GISC 6387 PROJECT: PREDICTING OIL WELL YIELDS IN RUSSIA

Rachel Brasier

3 August 2020

PROJECT PURPOSE

Predict the yield of oil wells in Russia based on their similarity to other wells with known production.

REVIEW OF LITERATURE

- Petroleum resource modeling
- Geology of oil fields
- Soviet infrastructure

REVIEW OF LITERATURE: USGS

- United States Geological Survey
- 2000 World Petroleum Assessment
  - Worldwide oil and natural gas production capacity
- Methodology [Charpentier and Klett (2005)]
  - Geological factors
  - Exploration history
    - Larger accumulations generally found earlier, but
    - Economic, technological, and political factors can also affect exploration trend

REVIEW OF LITERATURE: HAMIDA, ET AL. (2017)

- Explores variety of geometry-based optimization approaches for well placement in oil fields
- Measurement for quantitative similarity between wells
- Journal of Petroleum Science and Engineering

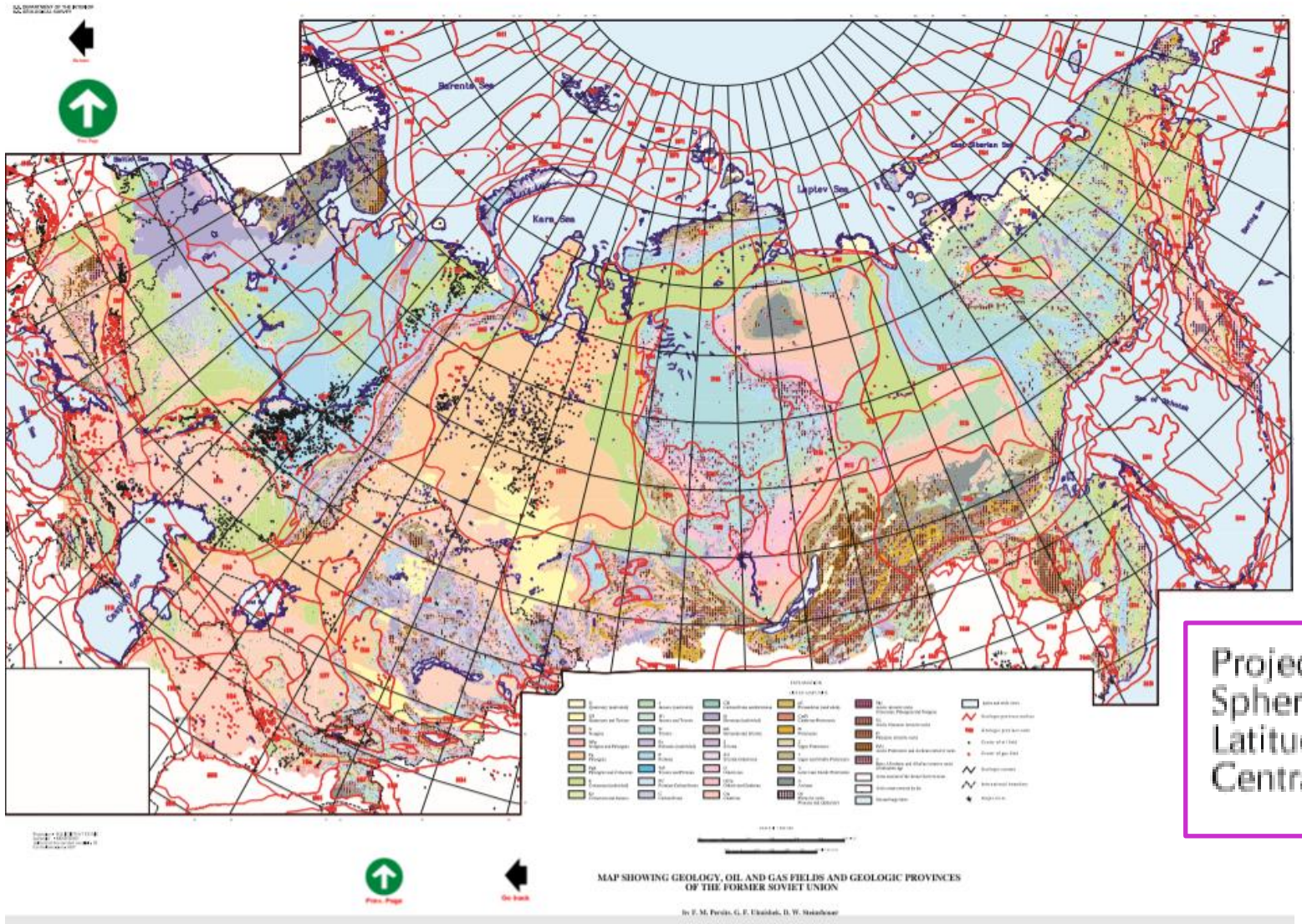
REVIEW OF LITERATURE: OTHER PAPERS

- Ivanov et al. (2018). *The main factors affecting the distribution of oil fields in the West Siberian Platform.*
  - Proximity to Triassic rifts
  - Thickness and geological age of the basement
- Campbell (1968). *Economic Reform in the USSR.*

STUDY AREA AND DATA SOURCES: USGS

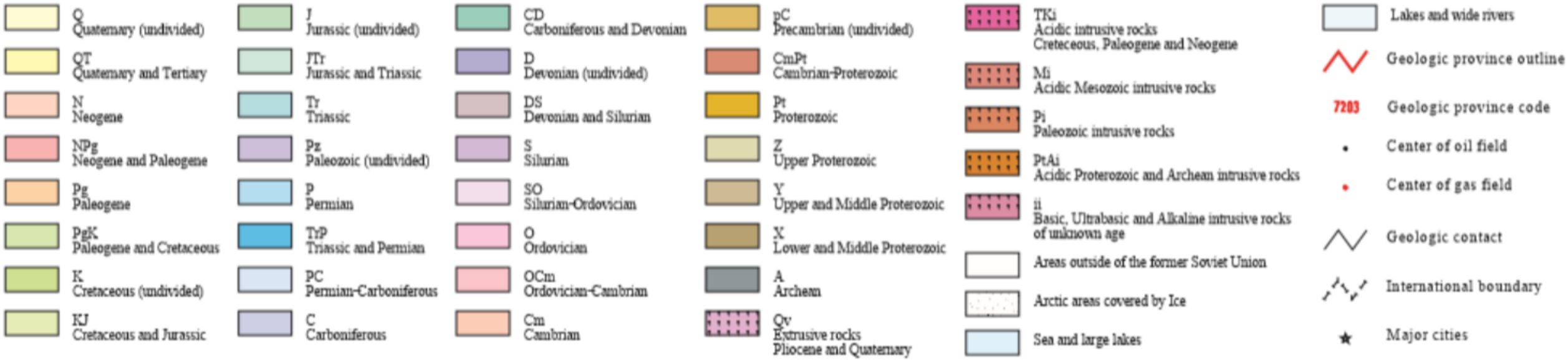
- USGS assessment of oil production capacity of oil fields in Former Soviet Union
- Map to a shapefile of geological provinces

