Lab 6

Antonio Jurlina

10/29/2020

The data under consideration can be seen in Figure 3. At first glance, there seems to be low values gathered at the top half of given data space with higher content samples being reported in the bottom half. After fitting the linear and quadratic trend models, estimates (Table 3) indicate that these two models account for 28 and 36% of data variation, respectively, with the y axis being relevant in determining soil calcium concentrations. A look at the residuals in Figure 4 indicates that both linear and quadratic models seem to detect some correlation between the changes in the y axis and calcium concentrations. According to Figures 5 and 6, and Tables 5 and 6, the residuals from the linear and quadratic models are not normally distributed, given that both the tests reject the normality assumption.

An ANOVA comparing the linear and quadratic models shows that the addition of 3 more parameters to the quadratic (compared to the linear) model lead to a significantly improved fit.

According to Figure 1 and Figure 2, any two calcium samples taken from the given area will be related to each other up to a distance of 300 (linear) and 180 (quadratic). The relatively high nugget indicates the likely presence of measurement errors or not enough data in the sample. The sill for both models can be visually estimated around 40. Fitting variogram models to these two empirical ones gives us fairly similar results (Table 4). Between the two models, adding more parameters to create the quadratic fit somewhat reduced the linear dependence of sampled data.

Finally, creating a generalized least squares model to estimate the spatial relationship of calcium samples across testing locations seems to provide a better fit. First, Table 1 shows that the GLS model has the lowest AIC and BIC values, as well as the highest Log Likelihood. Furthermore, according to Figures 5 and 6, as well as Tables 5 and 6, the residuals from the GLS model seem to be normally distributed (at least we fail to reject the normality hypothesis). Finally, both the x and y variables lose their significance in this approach, indicating an unlikely correlation between calcium and the geographic location of the site (and leaving room for other potential explanatory variables to take that place).

Table 1: Diagnostic result comparison

model	Log Likelihood	AIC	BIC
linear trend	-651.3071	1310.614	1323.341
quadratic trend	-640.2147	1294.429	1316.702
generalized least squares	-636.3316	1284.663	1303.652

Table 2: ANOVA (linear vs. quadratic trend models)

res.df	rss	df	sumsq	statistic	p.value
175	15708.60	NA	NA	NA	NA
172	13867.87	3	1840.725	7.61003	8e-05

Table 3: Estimates across models

term	estimate	std.error	statistic	p.value	r_squared	model
(Intercept)	188.95976	22.07712	8.55908	0.00000	0.27713	linear trend
x	0.00382	0.00296	1.29285	0.19777	0.27713	linear trend
y	-0.03047	0.00374	-8.15549	0.00000	0.27713	linear trend
(Intercept)	2925.64287	665.49305	4.39620	0.00002	0.36183	quadratic trend
x	-0.08357	0.11760	-0.71060	0.47829	0.36183	quadratic trend
y	-0.98057	0.21498	-4.56129	0.00001	0.36183	quadratic trend
I(x^2)	-0.00003	0.00001	-2.18530	0.03022	0.36183	quadratic trend
I(y^2)	0.00005	0.00002	2.99813	0.00312	0.36183	quadratic trend
I(x * y)	0.00008	0.00002	4.13598	0.00006	0.36183	quadratic trend
(Intercept)	169.43804	650.53691	0.26046	0.79482	NA	generalized least squares
x	-0.00423	0.03252	-0.12993	0.89677	NA	generalized least squares
y	-0.01793	0.03372	-0.53167	0.59563	NA	generalized least squares

Table 4: Variogram model estimates

model	psill	range	kappa	ang1	ang2	ang3	anis1	anis2
linear trend	38.44135	0.0000	0.0	0	0	0	1	1
linear trend	52.39640	292.1801	0.5	0	0	0	1	1
quadratic trend	35.64462	0.0000	0.0	0	0	0	1	1
quadratic trend	40.71863	181.8001	0.5	0	0	0	1	1

Table 5: Jarque Bera test

model	statistic	p.value	parameter
linear trend	13.01702	0.00149	2
quadratic trend	9.04228	0.01088	2
generalized least squares	2.10610	0.34887	2

