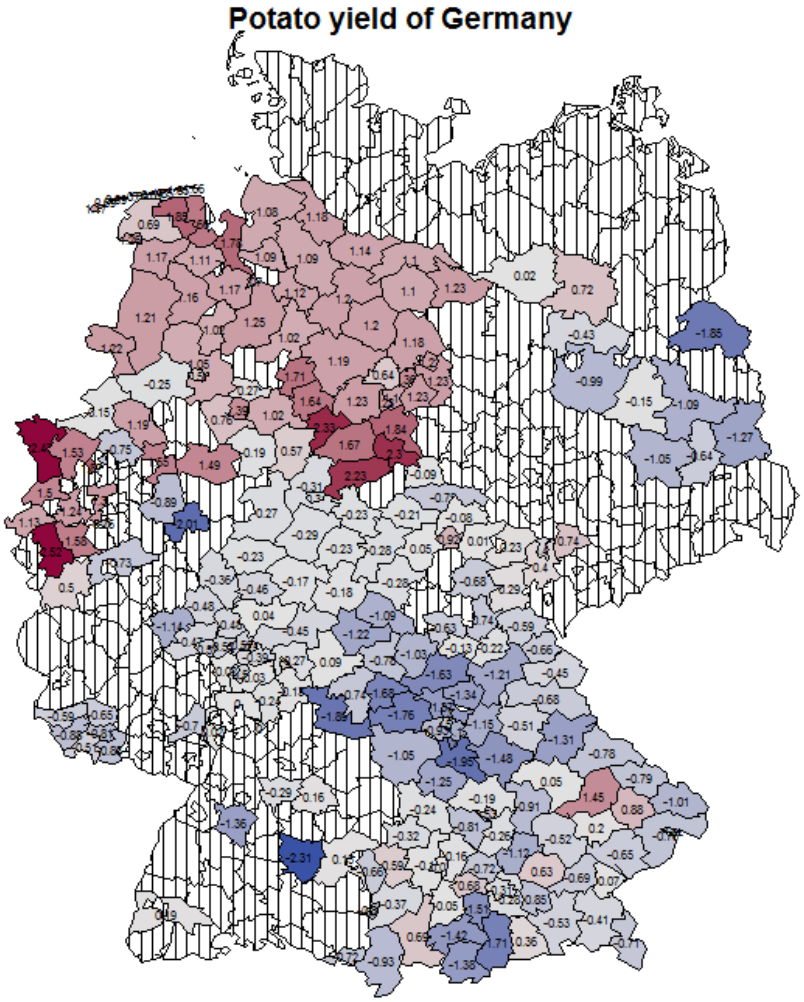
Agricultural yield estimation with spatial smoothing