STUDY AREA AND DATA SOURCES: USGS



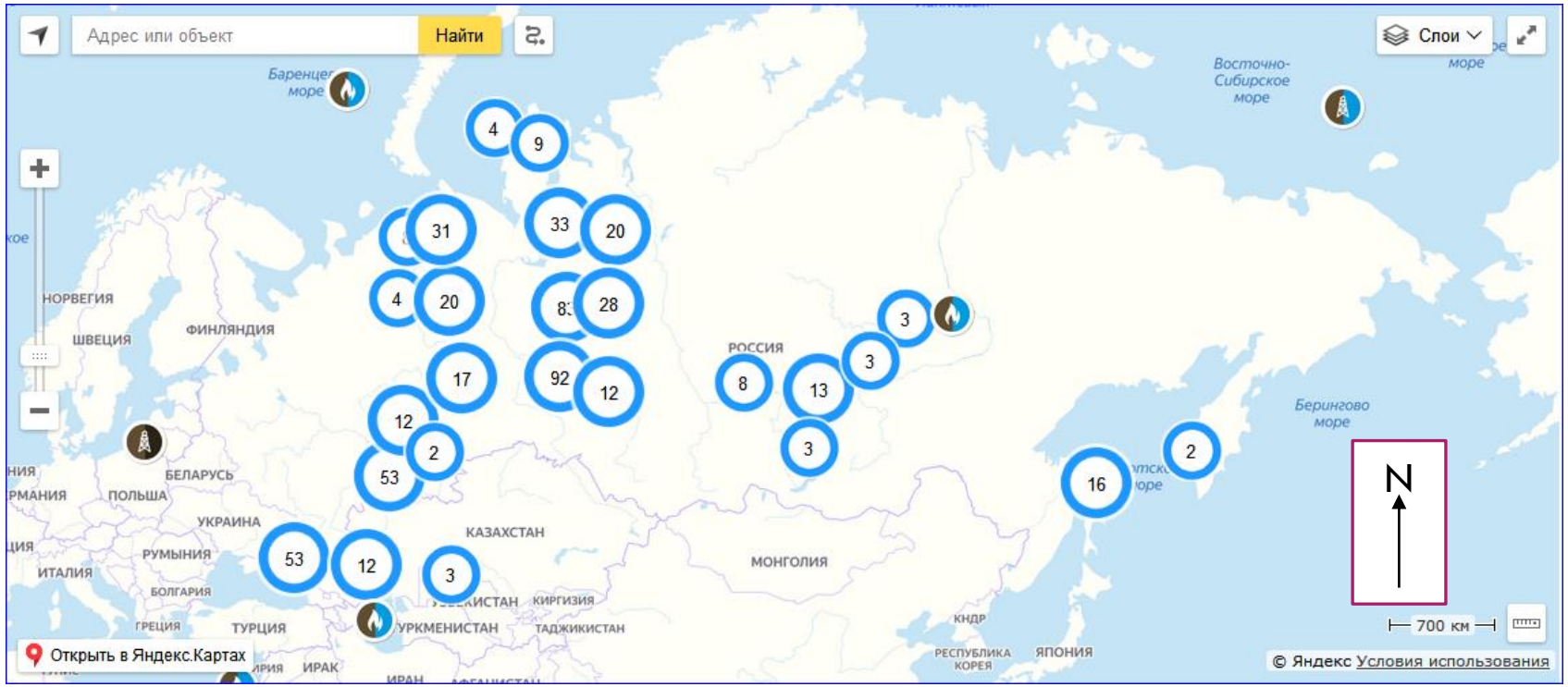
Projection - EQUIDISTANT CONIC
Spheroid - KRASOVSKY
Latitude of the standart parallel - 58
Central Meridian - 100°