Table 6: Shapiro-Wilk normality test

model	statistic	p.value
linear trend	0.97692	0.00470
quadratic trend	0.98101	0.01576
generalized least squares	0.99150	0.37790

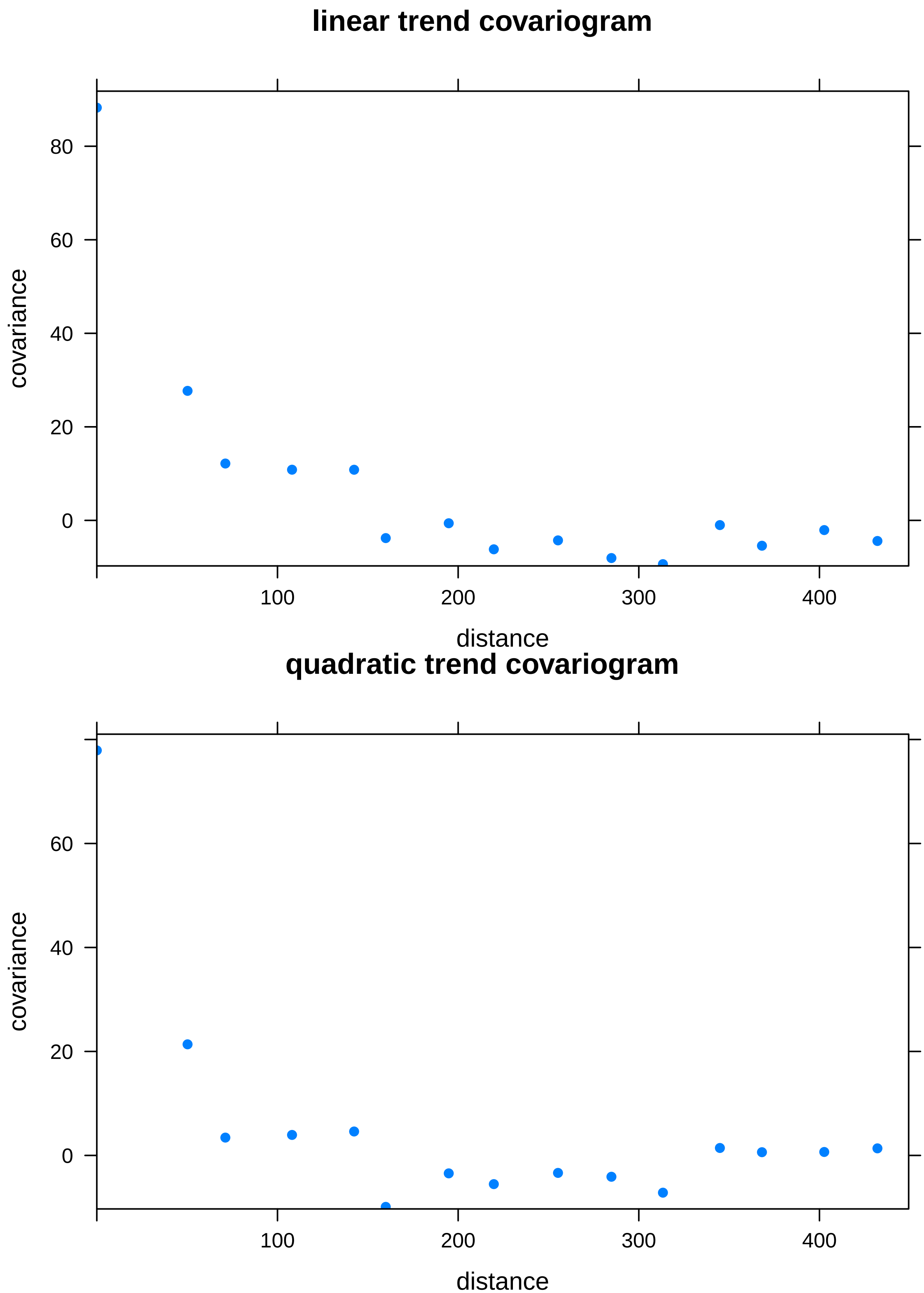

Figure 1: Covariograms

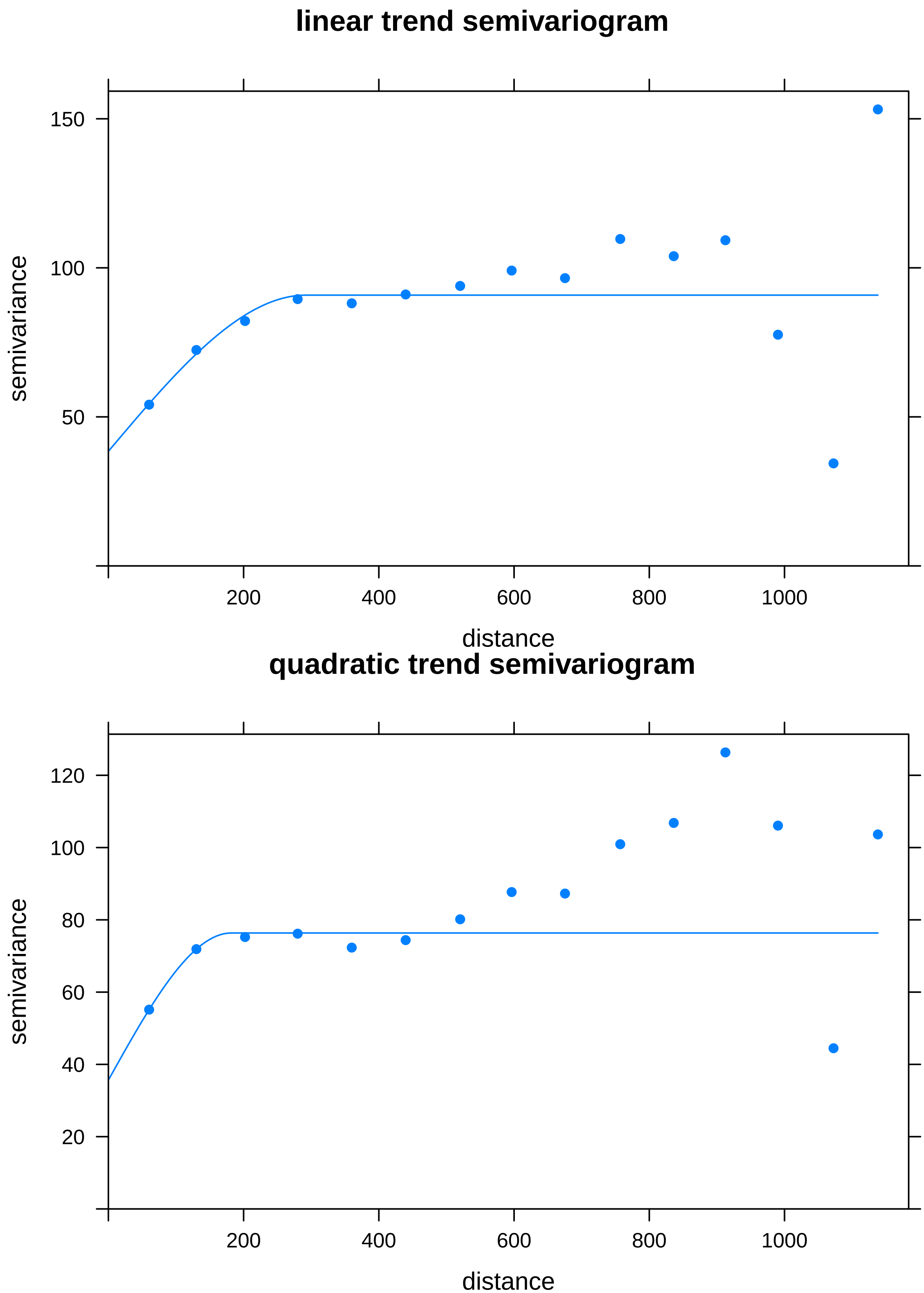

Figure 2: Variograms

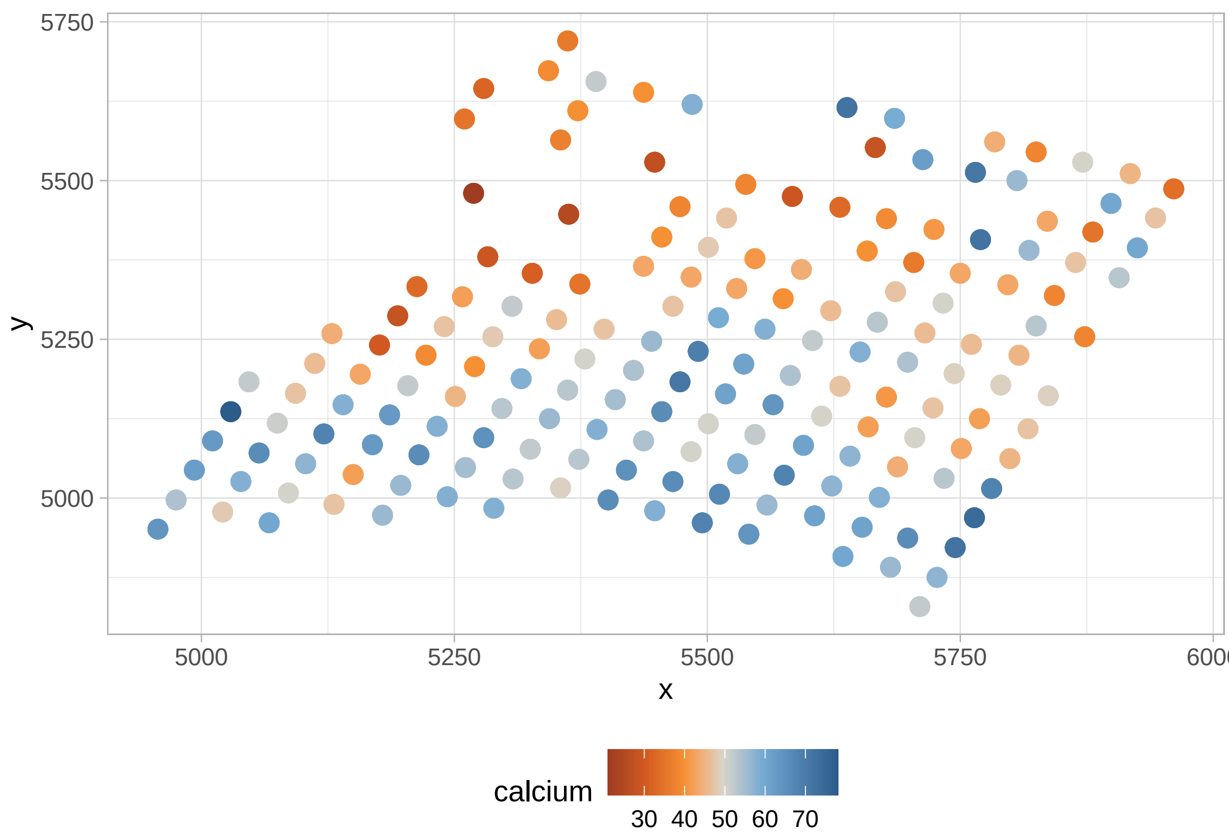

Figure 3: Calcium data

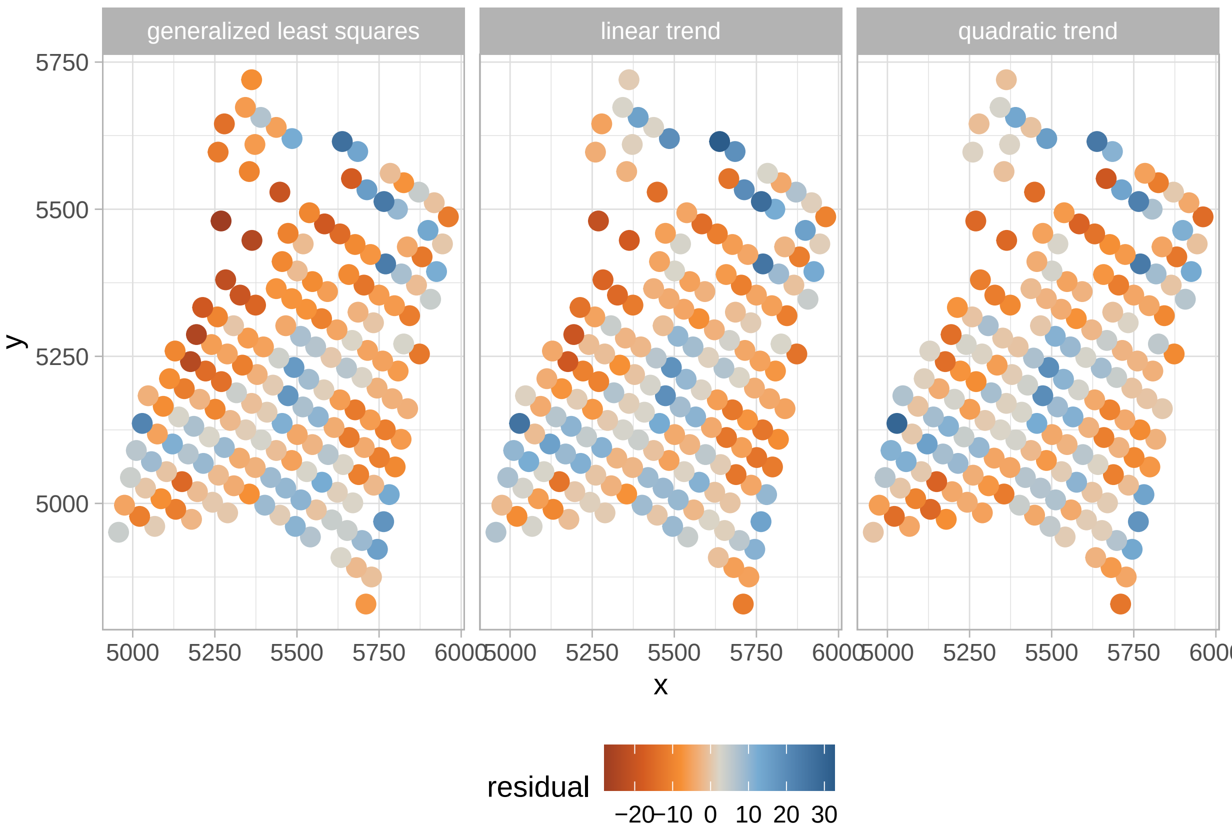

Figure 4: Model residuals

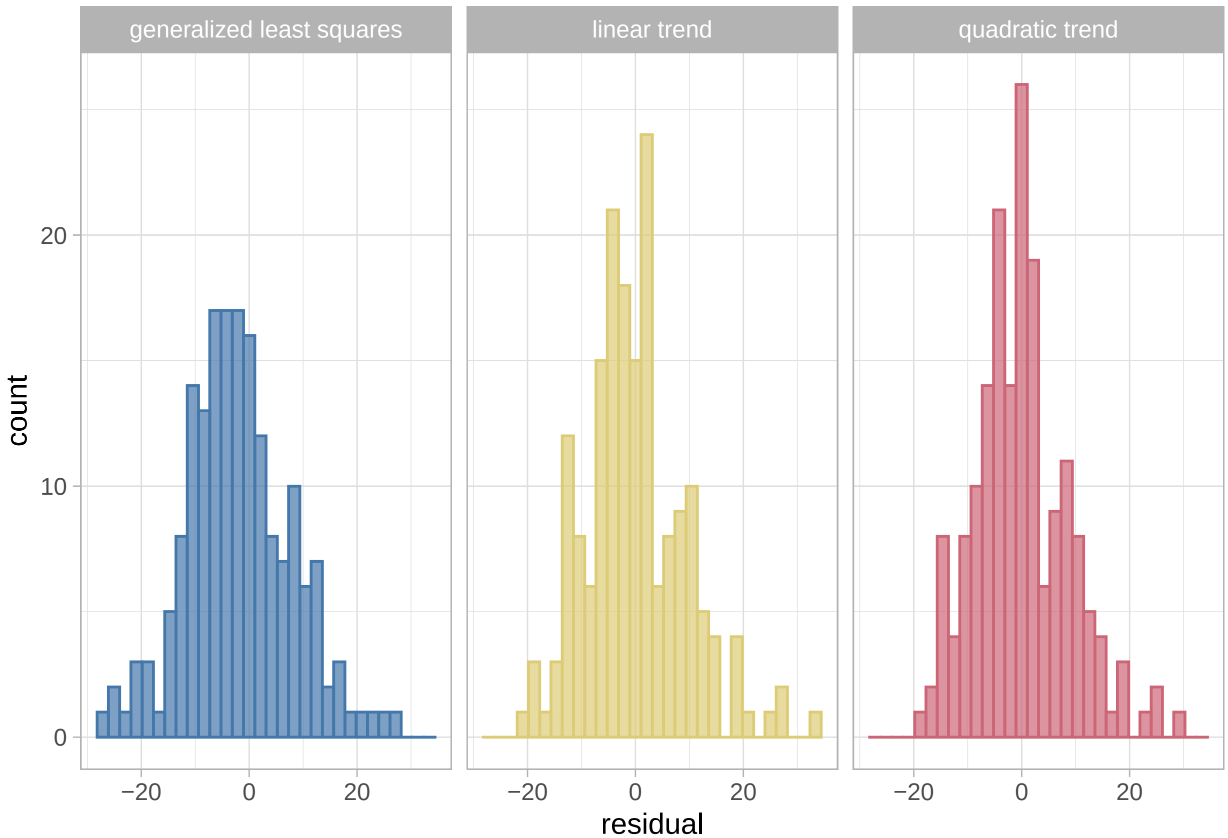