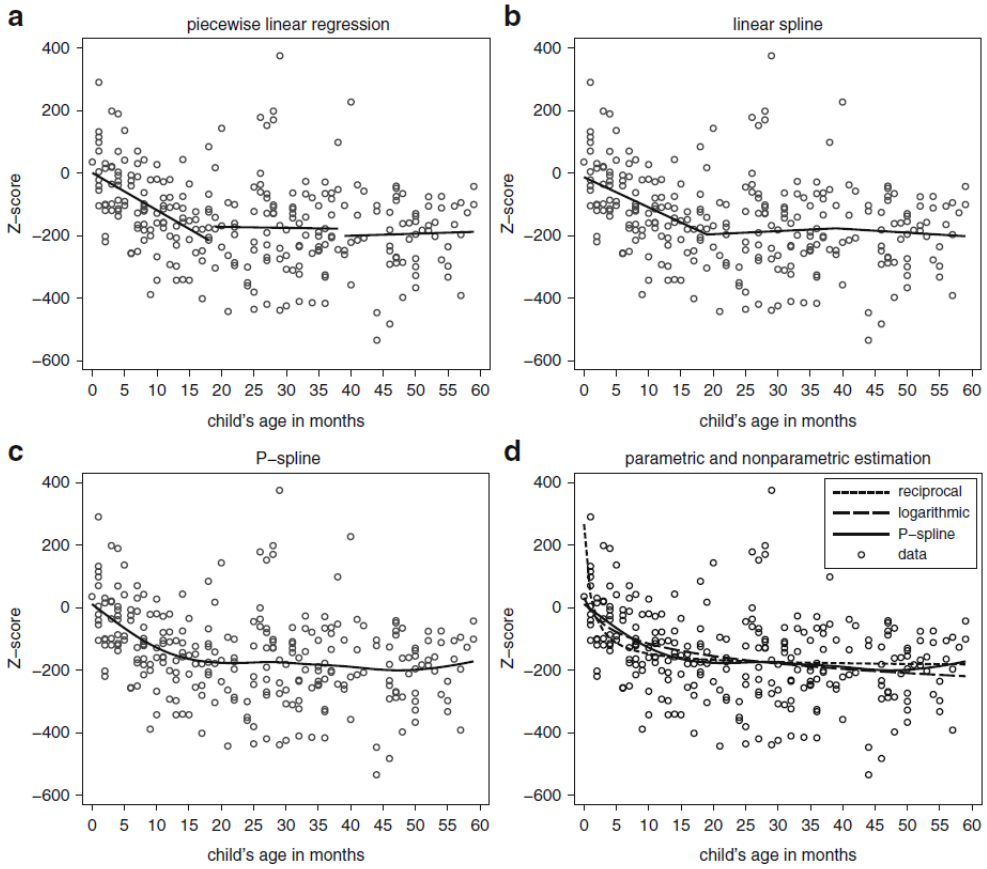
Caner Erdem

Goal



- Spatial smoothing
- Markov random field
- Penalized least square criterion
- One dependent variable and one covariate

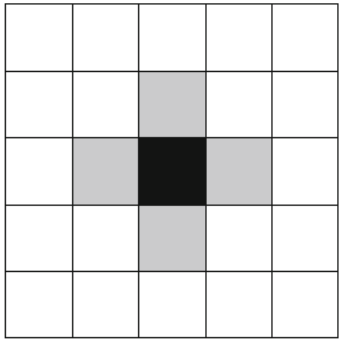
Background - 1

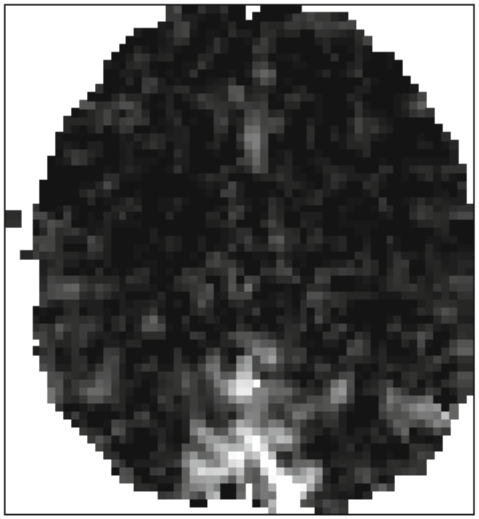


Source: Fahrmeir et al., 2013

- Nonparametric regression
  - Nonlinear effects
  - Difficult to fit parametric models
  - Data-driven & flexible
  - Exploratory data analysis

Background - 2



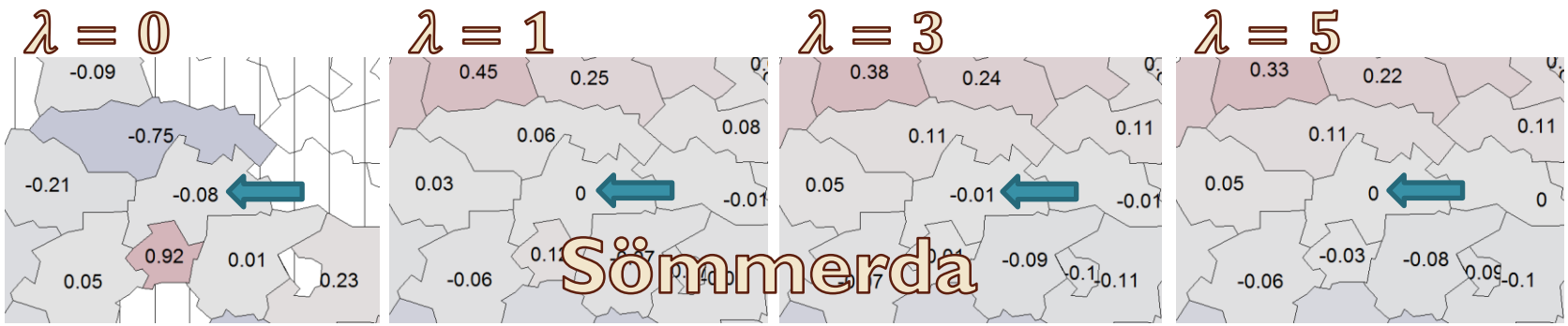


Source: Fahrmeir et al., 2013

- Spatial smoothing
  - Nonparm. method
  - Continuous location variables: geographic coordinates
  - Discrete spatial units: regions or grid
    - Sometimes overlap: human brain mapping
  - Euclidean distance vs. neighborhood

