Parameters

Foeke Boersma

May 2025

S1 Global

S1.1 Ensemble Trees

S1.1.1 Random Forest

For the random forest model, the number of estimators is set to 1000; the `min_samples_split` equals 10; the `min_samples_leaf` equals 5.

```
rf = RandomForestRegressor(n_estimators=1000, random_state=42,
    min_samples_split=10, min_samples_leaf=5, max_features=4,
    max_depth=10, bootstrap=True)
```

S1.1.2 LightGBM

For both the LightGBM and XGBoost models, the number of estimators is set to 50,000; the `reg_alpha` equals 2; the `reg_lambda` equals 0; the max depth equals 5; the learning rate is 0.0005.

```
model = lgb.LGBMRegressor(reg_alpha=2, reg_lambda=0, max_depth=5,
    learning_rate=0.0005, n_estimators=50000)
```

S1.1.3 XGBoost

Relevant code:

```
xg_reg = xgb.XGBRegressor(gamma=5, reg_alpha=2, reg_lambda=0,
    max_depth=5, learning_rate=0.0005, n_estimators=10000,
    random_state=42)
```

S1.2 Lasso/Ridge

Formula for the linear model:

```
linear_urb = Lopend_gemiddelde ~ 1 + population_3000 + road_class_3_3000
+ trafbuf25 + population_1000 + nightlight_1000 + nightlight_3150
+ trafbuf50 + road_class_3_300 + bldden100 + ndvi + road_class_2_25
+ trop_mean_filt_2019
```

Extended with:

```
model_lasso = Lasso(alpha=0.1)
model_ridge = Ridge(alpha=0.1)
```

S2 Local

S2.1 Linear-Separating Spatial Groups

Related formulas:

```
linear_urb = lm(Lopend_gemiddelde ~ 1 + nightlight_450 + nightlight_4950 +
population_3000 + road_class_1_5000 + road_class_2_1000 +
road_class_2_5000 + road_class_3_100 + road_class_3_300 +
trafBuf50, data=Urban)

linear_lp = lm(Lopend_gemiddelde ~ 1 + nightlight_450 + nightlight_4950 +
population_3000 + road_class_1_5000 + road_class_2_1000 +
road_class_2_5000 + road_class_3_100 + road_class_3_300 +
trafBuf50, data=Lowpopulation)

linear_ffr = lm(Lopend_gemiddelde ~ 1 + nightlight_450 + nightlight_4950 +
population_3000 + road_class_1_5000 + road_class_2_1000 +
road_class_2_5000 + road_class_3_100 + road_class_3_300 +
trafBuf50, data=FarFromRoad)
```

S2.2 Linear Model

Related code:

```
model <- lm(Lopend_gemiddelde ~ 1 + nightlight_450 + nightlight_4950 +
population_3000 + road_class_1_5000 + road_class_2_1000 +
road_class_2_5000 + road_class_3_100 + road_class_3_300 +
trafBuf50, data = data)
```

S2.3 Mixed-Effects Model

Spatial character definitions:

```
grid100_sf$spachar = ifelse(grid100_sf$population_1000 > population1000_05 &
((grid100_sf$road_class_2_100 > 0 | grid100_sf$road_class_1_100 > 0) |
grid100_sf$road_class_3_100 > roadclass3_100_05), 1, 0)

grid100_sf$spachar = ifelse(grid100_sf$population_1000 < population1000_05 &
((grid100_sf$road_class_2_100 > 0 | grid100_sf$road_class_1_100 > 0) |
grid100_sf$road_class_3_100 < roadclass3_100_05), 2, grid100_sf$spachar)
```

```
grid100_sf$spachar = ifelse((grid100_sf$spachar == 1 | grid100_sf$spachar == 2),
grid100_sf$spachar, 3)

model <- lmer(Lopend_gemiddelde ~ 1 + nightlight_450 + nightlight_4950 +
population_3000 + road_class_1_5000 + road_class_2_1000 +
road_class_2_5000 + road_class_3_100 + road_class_3_300 +
trafBuf50 + (1|spachar), data = data_3035)
```

S2.4 Ordinary Kriging

Perform autofit variogram (for precipation data):

```
variogram_auto_lin = autofitVariogram(Lopend_gemiddelde ~ 1, data_sp)
plot(variogram_auto_lin)
autofit_params_lin <- variogram_auto_lin$var_model
print(autofit_params_lin)

m <- vgm(psill = 80.16775, "Ste", range = 10000)
```