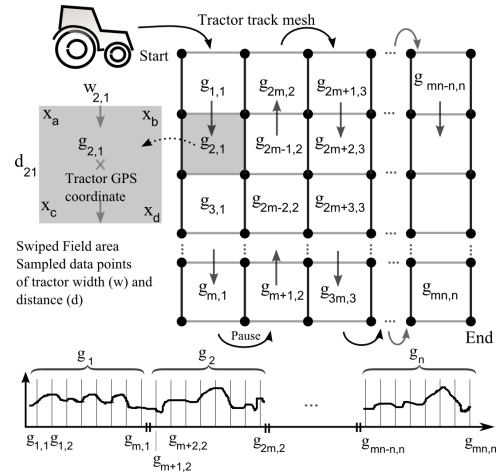


Fig. 2. The fertilization process and vectorization transformation of fertilizer deposits into an  $mn \times 1$  column vector obtained by stacking the columns of the matrix  $\mathbf{g}$  of data on top of one another

of fertilizer (in kg per hectare) deposited into the soil in the sampled areas are measured and specified using an ordered set of components in an  $mn$ -dimensional vector,  $vec(\mathbf{g}) \in \mathbf{R}^{mn}$

$$\mathbf{g} = [\mathbf{g}_1^T \quad \mathbf{g}_2^T \quad \mathbf{g}_3^T \quad \mathbf{g}_4^T \quad \dots \quad \mathbf{g}_{n-1}^T \quad \mathbf{g}_n^T] \quad (15)$$

which can be viewed as a vector  $\mathbf{g} \in \mathbf{R}^{mn}$ . Our objective is to determine an estimate of fertilizer deposited at a known distance  $h$  beneath the in the soil surface.

### Truncated SVD Regularization and Discrete Picard Condition of the Fertilization Process Model

The general principle of rank reduction based on SVD is treated in [6]. Given an integer  $k \leq n$ , we partition  $\mathbf{K}$  by SVD according to

$$\mathbf{K} = (\mathbf{U}_k, \mathbf{U}_0) \begin{pmatrix} \mathbf{D}_k & 0 \\ 0 & \mathbf{D}_0 \end{pmatrix} (\mathbf{V}_k, \mathbf{V}_0)^T \quad (16)$$

where  $\mathbf{D}_k = \text{diag}(\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_k)$  and  $\mathbf{D}_0 = \text{diag}(\mathbf{d}_{k+1}, \mathbf{d}_{k+2}, \dots, \mathbf{d}_n)$  consists of the  $k$  largest and  $(n - k)$  smallest singular values respectively. The matrix  $\mathbf{K}_k$  defined by

$$\mathbf{K}_k = \mathbf{U}_k \mathbf{D}_k \mathbf{V}_k^T$$

is considered to be an approximation to the original matrix  $\mathbf{K}$  with a corresponding decrease in the rank from  $n$  to  $k$ . For some large enough values of the discretization parameter  $n$ , the sequence of true data values  $\{\mathbf{u}_i^T(\mathbf{g} - \epsilon)\}$  goes to zero faster than the sequence of singular values  $\mathbf{d}_i$ . In this case, for some terms greater than or equal to some parameter  $k$ ,  $\mathbf{u}_i^T(\mathbf{g} - \epsilon) \approx \infty$ . The effect of these errors is that the singular values  $\mathbf{d}_i$  and the coefficients  $\mathbf{u}_i^T \mathbf{g}$  also called the Fourier coefficients do not decrease monotonically as expected.

From Eqn. (9), we decompose the naive solution  $\tilde{\mathbf{f}} = \mathbf{K}^{-1} \mathbf{g}$  by SVD and change the summation interval to a given choice of  $k$ . For a given choice  $k$ , we have the solution  $\mathbf{f}_k$  given by

$$\mathbf{f}_k = \sum_{i=1}^k \frac{\mathbf{u}_i^T \mathbf{g}}{\mathbf{d}_i} \mathbf{v}_i \quad (17)$$