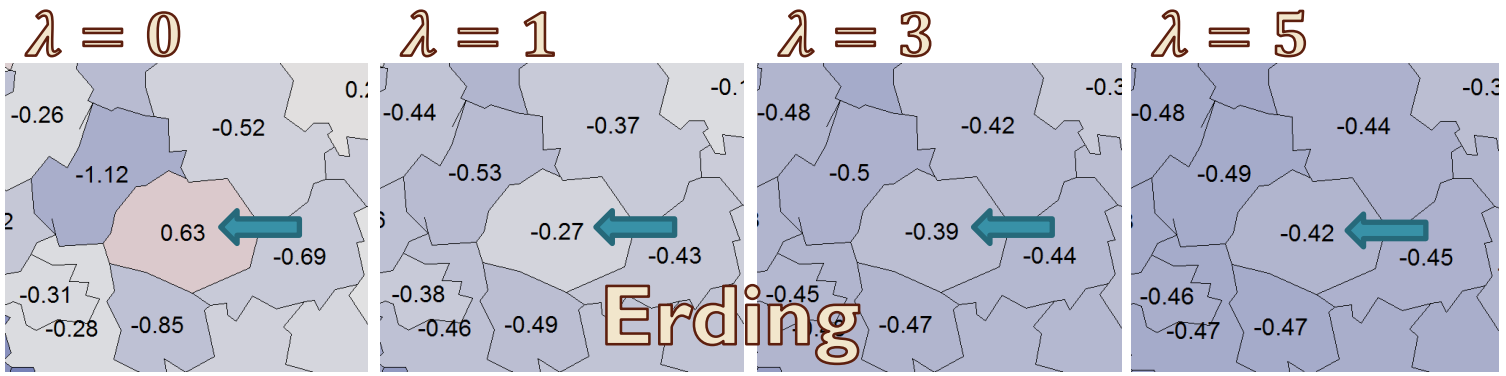
Background - 3

- Markov random field
  - Relationship between regions: neighborhood
    - A graph $G = (V, E)$
    - $V = \{1,2, \dots, d\}$ associated with γ_s for $s = 1, \dots, d$
    - Neighborhood of s is $N_s, r \in N_s \Leftrightarrow (s, r) \in E$
    - The property: $p(\gamma_s \mid \{\gamma_r\}_{r \in V \setminus s}) = p(\gamma_s \mid \{\gamma_r\}_{r \in N_s})$
    - Information transfer through local connections



Background - 4

- Penalized least square (PLS)
  - $s \sim r$ denotes region s and r are neighbors
  - Regression coeff. $f_{geo}(s) = \gamma_s, s = 1,2, \ldots, d$
  - PLS criterion:
    - $\sum_{i=1}^{n}(y_i - f_{geo}(s_i))^2 + \lambda \sum_{s=2}^{d} \sum_{r \in N(s), r<s}(\gamma_r - \gamma_s)^2$
    - Penalty term discourages large deviations in regression coefficients of nearby regions



Background - 5

```
> GermanyGraph[5:10,5:10]
      1051 1053 1054 1055 1056 1057
1051     4    0   -1    0    0    0
1053     0    7    0    0    0    0
1054    -1    0    2    0    0    0
1055     0    0    0    4    0   -1
1056     0    0    0    0    4    0
1057     0    0    0   -1    0    5
```

- Model and solution
  - Model: $y_i = f_{geo}(s_i) + \varepsilon_i\ ,\ \varepsilon \sim N(0,\ \sigma^2)$
  - Model in matrix notation: $y = Z\gamma + \varepsilon$
  - Design mat. $Z[i,s] = \begin{cases} 1 & if\ y_i\ is\ in\ region\ s \\ 0 & otherwise. \end{cases}$
  - Penalty mat. $K[s,r] = \begin{cases} -1 & s \neq r,\ \ s \sim r, \\ 0 & s \neq r,\ \ s \nsim r, \\ |N(s)| & s = r. \end{cases}$
  - PLS estimate: $\hat{\gamma} = (Z'Z + \lambda K)^{-1} Z'y$

