The Cryosphere Discuss., 3, S80–S91, 2009
www.the-cryosphere-discuss.net/3/S80/2009/

The Cryosphere
Discussions

TCD

3, S80–S91, 2009

Interactive
Comment

Interactive comment on "Estimating basal properties of glaciers from surface measurements: a non-linear Bayesian inversion approach" *by* M. J. Raymond and G. H. Gudmundsson

O. Eisen (Referee)

olaf.eisen@awi.de

Received and published: 23 March 2009

General comments

This finally is the paper on the retrievability of basal disturbances from surface measurements of non-linear ice masses in a series of papers. As before, the investigation builds on synthetic data from a finite-element forward model. As already announced in earlier papers (e.g. in TCD, 2, 413-445, 2008), this paper extends previous utilisations of a linear inverse method based on Bayes theory to cases of large perturbation, non-linear media and a non-linear sliding law. The results presented here finally show that the whole approach is applicable to real ice masses (with inherently non-linear properties), making it a worthy contribution to the journal. I have no criticism regarding the fundamental findings of this paper and therefore recommend publication. However, I have several questions/comments regarding the clarity of the presentation, which might make it more easily accessible to a wider readership.

Specific comments

Major issues

(Notation: page.line; [. . .]:deleted; bold: included/corrected)

- I think the outline of the inverse procedure in the introduction can be made clearer by stating (e.g. 184.4) that the actual inverse relation (the equation) between model and data results from Bayes theorem, which yields a number of solutions. From this posterior probability distribution the iterative optimization then selects the MAP from this distribution. This is also said so far but somewhat more distributed in the text.
- The Fréchet derivative matrix for a finite space is actually the Jabocian matrix. I assume the latter is known to a much wider readership than the Fréchet derivative, so it should be mentioned. Moreover, I wonder why the whole concept is actually discussed with validity for a quite general Banach space with Fréchet derivatives. (Maybe there is a reason, but I can't see it). As the authors will always use a finite N of discretization points, it would be possible to stick to a simpler description with a Jacobian matrix in $\mathcal{R}^N$ and thus likely increase the number of readers who will actually bother to go into the details of this work, rather then scaring some off with mathematical terms previsously not heard of.

- 204.17 "We determine . . . a posteriori error covariance matrix". This is a fundamental claim for (and partly the purpose of) an inverse approach. So far I did not find the calculation of the a posteriori error covariance matrix in the paper, only the (BLUE-updated) surface data covariance matrix. I assume that the authors refer to the error covariance matrix of the solution, the system state $\mathbf{m}$. If they indeed calculated it, but I overlooked it, then I would expect to have this error estimate displayed graphically somewhow for some of the model experiments. The only aspect of an a-posteriori estimate of accuracy I recognized is the size of $J(\hat{\mathbf{m}})$ discussed in section 5.10. The usage of the final value of $J(\mathbf{m})$ for the decision on whether the whole inverse method with the selected set of starting parameters etc. makes sense should be stressed more, as pointed out in section 5.10.

- I am somewhat unhappy with the structure of section 5. I would find it more natural to include all experiments with a step-wise increase of non-linear contributions with a constant rate factor in a subsection 5.3 and label the experiments a)-f) or alike. These labels should be listed in Table 1 for easier reference to the parameters used instead of identifying them in the text. New section 5.4 would contain the experiment with the non-constant rate factor (old 5.9) and 5.5. the comments on the forward model errors (old 5.10).

  In the same section: 200.5 or 201.4: Explicitly mentioning that "now noise-degraded data" are used somehow implies that this was not the case before (which I doubt from what is stated on 196.6). I was puzzled several times in a similar manner while going through these subsections presenting the results, as for some experiments properties were mentioned, but for other not, although they were apparently the same. I therefore recommend that in these subsections only those physical properties/parameters are mentioned, which differ from other experiments. Clear separation should be made between the experiments already presented in earlier papers. For instance, note explicitly in Table 1 which experiments are already discussed in Gudmundsson & Raymond 2008 (e.g. in an additional column). This helps the reader to focus on the new experiments.

