This is Appendix G from the Supplementary Material from Rousset and Ferdy (2014) (cross references were modified). Full Supplementary Material is available here.

Experiments with available R procedures

This section describes experiments with an implementation of the Matérn model suitable for use with R procedures such as `lme`, `glmmPQL` or `gls`. In particular, the `glmmPQL` function from the R package `MASS` can fit GLMMs with spatial correlations, and provides estimates based on penalized-quasi likelihood approximations (PQL), which we compared to estimates obtained by the methods implements in spaMM. The details are fastidious but illustrate problems commonly encountered in practice with existing procedures. We do not discuss here procedures that would depend on additional user input, such as specification of the number of iterations of a Markov chain or of prior distributions.

Input and output

Syntactically, `glmmPQL` does not fit models where the only random effect is spatial. It shares this with the `lme` function from the `nlme` package, on which `glmmPQL` is based. It has been suggested to add a spurious random effect with a single level in order to fit the model. Then, the call is of the form

```
glmmPQL(<fixed effect formula>,random=~1|dummy,
   correlation=corExp(form=~x+y|dummy),data=<data frame>)
```

where "`dummy`" is a spurious random effect with only one level for all the data analyzed. In a Gaussian LMM, the specification `corExp(form=~x+y|dummy)` means that there is an exponential correlation, with respect to the Euclidian distance between points with coordinates `x` and `y`, of the *residual* error among points within the same level of the conditioning dummy variable, that is, among all points in the present case. The "residual" variance reported by lme or glmmPQL is then the estimated variance of random effects, while the reported "intercept" variance is the variance of an uncorrelated residual error. In the Poisson or binomial GLMMs, the "residual" variance reported by glmmPQL is an overdispersion estimate for a quasipoisson of quasibinomial model.

The use of such a spurious random effect may look suspicious, but the fits it returns can be compared either to those obtained by analysing data that are duplicates of the original data, with two levels of the dummy variable, or to those obtained

by our ML or PQL procedures. The glmmPQL-based procedure has also been described (Dormann et al., 2007, Appendix) as reproducing the results of the GLIMMIX procedure in SAS (which was also based on quasi-likelihood methods; SAS, 2006). Further, for LMMs, `lme`, `glmmPQL` and the ML procedures should give consistent results as PQL estimation of fixed effects is equivalent to ML estimation in this case.

lme

Unexpectedly, lme systematically returned a practically null "intercept" variance in ML (as opposed to REML) fits. This lme result also differs from ML fits returned by our procedures, but the "residual" variance was correspondingly increased, so that sums of the two variances obtained by each ML implementation are similar. We then expect, and have checked, that this usage of lme gives results similar to the ML results on data simulated with low residual variance, namely the samples for the migration gene study design. We also implemented Matérn correlations within this framework. However, a substantial fraction of samples could then not be fitted due to lack of convergence of the optimization procedure called within lme.

If there are no repeated observations in given geographic locations, the residual variance can be modelled as a so-called nugget effect in the correlation function. However, we found that the optimization procedure almost always diverged when nugget estimation was attempted. Thus, one cannot analyze the simulated gaussian data sets from Table **??** under the model they where simulated from without additional programming. Moreover, this solution would not be appropriate for data sets with repeated observations in the same geographic location, because in this case the nugget and the residual variance are separable parameters.

glmmPQL

For GLMMs the PQL and ML results could be different, depending on the magnitude of random effects and the sample size. However, by comparing ML fits, glmmPQL fits and PQL/L fits, we found that glmmPQL performed much worse than PQL/L, which is very close to ML. for samples for which the glmmPQL fits converged, and for a low variance of random effects, estimates of fixed effects were close by both approaches and the `glmmPQL`-based analyses provided a good test of fixed effects. For a higher variance, glmmPQL fixed effect estimates depart from ML and PQL/L estimates, and the distribution of p-values was clearly

distorted.

Estimation performance is more easily compared on larger samples, and we turned to samples simulated under the design of the onchocerciasis prevalence study for such a comparison. In this case, all glmmPQL fits converged. ML and PQL/L fits are practically identical, with >0.9999 correlation of estimates of the different parameters, while PQL/L and glmmPQL differ as shown in Fig. 1. The performance of the tests is shown in Rousset and Ferdy (2014, Fig. 3, right). We checked that these discrepancies was not due to our implementation of the Matérn correlation for use with glmmPQL, by performing the same comparisons under a exponential correlation model (all other parameters unchanged), with similar results (Fig. 2).