STUDY AREA AND DATA SOURCES: USGS



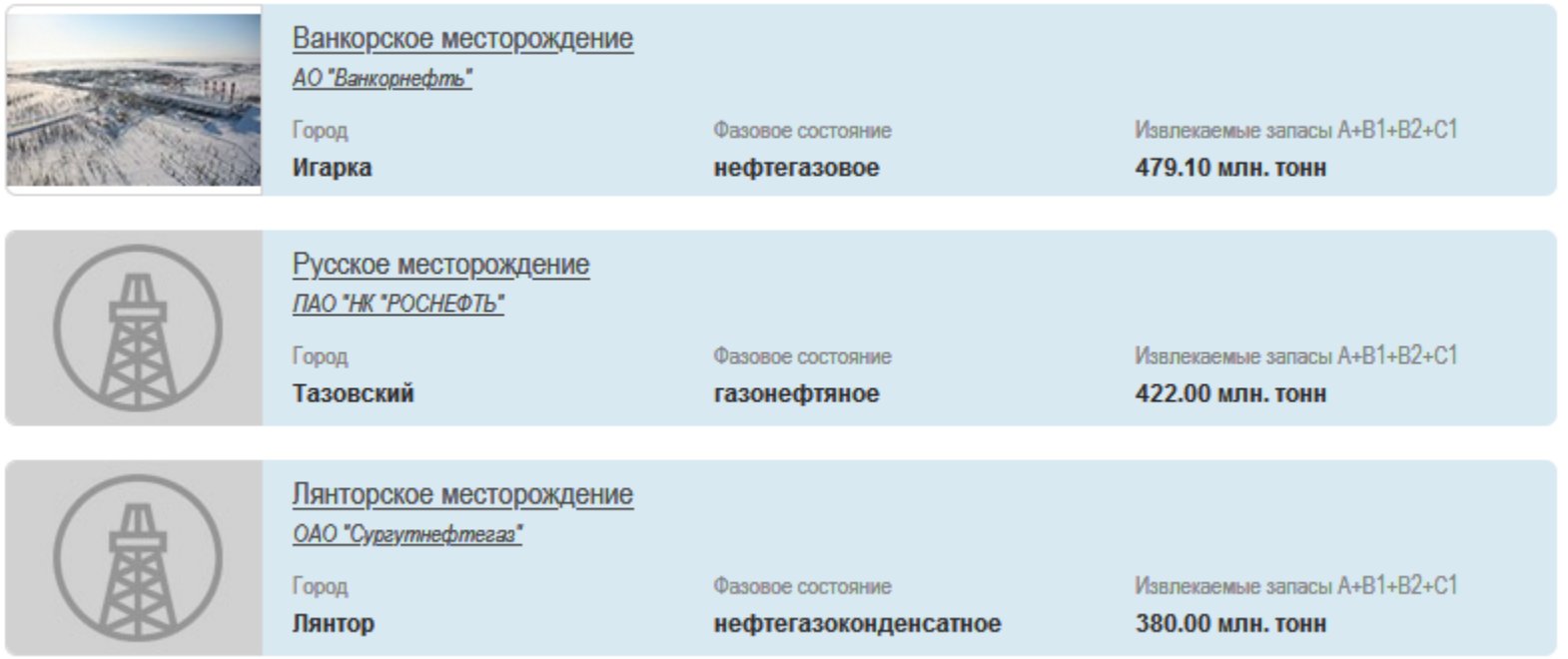
STUDY AREA AND DATA SOURCES: ENERGYBASE.RU

- Energybase.ru provides the location (city) of 558 currently operating oil and gas wells in Russia



STUDY AREA AND DATA SOURCES: ENERGYBASE.RU

- Oil/gas production data available for 98 wells
- Example:



WEB SCRAPING ENERGYBASE.RU

```
wellsdat <- NULL
for (pg in 1:28){
  webpage <- paste0("https://energybase.ru/oil-gas-field/index?page=", pg) %>%
    read_html(encoding = "utf-8")
  out <- webpage %>%
    html_nodes("div.name, div.info") %>%
    html_nodes("a, small a, small, div.value") %>%
    html_text()
  out <- out[grepl("\n", out) == FALSE]
  wells <- as.data.frame.vector(out)
  wellsdat <- rbind(wellsdat, wells)
}
write.csv(wellsdat, "wellsdat.csv", row.names = FALSE)
```