- Issue of terminology:
  In the past a native english referee (working with inverse models) recommended to me to pay more attention to the way "inversion", "invert for", etc. are used. I pass this recommendation on to you. For example, what does it mean if we perform a "surface inversion on ice streams"? Strictly speaking, don't we turn the surface upside down? I know an exact formulation is usually longer than this rather sloppy usage, but it improves clarity. It would be more accurate to use a running title such as "Non-linear inverse approach applied to synthetic data" (although this might be too long). In the title it would be more appropriate to use "Bayesian inverse approach". Likewise, use "inverse method" instead of "inversion method" in the text.

- Inversion vs. optimization procedure:
  The iteration is part of the whole procedure merely to select the MAP from the set of solutions (or the pdf), usually referred to as optimization in this manuscript. At several places the meaning is not entirely clear:
  191.20 "the inversion is most easily done in Fourier space": Unclear what is meant with inversion. The actual inversion of the forward problem takes place in equation (7). I think the analysis in Fourier space merely applies to the iteration (optimization) and the determination of the MAP solution, right?
  194.4 "optimization of the objective function": isn't it rather "the optimization to find the MAP solution of the objective function"?
  195.1 rather then using "inversion step" for point iv I would suggest to use optimization/iteration, as referred to in the text several times.
  204.3 again, from the the terminology used I suggest that the optimiation/iteration has been stopped rather than the inversion.

TCD

3, S80–S91, 2009

Interactive Comment

- Another comment on terminology (not validity of the whole method):
  I am somewhat puzzled that the authors claim to use the Gauss-Newton method. As I am not an expert in numerical mathematics, I might be mistaken here, but to my understanding one characteristic of the Gauss-Newton method is that one does not need the second derivative of the cost function (i.e. the Hessian used in eq. (13) and later on) but Gauss-Newton uses the normal equation in the interation instead. Maybe I get something wrong in the first paragraphy of p.191, an approximation or alike, but then it should be clarified in the text.

- Convergence 1:
  It is known that the degree of convergence of the Newton and Gauss-Newton methods heavily depend on the initial values. This becomes evident in several experiments in section 5. The authors use the maximum a-posteriori probability. With the considerable number of parameters involved in dermining this maximum, it seems likely that several local probability maxima exist, corresponding the minimum criterion of the cost function J. The author somehow confirm this in their discussion of the forward model parameter error in section 5.10. I therefore ask the authors for some more discussion of this issue in the manuscript, i.e. can we indeed talk about MAP although it is might just be a local maximum? Overall, the feasibility of the MAP depends on the choice of a-prior parameters!

- Convergence 2:
  In several experiments (e.g. section 5.6) an oscillating behaviour of the iteration towards the true solution is observed. This behaviour is also known for Newton's method and can be treated by introducing a damping term, which decreases the sensitivity of the method in respect to the choice of initialization parameters. Luckily, in all examples presented here (sections 5.3-5.9), the iterations converge to the true solution, although in two cases external forcing had to be used (see 200.18, 201.18 below), which seems somewhat subjective. Did the authors spend any thoughts on whether this external forcing could be avoided by

Full Screen / Esc
Printer-friendly Version
Interactive Discussion
Discussion Paper

using a damping factor in the iteration (12)?

- Which linearization?
  I still cannot say right away which parts of the inverse procedure were linearized to obtain a solution. Obviously the Fréchet derivative by using the linear transfer functions is the essential one. But as in other equations some terms are neglected as well, this seems like further linearizations. Maybe a small wrap up somewhere in the manuscript on how many different steps of linearization were used might be helpful.

