# Supplementary Materials for "Gaussian process modeling of heterogeneity and discontinuities using Voronoi tessellations"

October 25, 2019

## 1 FURTHER DETAILS: Cloud modelling

In this section, we include further details of the emulator models fitted to the System for Atmospheric Modelling cloud-resolving model (Khairoutdinov and Randall, 2003; Feingold et al., 2016) as detailed in Section 6 of the article. After building an emulator based on 105 training points, we found that having two distinct regions was most likely. In particular, there was a large region associated with high values for the output proportion and a much smaller region associated with low output proportions. The MAP estimates for the parameters of the Gaussian process models in each region are given in Table 1.

| Region | Diagonal of roughness matrix, $B$ | | | | | | $\hat{\beta}$ | $\hat{\sigma}^2$ |
|---|---|---|---|---|---|---|---|---|
| Large region | 2.38986 | 0.66088 | 2.94809 | 2.07238 | 4.19353 | 31.77767 | 0.83469 | 0.05940 |
| Small region | 1.21890 | 0.50040 | 0.85465 | 1.08899 | 1.61007 | 20.84341 | 0.32239 | 0.00469 |

Table 1: MAP parameters for the two regions

To test the predictive ability of our initial model, we created 35 additional validation points using a Latin hypercube design as recommended in Bastos and O'Hagan (2009). The complete input configurations for the 35 validation points are given in Table 2. Given the size of the larger region, 33 of the 35 points fell within its boundaries (with respect to the MAP-estimate partition). Figure 1 shows the performance of the emulator at the 35 points. Generally speaking, the level of uncertainty and agreement with the true simulator output is acceptable for the 33 points with relatively high output proportions. For the two points that correspond to lower proportions of cloud cover in the model, the performance of the emulator is much worse: in one of the cases, the point is in an uncertain region of the input space where it is not known which region the point belongs and, in the other case, the point is misclassified as being in the region corresponding to larger proportions. Despite the poor fit for the lower proportions, our method performed better at predicting these validation points (MSE = 0.016) than the Treed GP (MSE = 0.032) and a standard Gaussian process model with no partitioning (MSE = 0.025). However, the issues with poor prediction for the location of one region motivated the use of a sampling regime to better locate the boundaries of the regions (as is detailed in Section 4 of the paper).

The second model fit as described in Section 6 of the paper gives much more certainty to the location of the partition between the two output regimes. This is achieved by targeting the location using the design algorithm in Section 4 of the paper. The resulting validation of the model show a clearer distinction between the two regimes and less uncertainty with regards to the low output proportion region because many more input configurations now fall within that region and the emulator is more confident in matching the model behaviour there.