WEB SCRAPING ENERGYBASE.RU

```
sites <- NULL
for (pg in 1:28){
  webpage <- paste0("https://energybase.ru/oil-gas-field/index?page=", pg) %>%
    read_html(encoding = "utf-8")
  out <- webpage %>%
    html_nodes("div.name > a") %>%
    html_attr("href")
  site <- as.data.frame.vector(out)
  sites <- rbind(sites, site)
}
write.csv(sites, "sites.csv", row.names = TRUE)
```

WEB SCRAPING ENERGYBASE.RU

```
<div class="item">
    <small>
        Координаты:
    </small>
    <br>
    <a href="#yandex-map">Широта: 45.000278</a><br><a href="#yandex-map">Долгота: 48.561944</a>          </div>
```

```
sites <- read.csv("sites.csv")
coords <- NULL
for (i in 1:nrow(sites)){
  site <- sites$out[i]
  webpage <- paste0("https://energybase.ru", site) %>%
    read_html(encoding = "utf-8")
  out <- webpage %>%
    html_nodes("section.contacts div.item > a") %>%
    html_text()
  coord <- as.data.frame(cbind(out[1], out[2]))
  coords <- rbind(coords, coord)
}
```

```
colnames(coords) <- c("lat", "long")
coords$id <- as.numeric(row.names(coords))
coords$lat <- str_extract(coords$lat, "[[:digit:]].*$")
coords$lat <- as.numeric(coords$lat)
coords$long <- str_extract(coords$long, "[[:digit:]].*$")
coords$long <- as.numeric(coords$long)
write.csv(coords, "coords.csv", row.names = FALSE)
```

TRANSLATING WELL TYPES IN ENERGYBASE.RU

```
types_ru <- unique(wells_all$type)
types_ru <- paste0("^", types_ru, "$")
types_en <- c("oil-gas", "gas-oil", "oil-gas-condensate", "oil",
"gas-condensate", "gas")
wells_all$type_ru <- wells_all$type
for (i in 1:length(types_ru)){
  wells_all$type <- gsub(types_ru[i], types_en[i], wells_all$type)
}
```

Type	Known Production	Unknown Production
Gas	0	49
Gas-condensate	4	77
Gas-oil	3	4
Oil	54	195
Oil-gas	12	48
Oil-gas-condensate	22	82

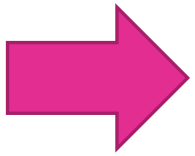
METHODOLOGY: INVERSE-DISTANCE WEIGHTING

- Inverse-distance weighting of petroleum yield using
- $p_i = \sum_{j=1}^{n} \frac{1}{{d_{ij}}^2} p_j / \sum_{j=1}^{n} \frac{1}{{d_{ij}}^2}$,

where p_i is predicted petroleum yield for well i, p_j is known petroleum yield for well j, and d_{ij} measures the distance between two wells i and j

METHODOLOGY: INVERSE-DISTANCE MATRIX

$$\frac{1}{A_{249,249} \bullet A_{249,249}}$$



$$D_{249,249}$$

Distance Matrix
(Great-circle distance,
in kilometers)

Inverse Distance Matrix
(symmetrical)

```
writeDist(rdist.earth(oil.mtrx, miles = FALSE, R = 6366.71), file = "oil.mtrx.csv")
oil.mtrx <- read.delim("oil.mtrx.csv", header = FALSE, sep = " ")
oil.mtrx <- as.matrix(oil.mtrx[2:250,2:250])
for(i in 1:dim(oil.mtrx)[1]) {oil.mtrx[i,i] = 0}
mtrx.dist.inv <- ifelse(oil.mtrx!=0, 1/(oil.mtrx*oil.mtrx), oil.mtrx)
dist.inv <- as.data.frame(mtrx.dist.inv)
dist.weights.sum <- as.data.frame(rowSums(dist.inv))
```

METHODOLOGY: GEOLOGICAL PROVINCE MATCH

GLG Vector (Character)