Hence, if there exists a square integrable solution  $\mathbf{f}$  to the integral equation in Eqn. (1), then the left-hand side  $\mathbf{g}$  must satisfy;

$$\sum_{i=1}^{\infty} \left( \frac{\mathbf{u}_i, \mathbf{g}}{\mathbf{d}_i} \right)^2 < \infty$$

where  $(\cdot, \cdot)$  is defined by

$$(\phi, \psi) = \int \phi(t) \psi(t) dt$$

### Tikhonov Based Regularization from Numerical Methodology Viewpoint

The solution denoted  $f_{\lambda_n}$  is obtained by minimizing the weighted combination of the residual norm and the added smoothness constraint. The Tikhonov solution  $f_{\lambda_n}$  is given by

$$\nabla_{\mathbf{f}} \left\{ \frac{1}{2} \|\mathbf{g} - \mathbf{K} \mathbf{f}\|_2^2 + \frac{\lambda_n^2}{2} \|\mathbf{L}(\mathbf{f} - \mathbf{f}_0)\|_2^2 \right\} = 0 \quad (18)$$

where  $\nabla$  denotes the vector differential operator,  $del$  and  $\lambda_n$  is the numerical regularization parameter used to control the degree of smoothness of the solution. In this paper  $L = I$  and that we are also unaware of any initial estimate  $\mathbf{f}_0$  of the

solution. Hence  $\mathbf{f}_0 = 0$ .

The regularized solution  $\mathbf{f}_{\lambda_n}$  given by

$$\mathbf{f}_{\lambda_n} = (\mathbf{K}^T \mathbf{K} + \lambda_n^2 \mathbf{I})^{-1} \mathbf{K}^T \mathbf{g} \quad (19)$$

The parameter  $\lambda_n$  is the numerical methods approach regularization parameter. The SVD of the regularized solution of Eqn. 19

$$\mathbf{f}_{\lambda_n} = \sum_{i=1}^n \left( \frac{\mathbf{d}_i^2}{\mathbf{d}_i^2 + \lambda_n^2} \right) \frac{\mathbf{u}_i^T \mathbf{g}}{\mathbf{d}_i} \mathbf{v}_i \quad (20)$$

### Statistical Bayesian Based Regularization from Maximum aPriori (MAP) Methodology Viewpoint

We compute an estimate for the posterior probability of  $\mathbf{f}$  given  $\mathbf{g}$ ,  $\sigma^2$  and  $\lambda_s^2$  via Bayes' rule;

$$p(\mathbf{f}|\mathbf{g}, \sigma^2, \lambda_s^2) = \frac{p(\mathbf{g}|\mathbf{f}, \sigma^2) p(\mathbf{f}|\lambda_s^2)}{p(\mathbf{g}|\lambda_s, \sigma^2)} \quad (21)$$

where  $[p(\mathbf{g}|\lambda_s, \sigma^2) = \int p(\mathbf{g}|\mathbf{f}, \sigma^2) p(\mathbf{f}|\lambda_s^2) d\mathbf{f}]$  and the densities  $p(\mathbf{g}|\mathbf{f}, \sigma^2)$  and  $p(\mathbf{f}|\lambda_s^2)$  are the likelihood and prior probability respectively. The logarithm of the posterior;

$$\log p(\mathbf{f}|\mathbf{g}, \sigma^2, \lambda_s^2) = -\frac{1}{2} \left\{ \frac{1}{\sigma^2} \|\mathbf{g} - \mathbf{K} \mathbf{f}\|_2^2 + \lambda_s^2 \|\mathbf{f}\|_2^2 \right\} + \kappa \quad (22)$$

where  $\lambda_s$  is the statistical MAP regularization parameter used to control the smoothness of the solution and  $\kappa$  is a constant given by

$$\kappa = \log \left\{ \tilde{C}_l(\sigma) \tilde{C}_p(\lambda_s) [p(\mathbf{g}|\sigma^2, \lambda_s^2)]^{-1} \right\}$$

$\tilde{C}_l(\sigma)$  is the normalizing factor of the least squares paradigm given by  $\left\{ \int \exp -\frac{1}{2\sigma^2} \|\mathbf{g} - \mathbf{K} \mathbf{f}\|_2^2 d\mathbf{f} \right\}^{-1}$  and  $\tilde{C}_p(\lambda_s)$  is the normalization factor of the prior probability given by  $\left\{ \int \exp -\frac{\lambda_s^2}{2} \|\mathbf{f}\|_2^2 d\mathbf{f} \right\}^{-1}$ . The regularized solution  $\mathbf{f}_{\lambda_s}$  lies on the stationary point satisfying

$$\mathbf{f}_{\lambda_s} = (\mathbf{K}^T \mathbf{K} + \sigma^2 \lambda_s^2 \mathbf{I})^{-1} \mathbf{K}^T \mathbf{g} \quad (23)$$