Figure 5: Distributions of model residuals

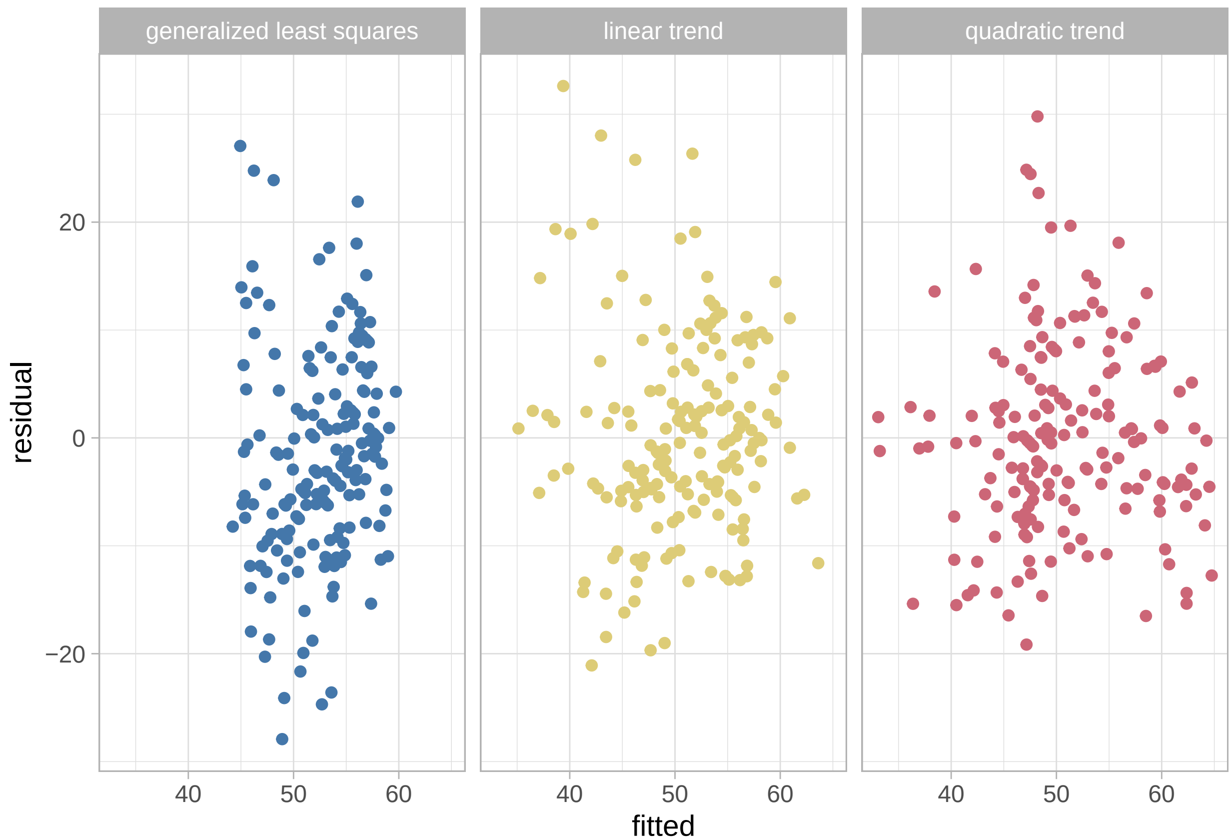

Figure 6: Residuals vs. fitted values across all models

```
## R version 3.6.2 (2019-12-12)
## Platform: x86_64-apple-darwin15.6.0 (64-bit)
## Running under: macOS Sierra 10.12.6
##
## Matrix products: default
## BLAS:   /Library/Frameworks/R.framework/Versions/3.6/Resources/lib/libRblas.0.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/3.6/Resources/lib/libRlapack.dylib
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
