Thank you for the opportunity to review this paper. The authors address an important problem in applying 4DEnVar to parameter calibration in complex land surface models, and demonstrate that the proposed regularisation can prevent implausibly large and/or unphysical parameter updates and improve the subsequent model simulations in both JULES and ORCHIDEE.

I have two main questions about the interpretation of the gamma hyperparameter. The mathematics here is close to the edge of my own comfort zone, so I would genuinely welcome clarification if I have misunderstood things.

We would like to thank the reviewer for their insightful comments, which substantially strengthen the article as well as our understanding of the method itself. We reply to each point individually below.

1. Relationship between gamma and the observation error covariance R

As I understand Eq. 9, increasing gamma increases the relative weight given to staying close to the prior. However, if the entire cost function is divided by gamma, the location of the minimum is unchanged, and the resulting expression appears to be exactly equivalent to standard 4DEnVar with gamma equal to 1 but with R multiplied by gamma.

If I have followed this correctly, the final analysis parameter vector obtained using gamma and R should therefore be identical to that obtained by leaving gamma at 1 and inflating R by the same factor. The authors note this relationship in the Discussion, but I wonder whether its implications deserve somewhat more attention.

For example, with the prescribed error standard deviation of 0.35 gC m-2 d-1, gamma values of 60 and 100 correspond, for determining the analysis mean, to effective standard deviations of about 2.7 and 3.5 gC m-2 d-1, respectively.

These are very large changes in the relative weighting of the prior and observational terms. I therefore wonder whether the experiments as presented can distinguish between gamma acting primarily to restrict the update to a region where the local linear approximation remains valid, and gamma compensating for the weighting implied by the prescribed R. These seem to have somewhat different implications for how a future user should choose R and gamma.

Could the authors discuss this distinction more explicitly? The simple nonlinear example earlier in the manuscript might perhaps provide a useful controlled setting in which to demonstrate specifically how regularisation addresses the nonlinear or linearisation problem when R is known.

We thank the reviewer for this comment, which is correct. We thus increase the discussion section to discussed all the very important point the reviewer rase:

> **It should be noted that γ and R are not independent. Dividing Eq. (9) by γ does not change the location of its minimum and yields the original cost function (Eq. 3) with R replaced by γ R. The analysis obtained with a given γ is therefore identical to the one obtained with γ = 1 and an observation error covariance matrix inflated by a factor γ. For the diagonal R-matrix used here, the optimal values γ = 60 (JULES) and γ = 100 (ORCHIDEE) correspond to effective observation error standard deviations of ≈ 2.7 and 3.5 gC m−2 d−1, respectively, instead of the prescribed 0.35 gC m−2 d−1. This equivalence allows two interpretations of γ, which cannot be distinguished within a single linearised analysis step such as the one used in this study.**
>
> **In the first interpretation, γ restricts the size of the update so that the analysis remains in the region where the linear approximation of the model is valid. Indeed, Eq. (10) is identical to the first step of a Levenberg–Marquardt minimisation starting from the prior, with a damping parameter λ = γ − 1 (Levenberg, 1944; Marquardt, 1963). In an iterative scheme, where the ensemble is regenerated around successive estimates as in Bocquet and Sakov, (2012,2013) , this damping would only control the step size, and the converged solution would remain the optimum for the prescribed B and R. However, regenerating the ensemble at each iteration requires additional model runs, which can be prohibitive for computationally expensive models, and prevents a single ensemble from being reused for multiple assimilation experiments. Moreover, the 2D example of Sect. 2.2.2 illustrates that the linear approximation can fail even when R is correctly specified, so that regularisation may be needed independently of any misspecification of R.**
>
> **In the second interpretation, γ compensates for an underestimation of R, which incorporates both observation and model errors that are difficult to quantify. While observation errors can, in principle, be prescribed, obtaining a reliable quantification of model errors remains a complex and challenging task. Our results suggest that this second mechanism contributes substantially in our experiments: the lowest posterior RMSEs (1.15 and 1.12 gC m−2 d−1 for JULES and ORCHIDEE, respectively) remain more than three times larger than the prescribed error standard deviation, and the diagonal R-matrix neglects both temporal error correlations in the monthly means and model structural errors.**
>
> **[...]**
>
> **In practice, we therefore recommend first specifying R as realistically as possible, including model and representation errors, and then using the γ**

as a safeguard against failures of the linear approximation. With a well-characterised R, the remaining need for regularisation would more directly reflect model nonlinearity.