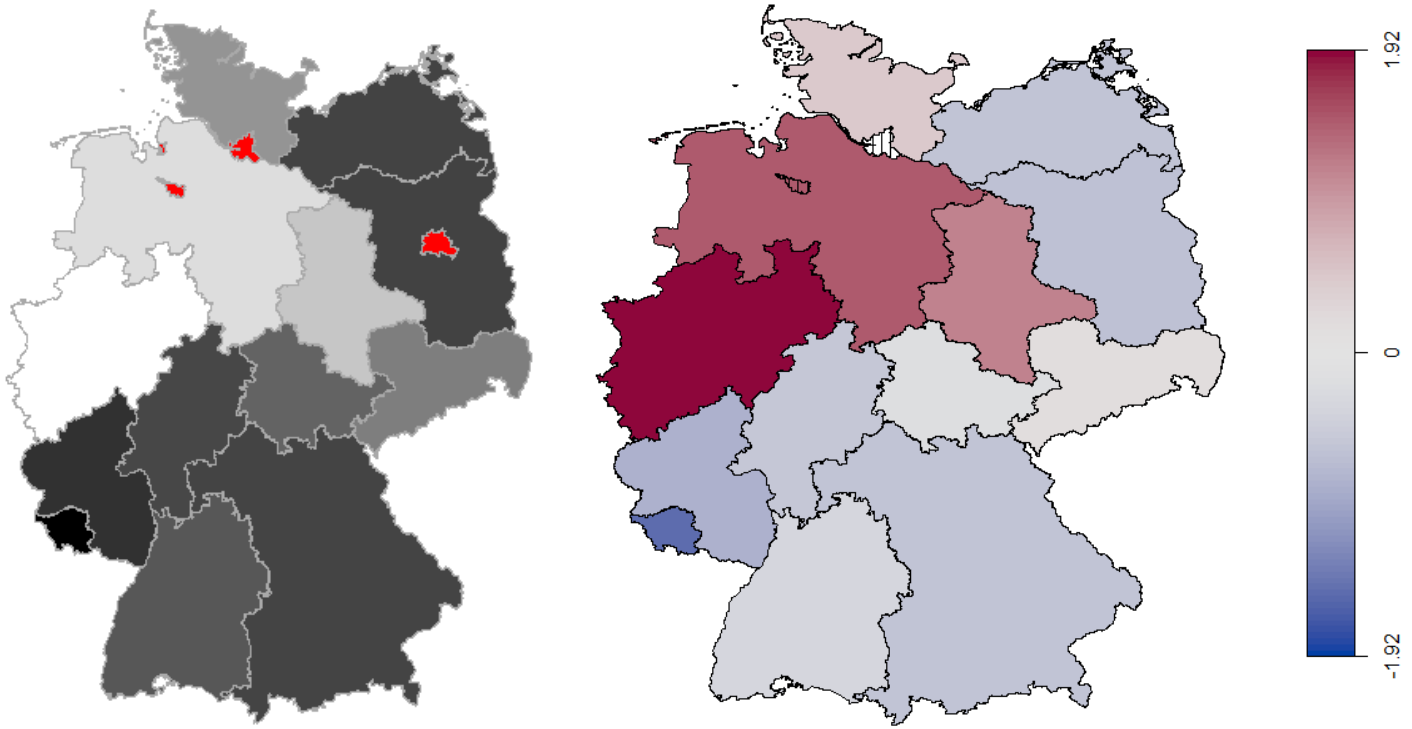
Background - 6

- Concerns about this method
  - Simplicity
  - PLS estimate $\hat{\gamma}$: all in one
    - i.e. spatial and non-spatial covariates
  - Simultaneous est.
    - Mixed models
    - Bayesian approaches
    - BayesX software?