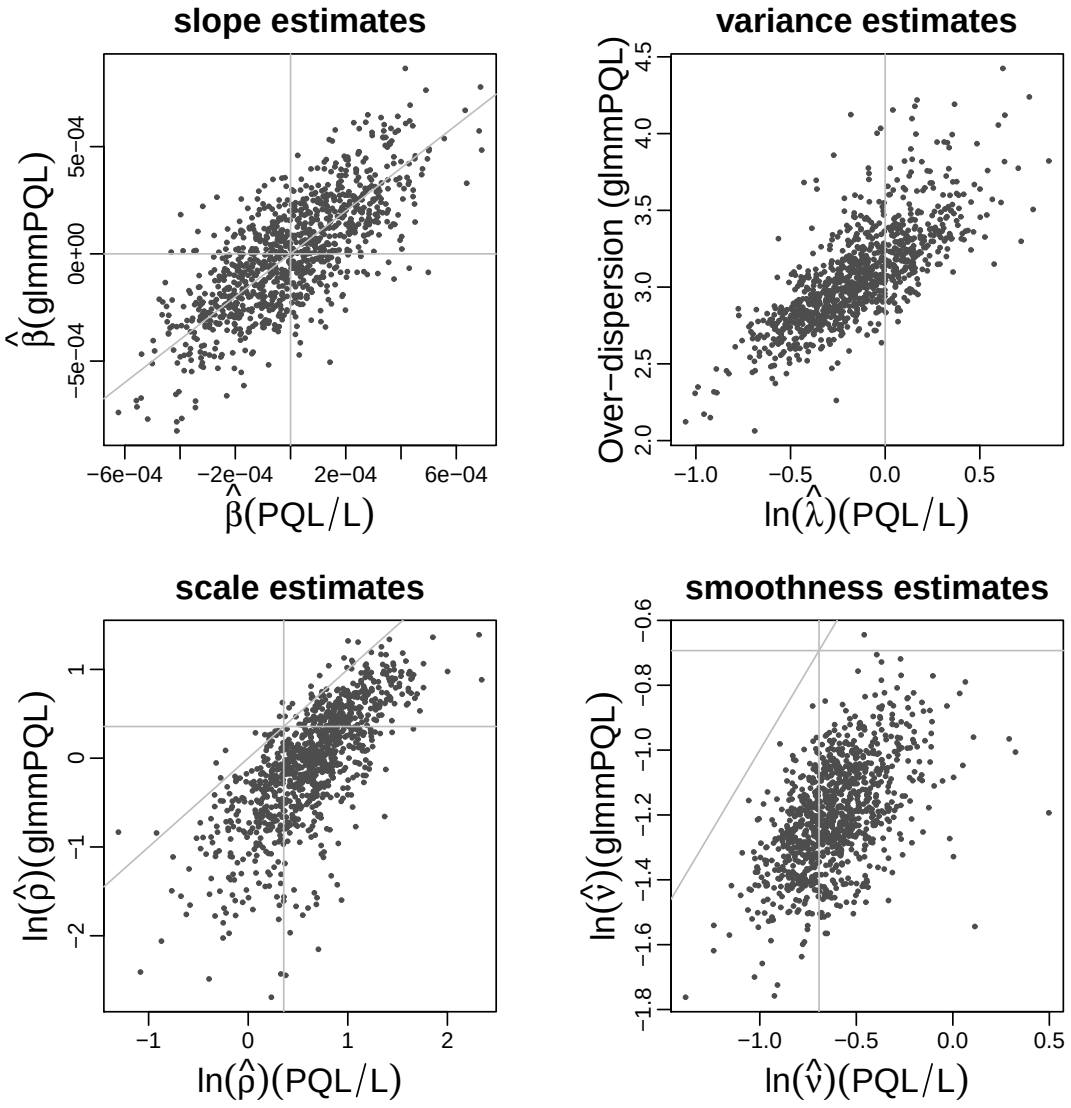

Figure 1: Joint distributions of parameter estimates by PQL/L and glmmPQL methods for data simulated according to the onchocerciasis study design.
The diagonal line is the 1:1 line, and the horizontal and vertical gray lines mark the position of the true value, except for over-dispersion in the glmmPQL analysis, where true value is hard to defined in all but the simplest designs.