Minor issues

- 190.24 Without wanting to go into too much mathematical detail: "the inverse is a matrix inverse". Although I think that this matrix inverse exists in most cases, it probably does not always exist. This would then require a pseudo-inverse and thus considerably increase the calculation time for the interation. Did you spend any thoughts on this?
- As in the short title of the screen version, it would be appropriate to add "with synthetic data" also in the title. This implies directly from the beginning that we basically know what the solution should look like. Moreover, the running title of the print version uses "ice stream", but the title uses "glaciers". In the text is is stated (183.27) "We focus . . . on active ice streams." A comparable issue has been raised by Andreas Vieli in the Discussion of an earlier paper (TCD,2, S269, 2008). So I ask the authors to decide if they find their method applicable to all ice masses or just a subset thereof, explain why, and then be consistent in the terminology used.
- The authors should mention more explicitly in the abstract and introduction that

TCD

3, S80–S91, 2009

Interactive Comment

they employ the inverse approach to two dimensions (x, z).

- Partly I had a difficult time to separate "a" as in "a priori" from article usage as in "a prior estimate". It might be better to italicize (or something comparable) expressions like "a priori" and "a posterior" for improved readibility.
- The authors use different notations for differentials without defining any of these. Knowing their previous papers it is obvious that e.g. $\sigma_{ij,j}$ refers to $i, j = x, z$ and $_{,j}$ to the derivative i.r.t. j and that Einstein's summation convention applies, but then they also use ∂_t and ∂_x and $\partial/\partial b$ (Eq. (16)) etc. One sentence in the notation section 2 could clarify all this.
- At one point it would be nice to have the numbers for the system used, i.e. number of unkowns and knowns to judge how overdetermined the system actually is. This could be especially of advantage if later studies aim at comparing results with different inverse methods (maybe a glaciological inverse model intercomparison project?).
- Please be consistent with the usage of nonlinear and non-linear. In Gudmundsson & Raymond 2008 non-linear is used throughout.
- 182.15 The "problem" does not have this goal, rather the "geophysical inverse method" or "approach".
- 182.26 The acronym MAP is defined several times after this location, I think once is enough.
- 183.21 "The main focus[es] of this paper is . . . "
- 183.22 ". . . in situation**s**"
- 183.25 Provide order of magnitude in percentages for how strong "moderately" actually is.

- 183.21, 184.17-22: At these two locations the purpose/focus of the paper is introduced. Should be concentrated at a single location in the introduction.

- 184.11 "... that solve**s a** reduced ..."

- 184.24 "a forward numerical model": This has already been introduced above as the non-linear ... FE model, so to avoid confusion if yet another model is used I would suggest "with [a] **our** numerical forward model".

- 184.25 "finite amplitude effects": of what?

- 184.25 "of convergence": presently not clear to what the convergence refers to, as the iterative approach has not been introduced yet. (As an inverse approach can also involve a single calculation/step, there is no implicite meaning of convergence in this context unless several iterative steps are used.)

- 185.9 "bold face**d**"

- 186.3 "model [model]"

- 186.4 "model is **a** two-dimensional"

- 186.10 So far the vector g has been used as the forward function. Now the accerleration due to gravity is also introduced as a vector with components g_i (which are $\neq 0$ because of the tilted coordinate system. Please change to avoid confusion.

- 186.24 "This parameter ...". It should be noted somewhere for which experiments this parameter has been used, e.g. in Table 1.

- 186.27 Provide the value respective expression for the rate factor.

- 187.2 Add that m only takes the values 1 and 3, depending on experiment.

TCD

3, S80–S91, 2009

Interactive Comment

- 187.15 "toward**s**"
- 188.17f With perturbations you actually mean $\Delta(s, u, w, b, c)$? Clarify in this sentence.
- 188.34 "**pdf[']s**" (plural, no apostroph)
- 188.24 Do you actually need the braces $(\ldots)$ in $P(m|(\ldots))$? This would remove superfluous characters in several equations and expressions in the text.
- 189.3 "[with] **and** the data"
- 190.6 "…Eqs. (8) and (9) into Eq. (7) **and omitting** $P(d)$ we obtain …"
- 190.12 Can you add a reference with (page numbers in case of a book) to MAP here?
- 190.15 "To find the [minimizers] **minimum** of …derivative **in respect to** m of …"
- Eq. (11) Somehow I am missing the contribution of the second part of the derivative i.r.t. $\mathbf{m}$ of both product terms when trying to carry out the differentiation with the product rule. What happened to them?
- 190.xx Both, $K^T(m)$ and $K(m)^T$ are used in the manuscript. Please unify for consistency.
- 190.22 "Newton's method **for systems of equations** via the iteration"
- 191.1 "**The term** $\nabla_m \Phi(m)$ …"
- 191.3 "whose resulting product is small": the product of which terms? Not clear which product results. In the denominator of (14) the term $-\nabla_m \mathbf{K}^T \mathbf{C}_D^{-1}(d-g)$ has been neglected, is this the product you referred to? Is this part of a further linearization?