##  [1] knitr_1.29      tseries_0.10-47 nlme_3.1-149    broom_0.7.0
##  [5] gstat_2.0-6     ggthemes_4.2.0  rgdal_1.5-16    rebus_0.1-3
##  [9] sf_0.9-5        sp_1.4-2        forcats_0.5.0   stringr_1.4.0
## [13] dplyr_1.0.2     purrr_0.3.4     readr_1.3.1     tidyr_1.1.2
## [17] tibble_3.0.3    ggplot2_3.3.2   tidyverse_1.3.0
##
## loaded via a namespace (and not attached):
##  [1] fs_1.5.0             xts_0.12-0           lubridate_1.7.9
##  [4] httr_1.4.2           rprojroot_1.3-2      tools_3.6.2
##  [7] backports_1.1.9      R6_2.4.1             KernSmooth_2.23-17
## [10] DBI_1.1.0            rebus.base_0.0-3     colorspace_1.4-1
## [13] withr_2.2.0          tidyselect_1.1.0     curl_4.3
## [16] compiler_3.6.2       cli_2.0.2            rvest_0.3.6
## [19] xml2_1.3.2           labeling_0.3         scales_1.1.1
## [22] classInt_0.4-3       quadprog_1.5-8       digest_0.6.25
## [25] rebus.unicode_0.0-2  rmarkdown_2.3        