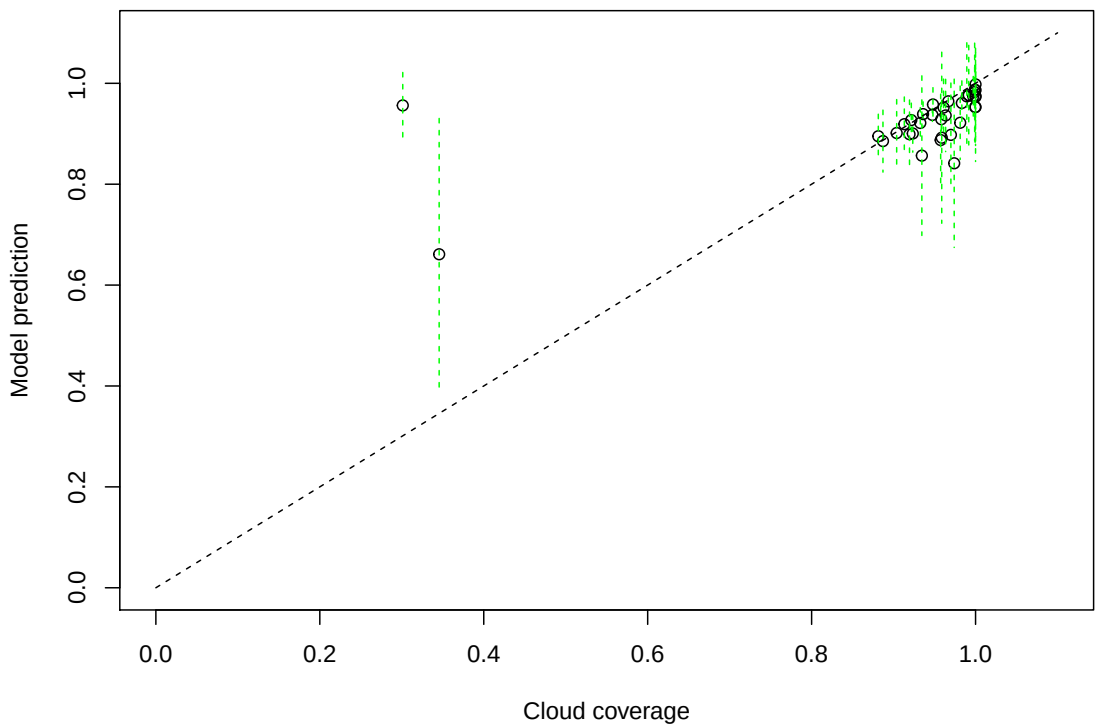


Figure 1: The predicted values (posterior means and two standard deviation ranges) plotted against the true value of the cloud fraction (those values gained from running the simulator).

| Index | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ |
|---|---|---|---|---|---|---|
| 1 | 286.120331 | 6.896842 | 1105.825684 | 6.766926 | 8.837931 | 123.511589 |
| 2 | 290.900452 | 8.704780 | 1238.534546 | 6.613815 | 6.059642 | 322.657379 |
| 3 | 289.877350 | 9.916729 | 881.390198 | 8.815986 | 6.335011 | 248.487091 |
| 4 | 287.186310 | 7.313122 | 1077.944824 | 8.986353 | 7.164250 | 201.089767 |
| 5 | 293.596558 | 10.422102 | 1184.299805 | 6.934781 | 9.394021 | 141.447266 |
| 6 | 287.653473 | 8.947084 | 688.814148 | 8.371425 | 6.810549 | 307.157623 |
| 7 | 284.266571 | 7.176429 | 697.652771 | 6.086229 | 6.856291 | 55.463993 |
| 8 | 290.681305 | 8.823312 | 1087.004517 | 6.375655 | 7.325428 | 77.907173 |
| 9 | 287.300934 | 9.377343 | 536.270142 | 6.902112 | 8.835855 | 258.708252 |
| 10 | 288.070679 | 9.976534 | 674.181885 | 6.489483 | 8.629810 | 187.722443 |
| 11 | 291.965179 | 10.233494 | 1133.300537 | 8.602721 | 6.487415 | 486.610962 |
| 12 | 285.561401 | 8.396331 | 1219.780762 | 9.970832 | 8.346141 | 119.727188 |
| 13 | 286.975067 | 7.441350 | 1188.145142 | 9.653389 | 7.268930 | 169.915665 |
| 14 | 288.093445 | 9.040697 | 1274.174561 | 8.731211 | 8.061809 | 288.946320 |
| 15 | 288.7691 | 8.878672 | 1192.7964 | 7.269598 | 8.892009 | 436.92209 |