2. Interpretation of the posterior uncertainty

My second question concerns Eq. 11 and the narrow posterior distributions shown in Fig. 3.

As a simple check, consider the limiting case in which the observations provide no information about the parameters. In that case, Eq. 11 appears to imply that the posterior perturbations are reduced by a factor of the square root of gamma, and therefore that the posterior covariance is reduced by a factor of gamma.

If this interpretation is correct, then with gamma values of 60 or 100 the posterior variance would be reduced by factors of 60 or 100 even in the absence of any observational information. More generally, it appears that the posterior covariance from Eq. 11 can never be larger than the original prior covariance divided by gamma.

This makes me uncertain about how to interpret the narrow posterior distributions in Fig. 3. A substantial reduction in spread relative to the stated prior appears to be imposed directly by the choice of gamma, irrespective of any additional constraint provided by the GPP observations.

I can see that Eq. 11 would have a consistent probabilistic interpretation if introducing gamma is equivalent to redefining the prior covariance as the original prior covariance divided by gamma. However, I had understood gamma in the manuscript primarily as an algorithmic regularisation parameter used to keep the update within the region where the local linear approximation is useful.

Could the authors therefore clarify what statistical quantity the posterior uncertainties in Fig. 3 represent? In particular, I am not sure that the narrowness of these distributions can straightforwardly be interpreted as evidence that the observations have tightly constrained the parameters unless the probabilistic role of gamma is made clearer.

We would like to thank the reviewer for spotting this. This is completely true, and we have modified the statistical interpretation of γ accordingly. As also discussed in our reply to the first point, we believe that the regularised solution (Eq. 10) is identical to the first step of a Levenberg–Marquardt minimisation with damping parameter $\lambda = \gamma - 1$. Following this interpretation, γ is an algorithmic device that limits the size of the update rather than a modification of the prior. We therefore now compute the posterior covariance from the Hessian of the original cost function (Eq. 5), which does not depend on γ, and have removed Eq. (11). We have also added the following text at L236:

Eq. (10) is also identical to the first step of a Levenberg–Marquardt minimisation starting from the prior, with a damping parameter λ = γ − 1 (Levenberg, 1944; Marquardt, 1963), as in Bocquet and Sakov (2012). In this interpretation, γ is an algorithmic device that limits the size of the update, and is not part of the statistical description of the prior or observation errors. The posterior error covariance is therefore computed from the Hessian of the original cost function (Eq. 3), i.e. using Eq. (5), and does not depend on γ.

In the revised Fig. 3, the posterior uncertainties are consequently identical for the γ = 1 and optimal-γ optimisations, and only the means differ. We have removed or modified the corresponding statements in Sect. 3.2 (L315):

Since the posterior error covariance (Eq. 5) does not depend on γ, the posterior uncertainties are identical in both optimisations, and only the posterior means differ.

and in the Discussion (L349):

In both cases, the regularised optimisation yields posterior means within physically plausible ranges, whereas several means obtained with γ fixed at 1 (i.e. the original 4DEnVar cost function) fall outside their prescribed bounds, indicating that a γ value different from 1 may improve parameter calibration exercises in land surface modelling.

We also added at the end of the Discussion paragraph revised for point 1 the following paragraph (L)

While the choice of interpretation of γ does not affect the analysis mean, it may affect the interpretation of the posterior uncertainty. The posterior error covariance used in this study (Eq. 5) follows the first interpretation and does not depend on γ. Under the second interpretation, however, this posterior error covariance would be underestimated, since it is computed with an observation–model error covariance matrix R that is itself underestimated. As both, model nonlinearity and the misspecification of R, are likely to contribute to the optimal values of γ found here, and as their effects are difficult to disentangle, the posterior uncertainties shown in Fig. 3 should be interpreted with caution.

Indeed, as the reviewer pointed out, we believe that γ is more of a practical choice that makes the use of 4DEnVar possible than a genuine statistical choice.

Overall, I found the paper interesting and the practical improvement from regularisation convincing. My questions are mainly about how that improvement should be interpreted statistically. A clearer discussion of the relationship between gamma and R, and

particularly of the meaning of the posterior uncertainties when gamma is greater than 1, would substantially strengthen the manuscript.

We would like to thank the reviewer again for their careful reading and for raising two important points, which we believe we have addressed in the revised manuscript and which have substantially strengthened it.