Geological Match Matrix (binary and symmetrical)

$$G_{249,249}$$

```
mtrx.glg <- matrix(NA, nrow = 249, ncol = 249)
for (i in 1:nrow(oil.wells)){
  for (j in 1:nrow(oil.wells)){
    mtrx.glg[i,j] <- ifelse(
      oil.wells$GLG[i] == oil.wells$GLG[j], ifelse(
        i != j, 1, 0),
       0)
  }
}
glg <- as.data.frame(mtrx.glg)
glg.weights.sum <- as.data.frame(rowSums(glg))
```

METHODOLOGY: PRODUCTION WEIGHTS

Production Vector (numeric) Production Weights Vector (binary)

$O_{249,1}$  $V_{249,1}$

Min: 1.80
Mean: 63.87
Max: 360.0
Not Missing: 54
Missing: 195
N: 249

```
prod <- as.matrix(oil.wells$prod)
prod[is.na(prod)] <- 0
prod.weights <- prod
prod.weights[is.na(prod.weights)] <- 0
prod.weights[prod.weights > 0] <- 1
```

METHODOLOGY: WEIGHTS CALCULATION

Inverse Distance Matrix

Geological Match Matrix

Production Weights Vector

Weights vector

$$D_{249,249}G_{249,249}V_{249,1} = W_{249,1}$$

dist.glg.weights.all <- mtrx.dist.inv %*% mtrx.glg %*% prod.weights

METHODOLOGY: WEIGHTED PREDICTIONS

Inverse Distance Matrix

Geological Match Matrix

Production Vector

Weights vector

Predicted production vector

$$P_{249,1} = \frac{D_{249,249} G_{249,249} O_{249,1}}{W_{249,1}}$$

```
dist.glg.weighted.pred <- mtrx.dist.inv %*% mtrx.glg %*% prod
dist.glg.pred <- dist.glg.weighted.pred/dist.glg.weights.all
```