| 16 | 291.1294 | 9.890397 | 1264.6671 | 8.542641 | 6.535751 | 480.32889 |
| 17 | 287.4935 | 9.418394 | 1033.1639 | 9.741000 | 7.526431 | 342.75906 |
| 18 | 285.0430 | 7.662993 | 1084.1565 | 8.112438 | 6.009365 | 283.14258 |
| 19 | 289.3911 | 10.211799 | 862.2623 | 6.512305 | 8.251513 | 413.94525 |
| 20 | 285.6145 | 8.251071 | 786.8574 | 6.154116 | 9.888325 | 303.79385 |
| 21 | 291.9070 | 9.642790 | 919.7665 | 7.145650 | 9.038586 | 164.66283 |
| 22 | 286.0617 | 7.340646 | 753.2080 | 6.823638 | 6.784937 | 128.04608 |
| 23 | 289.9971 | 8.102586 | 1025.3121 | 7.023006 | 6.395765 | 174.56256 |
| 24 | 287.6086 | 8.363940 | 580.2039 | 9.023112 | 9.969320 | 149.89809 |
| 25 | 290.2240 | 9.315969 | 843.0035 | 8.340917 | 7.480579 | 496.92413 |
| 26 | 292.9937 | 9.230615 | 1204.1548 | 6.957496 | 9.338216 | 463.66156 |
| 27 | 292.6375 | 9.548503 | 1070.3263 | 7.384440 | 8.984981 | 348.31155 |
| 28 | 289.1604 | 8.669858 | 775.1762 | 9.355122 | 7.039664 | 192.46733 |
| 29 | 285.4333 | 8.124009 | 518.6815 | 8.439637 | 9.257391 | 30.75079 |
| 30 | 290.2913 | 9.477997 | 798.2083 | 7.982011 | 7.219952 | 368.44559 |
| 31 | 289.6282 | 8.169958 | 886.3463 | 9.127620 | 8.949938 | 447.49091 |
| 32 | 290.1033 | 9.062427 | 824.5475 | 7.611224 | 6.641779 | 224.59059 |
| 33 | 287.8524 | 8.065599 | 831.0386 | 9.320685 | 9.995889 | 233.12976 |
| 34 | 285.0476 | 7.393353 | 585.1801 | 7.915319 | 9.542266 | 474.21387 |
| 35 | 286.7267 | 7.080297 | 900.1011 | 6.234400 | 7.567740 | 445.32407 |