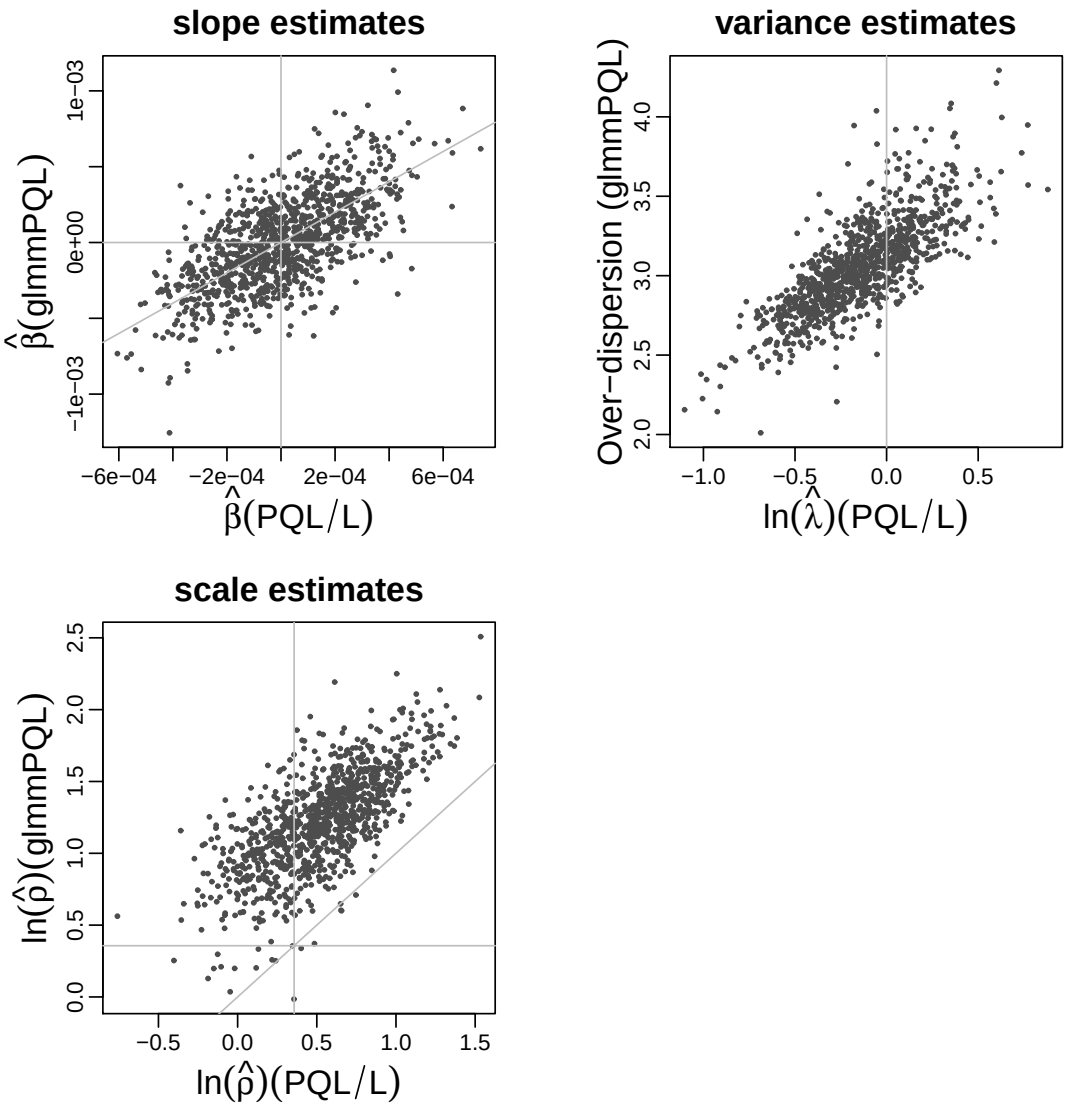

Figure 2: Same as Fig. 1 but for samples simulated an analyzed under an exponential correlation model

References

Dormann, C. F., J. M. McPherson, M. B. Araújo, R. Bivand, J. Bolliger, G. Carl, R. G. Davies, A. Hirzel, W. Jetz, W. D. Kissling, I. Kühn, R. Ohlemüller, P. R. Peres-Neto, B. Reineking, B. Schröder, F. M. Schurr, and R. Wilson. 2007. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. Ecography **30**:609–628.

Rousset, F., and J.-B. Ferdy. 2014. Testing environmental and genetic effects in the presence of spatial autocorrelation. Ecography **37**:781–790.

SAS. 2006. The GLIMMIX procedure, June 2006. SAS Institute Inc, Cary, NC.