The SVD of the regularized solution of Eqn. 23

$$\mathbf{f}_{\lambda_s} = \sum_{i=1}^n \left( \frac{\mathbf{d}_i^2}{\mathbf{d}_i^2 + \lambda_s^2 \sigma^2} \right) \frac{\mathbf{u}_i^T \mathbf{g}}{\mathbf{d}_i} \mathbf{v}_i \quad (24)$$

By comparison of Eqns. (20) and (24), we have

$$\lambda_n = \lambda_s \sigma \quad (25)$$

### Precision Matrix of the Regularized Solution and Error Bars

We take the second order partial derivatives of the log-posterior with respect to  $\mathbf{f}$  to obtain the curvature information;

$$J(\lambda_s, \sigma) = -\nabla_{\mathbf{f}}^2 \log [p(\mathbf{f}|\mathbf{g}, \sigma^2, \lambda_s^2)] = \frac{\mathbf{K}^T \mathbf{K}}{\sigma^2} + \lambda_s^2 \mathbf{I} \quad (26)$$

where  $J$  in (26) is independent of  $\mathbf{f}$  and the expectation of  $J$  denoted  $\Sigma_{\mathbf{f}_\lambda}$  is  $J$ .

The precision matrix which is the inverse matrix  $\Sigma_f^{-1}$  and the decomposition of  $\Sigma_f^{-1}$  by SVD yields;

$$\Sigma_f^{-1} = \left( \frac{\mathbf{K}^T \mathbf{K}}{\sigma^2} + \lambda_s^2 \mathbf{I} \right)^{-1} = \mathbf{V} \left( \frac{D^2}{\sigma^2} + \lambda_s^2 \mathbf{I} \right) \mathbf{V}^T \quad (27)$$

The construction of the error bars on  $\mathbf{f}_{\lambda_s}$  and the corresponding approximation is given by the expression

$$\mathbf{f}_{\lambda_s} \pm \sqrt{\text{diag}[\tilde{\Sigma}_{\mathbf{f}_{\lambda_s}}]} \simeq \mathbf{f}_{\lambda_s} \pm \sqrt{\text{diag}[\Sigma_f^{-1}]}$$

where  $\tilde{\Sigma}_{\mathbf{f}_{\lambda_s}}$  is the variance-covariance

$$\tilde{\Sigma}_{\mathbf{f}_{\lambda_s}} = \sigma^2 \sum_{i=1}^n v_i \left( \frac{d_i}{d_i^2 + \lambda_s^2 \sigma^2} \right)^2 v_i^T \quad (28)$$

We can choose to evaluate the parameters  $\lambda_s$  and  $\sigma$  using the Evidence Framework [7] or choose to optimize the parameters using Variational Methods [3] and [7]. Furthermore, we can also choose to estimate/determine the parameter  $\lambda_n$  by either using the L-Curve method or be calculated from Eqn. 25 using the values obtained for  $\lambda_s$  and  $\sigma$ .

## B. Data Fusion

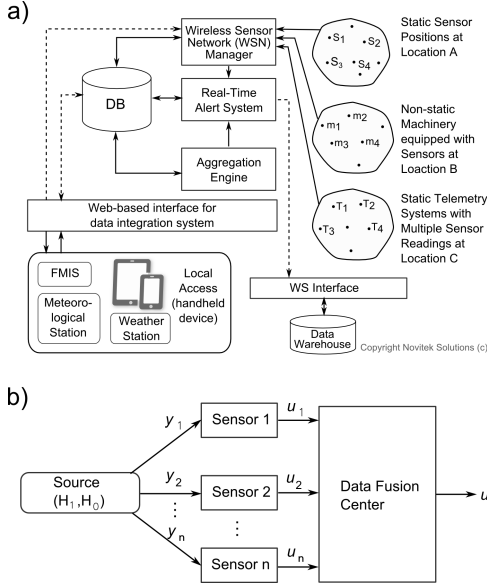


Fig. 3. a) A functional overview of the farm production process architecture. b) Data fusion function blocks for crop production process and application