Table 2: The 35 input configurations for the validation (and subsequent improvement) of the emulator

## 2 FURTHER APPLICATION: USA ammonia levels data

We also apply our method to data on recorded ammonia ($NH_4$) levels at locations across the USA, obtained from obtained from the National Atmospheric Deposition Program (National Atmospheric Deposition Program, 2007), which can be seen in Figure 2. The $NH_4$ was measured at 250 locations in the USA, with the two points in the bottom right corresponding to the United States Virgin Islands and Puerto Rico. On plotting the data in Figure 2, we have found that there is a drastic change in the output for certain areas of the USA. In other locations, however, the output does not change as drastically, suggesting that we may have heterogeneity. As this is real observed data, which is observed with error as opposed to a deterministic computer output as in Section 6 of the main paper, the error term $\sigma_\epsilon^2$ is included in our model.

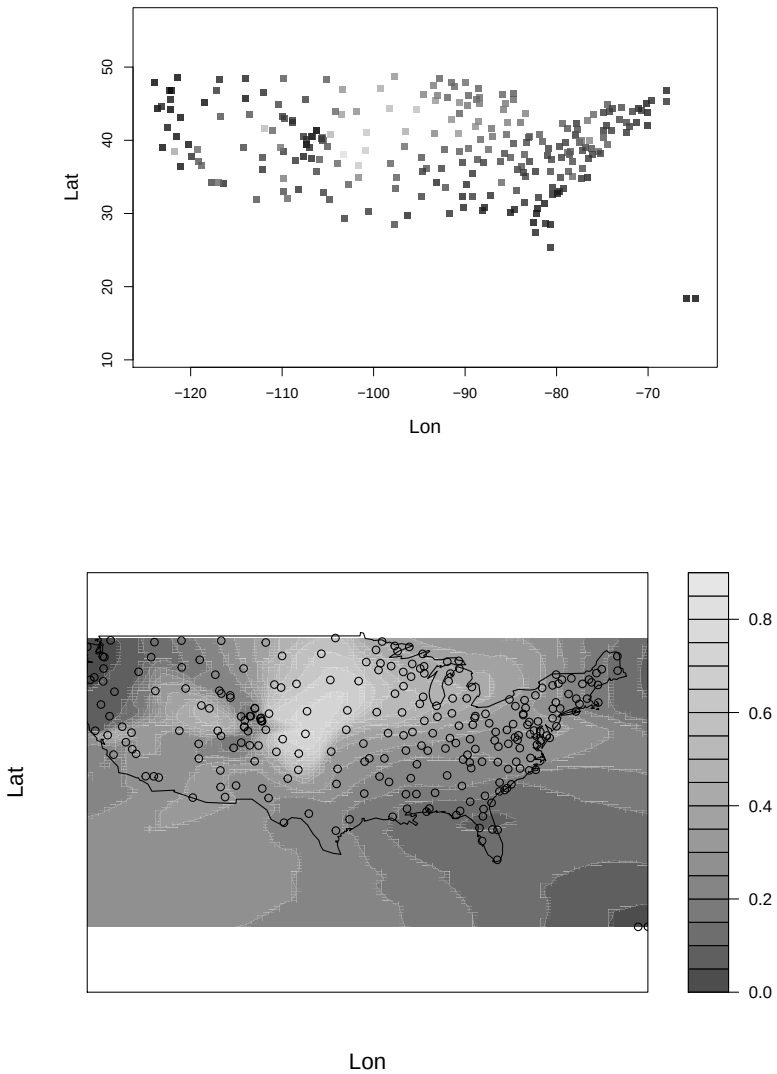


Figure 2: Analysis on the USA ammonia data. Top: The design points and output of the data; Bottom: The integrated surface from the application of our method.

The integrated surface that we obtain for this example, via application of our modelling approach,

is shown in Figure 2. This surface suggests that the north-central region of the USA has higher levels of ammonia compared to the rest of the country. We also notice that the north western region of the USA has much lower levels than the rest of the country.

Figure 3 shows our posterior distribution for the number of regions of different behaviour in $NH_4$ over the USA. We see that we have a bell shaped distribution that peaks at eight regions with an elongated tail towards the larger values, showing that there are most likely 8 different regimes over the spatial area. As with the previous examples in the paper, we test our method against the TGP and the standard Gaussian process modelling approaches. To do this here we use cross validation in which we randomly omit 50 training points (20% of the total data), and then use these as validation points on a model trained using only the remaining training points. Here, we again found that our method has a lower MSE (MSE = 0.0057) than both the standard Gaussian process (MSE = 0.0084) and the treed Gaussian process (MSE = 0.0059).

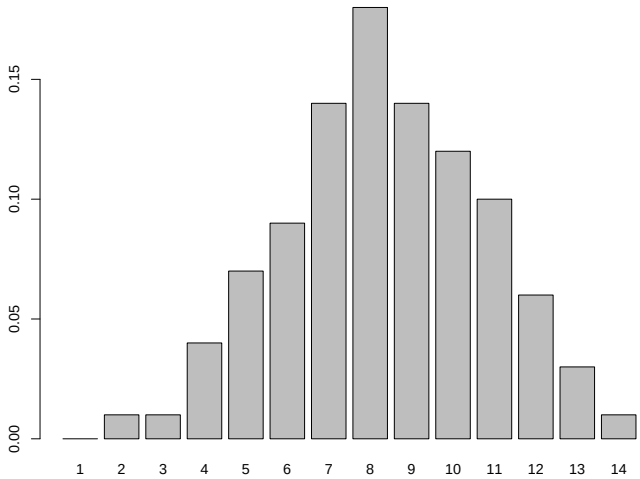


Figure 3: The posterior distribution for the number of regions for the USA ammonia level data.

# References

Bastos, L. S. and A. O'Hagan (2009). Diagnostics for Gaussian process emulators. *Technometrics 51*(4), 425–438.

Feingold, G., A. McComiskey, T. Yamaguchi, J. S. Johnson, K. S. Carslaw, and K. S. Schmidt (2016). New approaches to quantifying aerosol influence on the cloud radiative effect. *Proceedings of the National Academy of Sciences 113*(21), 5812–5819.

Khairoutdinov, M. F. and D. A. Randall (2003). Cloud resolving modeling of the ARM summer 1997 IOP: Model formulation, results, uncertainties, and sensitivities. *Journal of the Atmospheric Sciences 60*(4), 607–625.

National Atmospheric Deposition Program (2007). Ammonia monitering network. `http://nadp.sws.uiuc.edu/data/AMoN/`. Accessed: 2017-06-10.