METHODOLOGY: EVALUATING THE PREDICTIONS

Predicted production vector

Production Vector

$$\sum \left\{ P_{249,1} - O_{249,1} \right\}^2$$

Min: 8.332
Mean: 57.014
Max: 108.477
Not Missing: 54
Missing: 195
N: 249

Min: 1.80
Mean: 63.87
Max: 360.0
Not Missing: 54
Missing: 195
N: 249

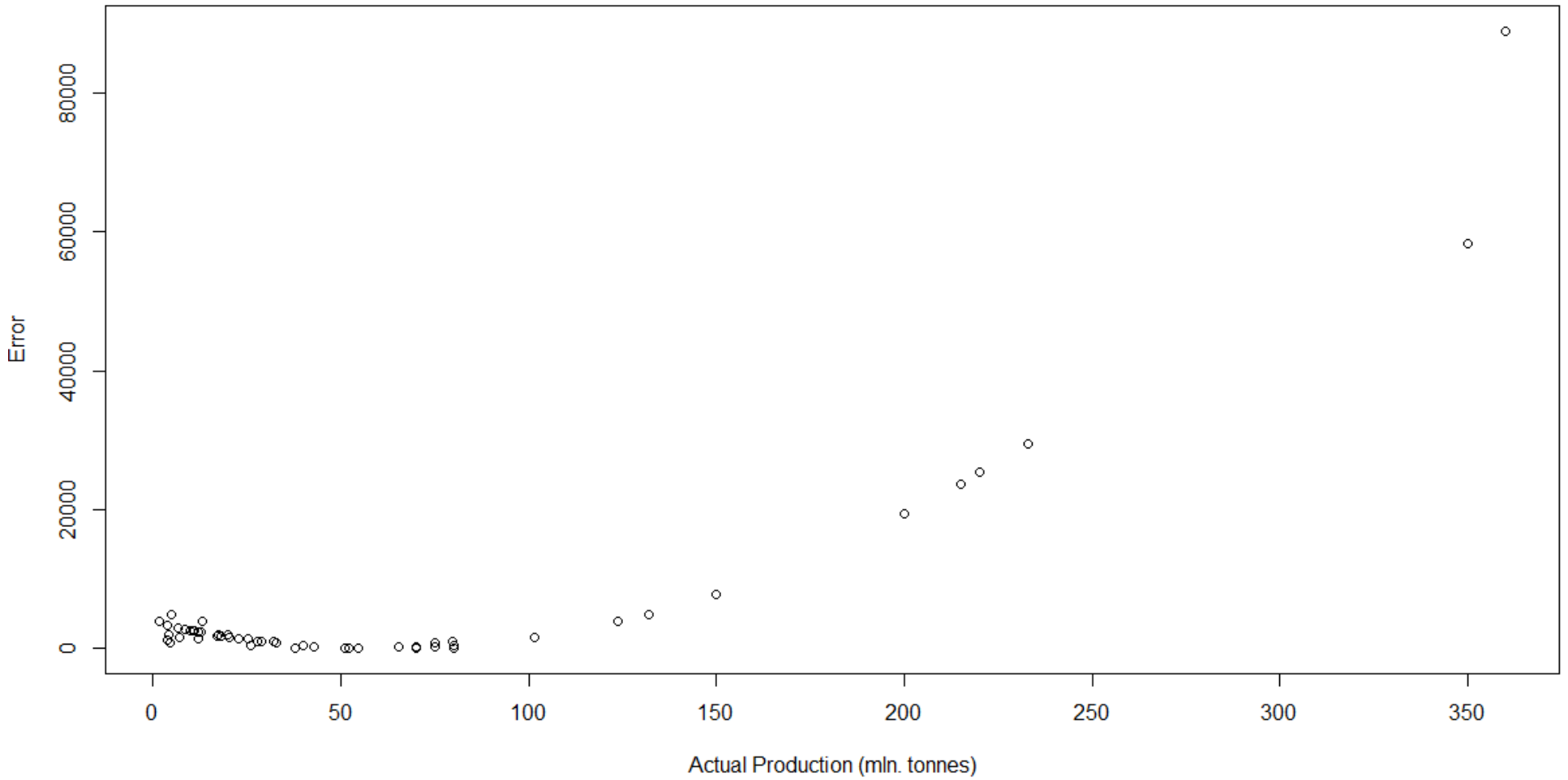
Sum of squared errors

Inverse-distance weighted: 513,153.1

Geological match weighted: 517,393.6

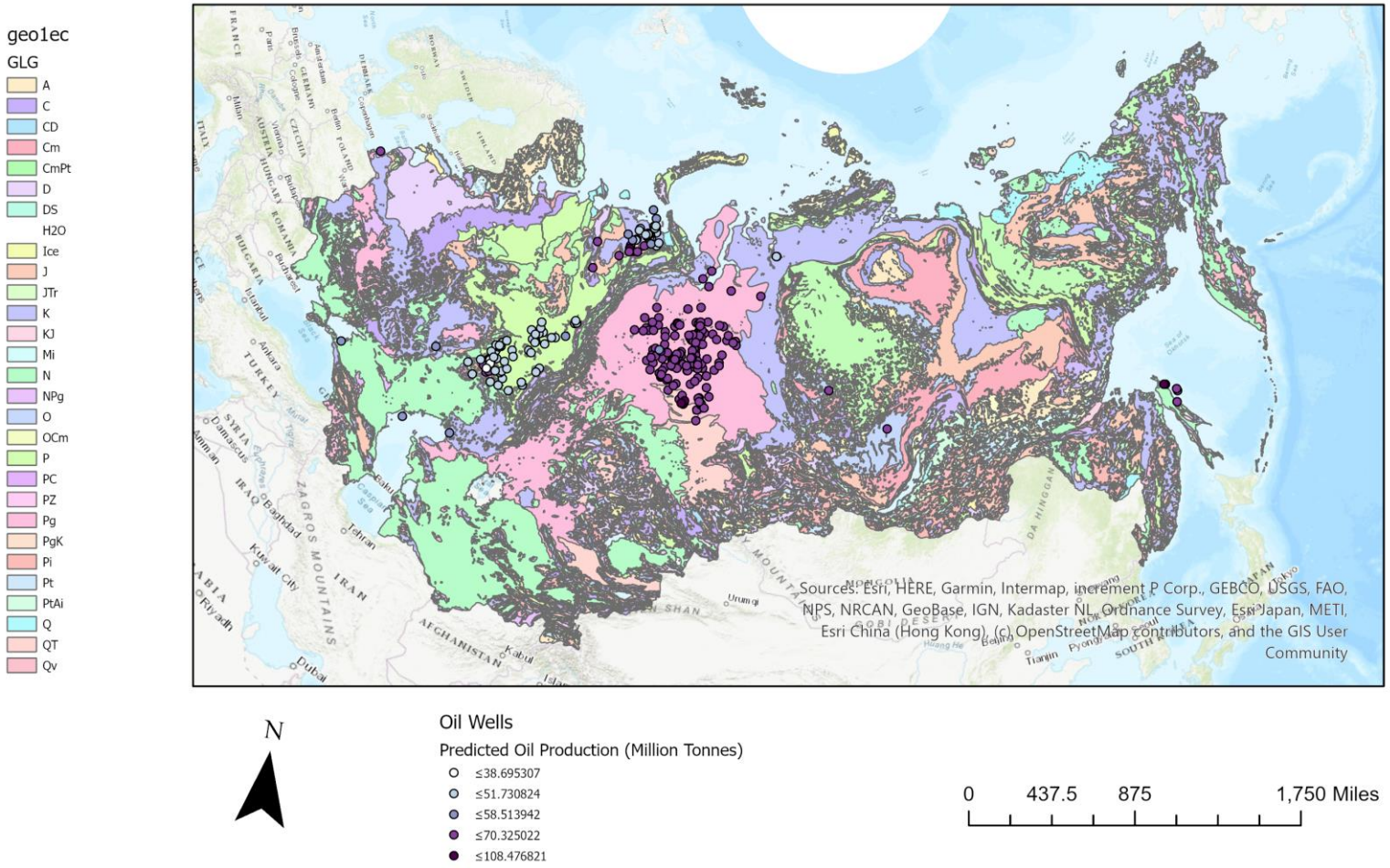
Weighted by both: 327,098.6

METHODOLOGY: EVALUATING THE PREDICTIONS



PREDICTED OIL PRODUCTION

Map of the Geological Provinces of the Former Soviet Union



SOURCES

Aguilera, R. F. (2011). Modeling petroleum resources in provinces of the Former Soviet Union. Energy Exploration and Exploitation, 29(4), 379-396.

Campbell, Robert. (1968). Economic Reform in the USSR. The American Economic Review, 58(2), 547-558. Papers and Proceedings of the Eightieth Annual Meeting of the American Economic Association.

Charpentier, R. R., and Klett, T. R. (2005). Guiding principles of USGS methodology for assessment of undiscovered conventional oil and gas resources. Natural Resources Research, 14(3), 175-186.

Hamida, et al. (2017). An efficient geometry-based optimization approach for well placement in oil fields. Journal of Petroleum Science and Engineering, 149, 383-392.

Ivanov, K.S., et al. (2018). The main factors affecting the distribution of oil fields in the West Siberian Platform. Doklady Earth Sciences, 481(1), 873-876.

Каталог нефтяных и газовых месторождений России. Energybase.ru. Retrieved June 22, 2020, from https://energybase.ru/oil-gas-field.

Persits, F.M., et al. (1999). Maps showing geology, oil and gas fields and geologic provinces of the former Soviet Union. U.S. Geological Survey Open-File Report 97-470-E, 13 p., https://doi.org/10.3133/ofr97470E.