As shown in Fig. 3, the architecture is built around a Database management system (DBMS), which provides the main repository for the data gathered from the Wireless Sensor Network (WSN) comprising; a) static sensors at Location A, b) Non-static machinery equipped with sensors at Location B, and static telemetry systems with multiple-sensor readings capabilities at location C. Includes, WSN operating rules, and three applications of a specific programming language (e.g., java): WSN manager, the Aggregation Engine and a Real-time Alert System, RTAS. The WSN manager links the WSN which produces the data, to the DBMS where the data are stored or aggregated. The WSN are the producers of the massive amounts of heterogeneous information. The

data aggregation engine takes the data produced by the WSN and summarises them by aggregation and/or perform signal processing. The Real-Time alert system (RTAS) reacts to both real-time sensor readings stored in the DBMS and the data produced by the aggregation engine. An integral part of the RTAS's job/functions is to check if the sensor readings are within certain (programmable) limits.

Based on the description of the system architecture of Fig. 3, complete sensor observations are available at a central location and a classical hypothesis testing procedures are employed for signal processing. The basic goal of data fusion is to improve system performance, (e.g., reliability or speed).

## Formulation of the Data Fusion Problem

Consider the detection of a known signal in additive Gaussian noise with  $n$  sensors. We will formulate the problem as a binary hypothesis testing problem with the following two hypotheses:

$$H_0 : \mathbf{y}_k = \mathbf{n}_k \quad ; \quad H_1 : \mathbf{y}_k = \mathbf{x}_k + \mathbf{n}_k \quad (29)$$

where  $H_0$  is the hypothesis when the signal is absent and  $H_1$  is the hypothesis when the signal is present [16].

## C. Crop Growth Modelling, Simulation and Visualization

The articles [15] and [12], presents dynamical models for crop growth. Especially [15] is interesting, because it presents the models supplemented by a computational environment based on the R language. Among others issues, it remains to obtain parameter sets for barley growth in the geographical region of interest. With respect to some data mining tools and programming languages (such as R, WEKA, RapidMiner), including visualization of data acquisitions, parameter estimations/fusions/decisions and the corresponding sensitivities, the articles [11] and [15] are of relevance. Furthermore, the references [15] and [14], in addition to the above mentioned modelling, might form a basis for further work in an open source environment, because of the R package ZeBook, for agricultural modelling and simulation.

## IV. SIMULATION RESULTS

We will assume that some mass of fertilizer (e.g., N,P or/and K) has been equally and evenly applied to a soil at a depth  $h = 0.25$  at a given farm location/area. The discretized data sequence  $(t_i); i = 1, 2, \dots, n$  (where,  $n = 60$  and  $n = 100$  are the two discretized parameter sequence to be used for the simulation). Moreover, we assume that the mass of controlled input of fertilizer is expected to be  $m(t_i) = 5g \quad \forall t_i > 0$ . We further make the assumption that a change in the mass of the controlled input occurs for each input at any area/location due to the flow patterns of the soil water and nutrients. For this simulation, we will define the true input data  $\mathbf{f}$  which is a function of time  $t$  as a superposition of the constant mass  $m$  and some additive noise;

$$f(t) = m(t) + 0.01 * randn(n, 1)$$

We will also assume that the measurement data (or observation data)  $\mathbf{g}$  is given by

$$\mathbf{g} = \mathbf{K}\mathbf{f} + 0.5 * \text{randn}(60, 1)$$

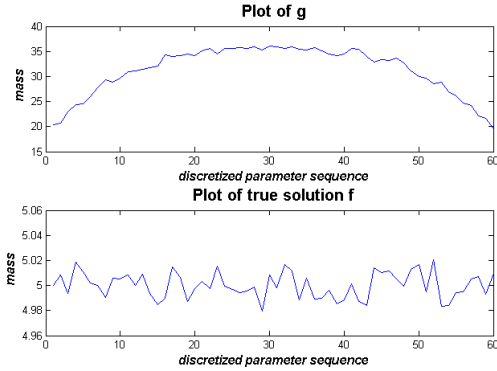


Fig. 4. *Top)* The output signal  $\mathbf{g}$ . *Bottom)* The true signal/data  $\mathbf{f}$