pkgconfig_2.0.3
## [28] htmltools_0.5.0      dbplyr_1.4.4         highr_0.8
## [31] rlang_0.4.7          readxl_1.3.1         TTR_0.24.2
## [34] rstudioapi_0.11      quantmod_0.4.17      FNN_1.1.3
## [37] generics_0.0.2       farver_2.0.3         zoo_1.8-8
## [40] jsonlite_1.7.0       magrittr_1.5         Rcpp_1.0.5
## [43] munsell_0.5.0        fansi_0.4.1          lifecycle_0.2.0
## [46] stringi_1.4.6        yaml_2.2.1           grid_3.6.2
## [49] blob_1.2.1           crayon_1.3.4         rebus.datetimes_0.0-1
## [52] lattice_0.20-41      haven_2.3.1          hms_0.5.3
## [55] pillar_1.4.6         spacetime_1.2-3      rebus.numbers_0.0-1
## [58] reprex_0.3.0         glue_1.4.2           evaluate_0.14
## [61] modelr_0.1.8         vctrs_0.3.4          cellranger_1.1.0
## [64] gtable_0.3.0         assertthat_0.2.1     xfun_0.16
## [67] e1071_1.7-3          class_7.3-17         intervals_0.15.2
## [70] units_0.6-7          ellipsis_0.3.1       here_0.1
```