- 191.11 Equation (16): Can you explain the derivatives $\partial g(m)_s/\partial b$ etc.? What do these subscripts refer to?

- 192.1 Please superscript H and T, as they could otherwise be mistaken for variables, e.g. as the ice thickness as used in tables (see below).

- 192.15 "is the covariance of **surface data** s_i and s_j **(as e.g. available from a number of repetitive measurements)**"

- 193.5 Can you provide a further argument or reference why "the correlation off-diagonal elements" should follow a Gaussian distribution"?

- 193.15 "stated above": provide reference to where exactly.

- 193.20 "pdf[']s"

- 193.21 "[minus] **negative** logarithm"

- 194.9 "or **a** maximum number of iternation**s** has"

- 194.10 Add a few words that this is related to the number of degrees of freedom.

- 196.14 "Data [was] **were** interpolated".

- 196.18 Could you describe in a few words what these (final) covariance matrices look like?

- section 5.2. A section heading like "Non-dimensionalisation" seems more appropriate.

- 197.14 "[or] **(** 5% . . . values **)** [,]"

- 197.15 "with **a** standard deviation"

TCD

3, S80–S91, 2009

Interactive Comment

- 197.16 "degrees" change to symbol °
- 197.18 "The **non-dimensionalized** surface data"
- 197.22 "[Surface] **Synthetic** measurements"
- 198.3 Are the correlation lengths the same in all experiments?
- 198.20 Unclear which standard deviation is meant, as the term "std.dev" is used several times above to describe the shape of the disturbances. According to the caption of Fig. 5, this std.dev. is taken from the (BLUE) covariance matrix and thus also includes some kind or interpolation error. Then it is far more than the initial std.dev. introduced in section 4.1. Please clarify.
- 198.22 "[by] **in** the first guess"
- 199.15 "in **the** next iteration step"
- 199.21 and 199.25: Somewhat repetitive (e.g. $n = 3$). Please shorten.
- 200.18 How (where) was this forcing implemented?
- 200.18, 201.18 It should be noted in Table 1 in an additional column or as a footnote for these two experiments that additional forcing was required to limit overshooting and allow for convergence.
- 200.24 "is in [a] the form"
- 200.25 "toward**s**"
- 201.14 "other[s] parameters are"
- 203.4 "are always **only** known"

TCD

3, S80–S91, 2009

Interactive Comment

- 203.10/11 "likely . . . likely": rewrite.
- 203.12/13 "We do not attempt here . . . We limit the discussion here": rewrite.
- 204.5 "a posteriori solution [] at the minimum **of** $J(\hat{m})$"
- 204 first paragraph: Is this finding (that one set of initial parameters yields a better solution than another one) also obvious from the related a posteriori error estimates of the solution?
- 204.16 "**emphasize**"
- Tables: Why is ice thickness given as H and not h?
- Table 1: It would also be helpful to state in Table 1 that all synthetic experiments consider noise-degraded data and that the rate factor is constant.
- Fig.3: Why is "z" on the y-axis? It is not depth. Caption: "velocities are [normed] **normalized** with"
- Fig.4: "The a priori was set to zero": The a priori what? Something is missing.
- Fig.5: "Note that the first and second": There is not first iteration shown?

Interactive comment on The Cryosphere Discuss., 3, 181, 2009.