The elements of the computed naive solution is

$$\tilde{\mathbf{f}} = \mathbf{K}^{-1}\mathbf{g}$$

The Least Squares solution [6] denoted  $f_{ls}$  of  $\mathbf{g}$  is computed using the solution to the normal equation given by

$$f_{ls} = (\mathbf{K}^T\mathbf{K})^{-1}\mathbf{K}^T\mathbf{g}$$

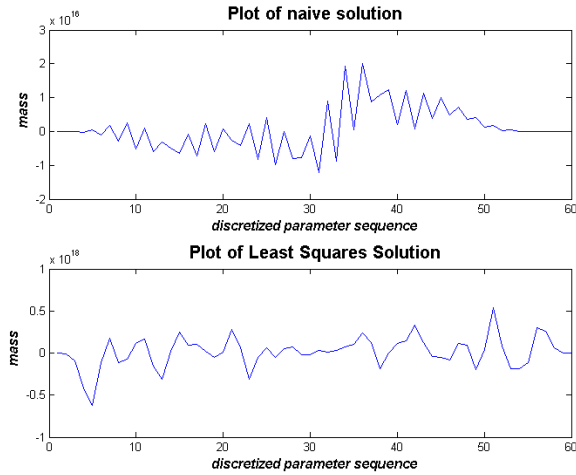


Fig. 5. *Top)* The naive solution  $\mathbf{f}_{naive}$ . *Bottom)* The least squares solution  $\mathbf{f}_{ls}$ .

From Fig. 4, it can be seen that the range of values of the measurement signal/data  $\mathbf{g}$  and its characteristics have no similarity with the true signal  $\mathbf{f}$ . The range of values of the true signal  $\mathbf{f}$  is between 4.5 and 5.5 (i.e.,  $4.5 < f < 5.6$ ) and the range of values of  $\mathbf{g}$  is between 15 and 40 (i.e.,  $15 < g < 40$ ).

Fig. 5, shows the estimates (or the reconstructed signal)  $\mathbf{f}_{naive}$  and  $\mathbf{f}_{ls}$  of  $\mathbf{g}$  computed using  $\mathbf{K}^{-1}\mathbf{g}$  and  $(\mathbf{K}^T\mathbf{K})^{-1}\mathbf{K}^T\mathbf{g}$ . It can be seen from Fig. 5 that the range of values of the estimated signals  $\mathbf{f}_{naive}$  and  $\mathbf{f}_{ls}$  based on direct inversion and the least squares solution gives a poor representation

of the true signal  $\mathbf{f}$  of Fig. 4-*Bottom*). Thus, both figures of  $\mathbf{f}_{naive}$  and  $\mathbf{f}_{ls}$  are no where near the true signal/data  $\mathbf{f}$ . Furthermore, the signal characteristics of both  $\mathbf{f}_{naive}$  and  $\mathbf{f}_{ls}$  has no resemblance of the true signal  $\mathbf{f}$ . The range of values of  $\mathbf{f}_{naive}$  and  $\mathbf{f}_{ls}$  are of the order of  $10^{16}$  and  $10^{18}$  respectively. The results of the reconstructed signals  $\mathbf{f}_{naive}$  and  $\mathbf{f}_{ls}$  shows that direct inversion methods and Least Squares solution methods cannot be used to reconstruct (or find a good estimate of) the true solution  $\mathbf{f}$ . The reason is that the integral equation is not square integrable due to the amplifications by the singular values  $d_i$  of the matrix  $\mathbf{K}$ . This explains why the Least Squares solution and direct inversion methods cannot be used to reconstruct or find a good estimate of the true signal  $\mathbf{f}$ . This can be seen from the Picard plot in Fig. 6.

Fig. 6, strongly shows that the Picard condition has not been satisfied for the integral equation of the fertilization process problem and that it is impossible to compute a numerical solution. All the the coefficients  $|\mathbf{u}_i^T\mathbf{g}|$  decay slower than the singular values  $d_i$ . The noise comes from both rounding errors and additive noise.

Fig. 7, shows the estimates  $\mathbf{f}_{\lambda_n}$  obtained for two different choices of the regularization parameters,  $\lambda_n^2 = \lambda_{rls}^2 = 0.097843$  and  $\lambda_n^2 = \lambda_{rls}^2 = 0.094469$  for the cases where  $n = 60$  and  $n = 100$  respectively using Eqn. (19). It can be seen that the range of values of the estimated signals for the two regularization parameters for both  $n = 60$  and  $n = 100$  are within the range  $4 < \mathbf{f}_{\lambda_n} < 5.6$  of the original signal  $\mathbf{f}$ . This shows that by computing the optimal parameters using the L-Curve method or the Evidence framework will definitely improve the estimates.