Source: eBay, 2017

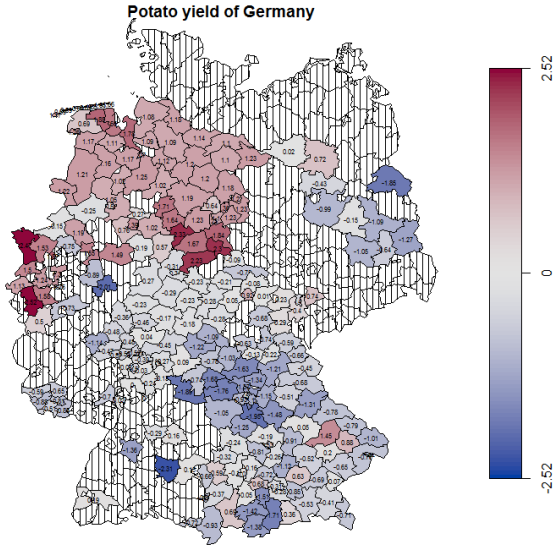
First steps

- state based potato yield (100kg/ha)
- only three states without official data
- gadm.org & regionalstatistik.de



Estimation of spatial effects

- Yield data for 230 out of 439 districts
- Estimation steps:
  - Penalty matrix K
    - Via neighborhood
  - Ruegen's neighbor
    - Nordvorpommern
  - Design matrix Z
    - Via yield data
  - PLS estimation
    - $\hat{\gamma} = (Z'Z + \lambda K)^{-1} Z'y$