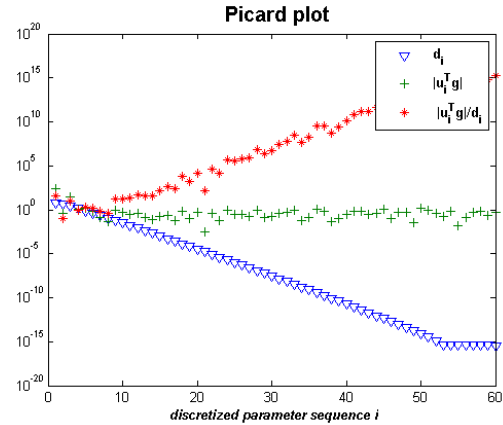


Fig. 6. Picard plot of the singular values  $d_i$  and the coefficients  $|\mathbf{u}_i^T\mathbf{g}|$  and  $|\frac{\mathbf{u}_i^T\mathbf{g}}{d_i}|$  for the fertilization process problem which has no square integrable solution. All the the coefficients  $|\mathbf{u}_i^T\mathbf{g}|$  decay slower than the singular values  $d_i$ . The noise comes from both rounding errors and additive noise.

## V. CONCLUSION

The focus of this paper was ill-posed inverse problems in crop and production systems, processes and related applications. We emphasized on the Tikhonovs functional form of regularization from both numerical and Statistical methods viewpoints. We extended the concept of regularization to the Bayesian methods framework. The Bayesian paradigm

provides a general unified framework to the ill-posed inverse problem. We showed that the conditional mean of the inverse problem in the statistical maximum a priori (MAP) framework is the same as that of the regularized normal equations in the numerical framework if we are dealing with multivariate Gaussian distribution. We can consider the regularized normal equations in the numerical methods framework as a special case of stochastic modelling theory when we are dealing with Gaussian random variables. Furthermore, the tuning parameter  $\lambda_n$  in the numerical methods framework is a product of the Bayesian statistical regularization parameter  $\lambda_s$  and the noise level  $\sigma$ . What remains to be treated are methods for computing the noise level  $\sigma$  and regularization parameters  $\lambda_n$  and  $\lambda_s$  and also be able to compute the error bars.

## VI. FUTURE WORK

The near and medium future development and research efforts will focus on the following areas

- Aim at reducing the problem of parameter estimation in the Bayesian framework to an optimization problem through the Variational methods approach by deriving an optimization algorithm based on the Variational Bayesian Inference approach for the inverse problem and show that any optimum of the Bayesian framework will also correspond to an optimum of the Variational Bayesian Inference algorithm.
- Aim at developing a new framework for Sequential Bayesian estimation in sensor networks, which will focus on measurements of localized sensors and fusion of both received measurements and missing ones at the fusion centre (FC).
- Aim at obtaining, for the present growth cycle, which starts approximately April 2015, a stratified sampling of a barley growth field such that a site specific characterization is obtained, and can be used for planning of the present and the future growth cycles.
- Aim at obtaining an initial model for deciding when the spraying process is carried out in a barley produc-

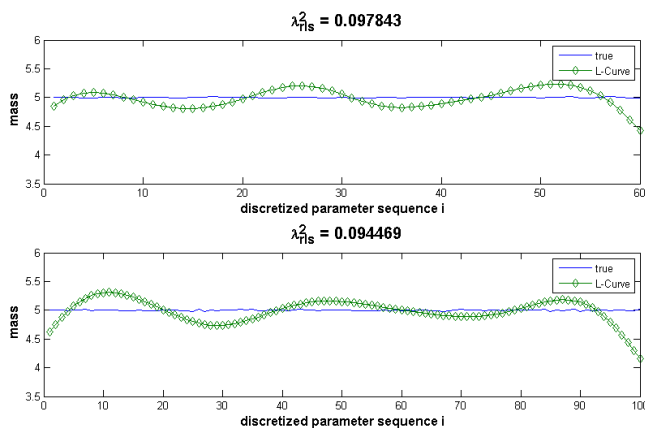


Fig. 7. Estimates  $\hat{\lambda}_n$  of the Tikhonov's based regularization for two different choices of the regularization parameters,  $\lambda_n^2 = \lambda_{r,l,s}^2 = 0.097843$  and  $\lambda_n^2 = \lambda_{r,l,s}^2 = 0.094469$  for the cases where  $n = 60$  and  $n = 100$  respectively using Eqn. (19).

tion cycle and research if the spraying could/should be augmented with a second simultaneous fertilization.

- Setting up an open source environment, which can be used for future modelling, simulation, visualization and statistical analysis and possibly also documentation, targeting a basis for Fig 3.
- Building a barley production cycle growth model, e.g. using the R package [14] and the corresponding models in [15] and [1].

## ACKNOWLEDGMENT

This project is carried out in close collaboration with Prof. Dr. Joseph Küng, Institute for Application Oriented Knowledge Processing, JKU, Linz. This research was funded by the FP7 - EU FRAMEWORK PROGRAMME under grant agreement No. 604659 (project CLAFIS).

## REFERENCES

- [1] V.I. Adamchuck, J.V. Hummel, M.T. Morgan, S.K. Upadhyaya, *On-the-go soil sensors for precision agriculture*, Computers and Electronics in Agriculture 44 (2004) 7191, Elsevier, 2004.
- [2] P.M. Anderson, E.A. Oelke, and S.R. Simmons, *Growth and development guide for spring barley*, 2013 [www.extension.umn.edu/agriculture/small-grains/growth-and-development/spring-barley/](http://www.extension.umn.edu/agriculture/small-grains/growth-and-development/spring-barley/)
- [3] Christopher Bishop, *Pattern Recognition and machine Learning*, Springer, 2006.
- [4] G.M. Wing, *A Primer on Integral Equations of the First Kind: the problem of deconvolution and unfolding* Society of Industrial and Applied mathematics, 1991.
- [5] J.L. Drouet, et al. *Analysis for models of greenhouse gas emissions at farm level. Case study of N<sub>2</sub>O emissions simulated by the CERES-ECG model*, Environmental Pollution, Elsevier 1-6, 2011.
- [6] Gene Golub, Charles F. Van Loan, *Matrix Computations*, 4th Edition, The John Hopkins University Press, 2013.
- [7] David J.C. Mackay, *Information Theory, Inference and Learning Algorithms*, Cambridge University Press, 2014
- [8] Federico Guerrini, *The future of Agriculture: Smart Farming*, 2015 [www.forbes.com/sites/federicoguerrini/2015/02/18/the-future-of-agriculture-smart-farming/](http://www.forbes.com/sites/federicoguerrini/2015/02/18/the-future-of-agriculture-smart-farming/)
- [9] Hak-Jim Kim, Kenneth A. Sudduth, John W. Hummel, *Soil macronutrient sensing for precision agriculture*, Journal of Environmental Monitoring, 2009, 11, pp. 1810-1824.
- [10] P. Kabat, R.W.A. Hutjes, R.A. Feddes, *The scaling characteristics of soil parameters: From plot scale heterogeneity to subgrid parameterization*, Journal of Hydrology 190 (1997) pp. 363-396.
- [11] Hilda Kosorus, Jürgen Hönlgl, Josef Küng, *Using R, WEKA and RapidMiner in Time Series Analysis of Sensor Data for Structural Health Monitoring*, 22nd International Workshop on Database and Expert Systems Applications pp. 306-310, IEEE Computer Society, 2011.
- [12] Allen R. Overman, Richard V. Scholtz III, *Mathematical Models of Crop Growth and Yield*, Dekker 2002.
- [13] Liisa Pietola, Risto Tanni, Paavo Elonen, *Responses of yield and N use of spring sown crops to N fertilization, with special reference to the use of plant growth regulators*, Agricultural and Food Science Finland, Vol. 8, No. 4-5, August 1999.
- [14] D. Wallach, D. Makowski, J. W. Jones, F. Brun, *ZeBook: ZeBook Working with dynamic models for agriculture and environment*, 2013. [cran.r-project.org/web/packages/ZeBook/index.html](http://cran.r-project.org/web/packages/ZeBook/index.html)
- [15] D. Wallach, D. Makowski, J. W. Jones, F. Brun, *Working with Dynamic Crop Models - Methods, Tools, and Examples for Agriculture and Environment*, Elsevier 2014.
- [16] Z. Chair, P. K. Varshney, *Optimal Data Fusion in Multiple Sensor Detection Systems*, IEEE Transactions on Aerospace and Electronic Systems, vol AES-22.