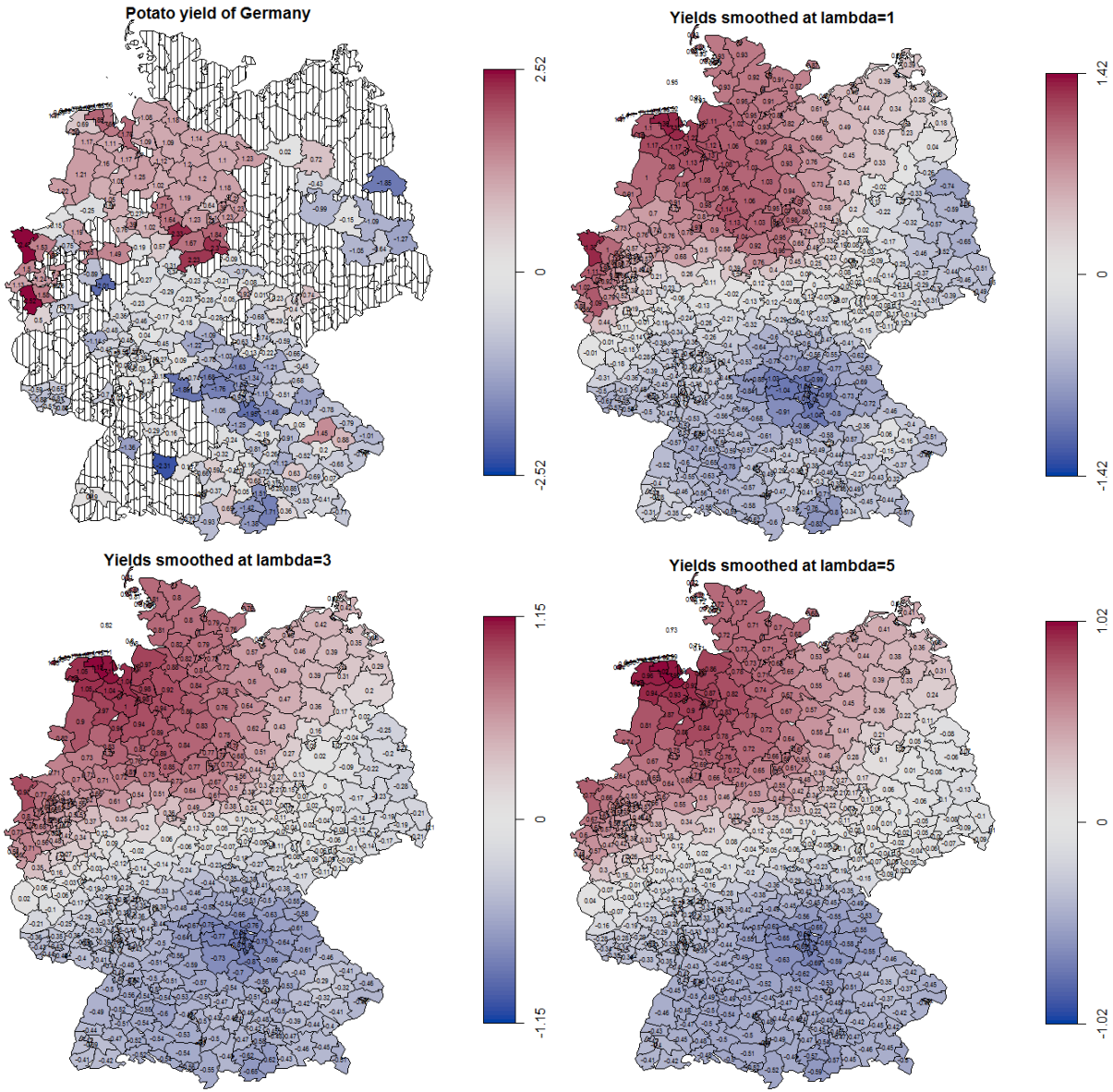


Before		
	13057	13061
13057	5	0
13061	0	0

After		
	13057	13061
13057	6	-1
13061	-1	1

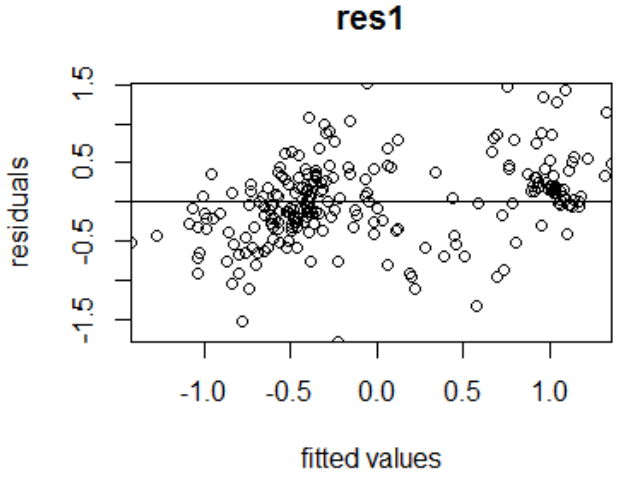
Results

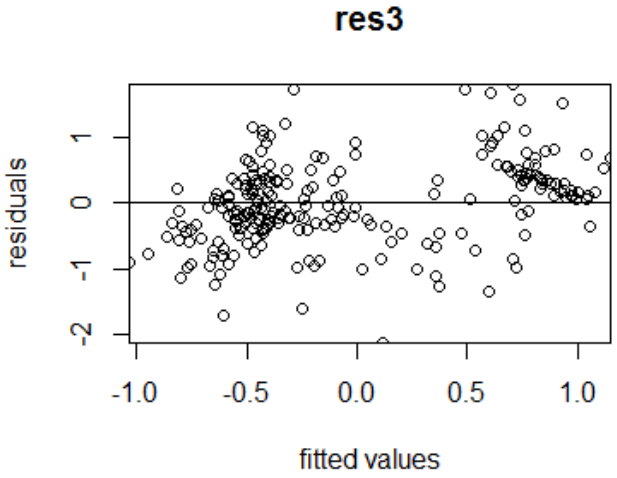
- as $\lambda \uparrow$ $smth \uparrow$
- but, $as\ \lambda \uparrow$ $\sigma_\varepsilon^2 \uparrow$

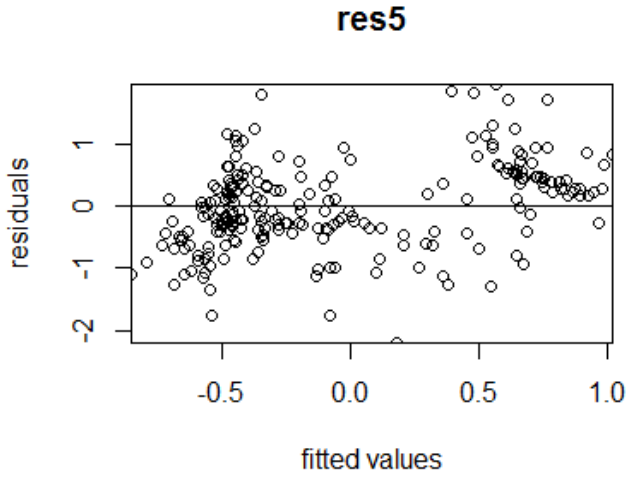


Assumptions

- ~ homoscedastic and linear residuals
- $\sigma^2_{\varepsilon|\lambda=1} = 0.2673956$
- $\sigma^2_{\varepsilon|\lambda=3} = 0.4040608$
- $\sigma^2_{\varepsilon|\lambda=5} = 0.4691135$

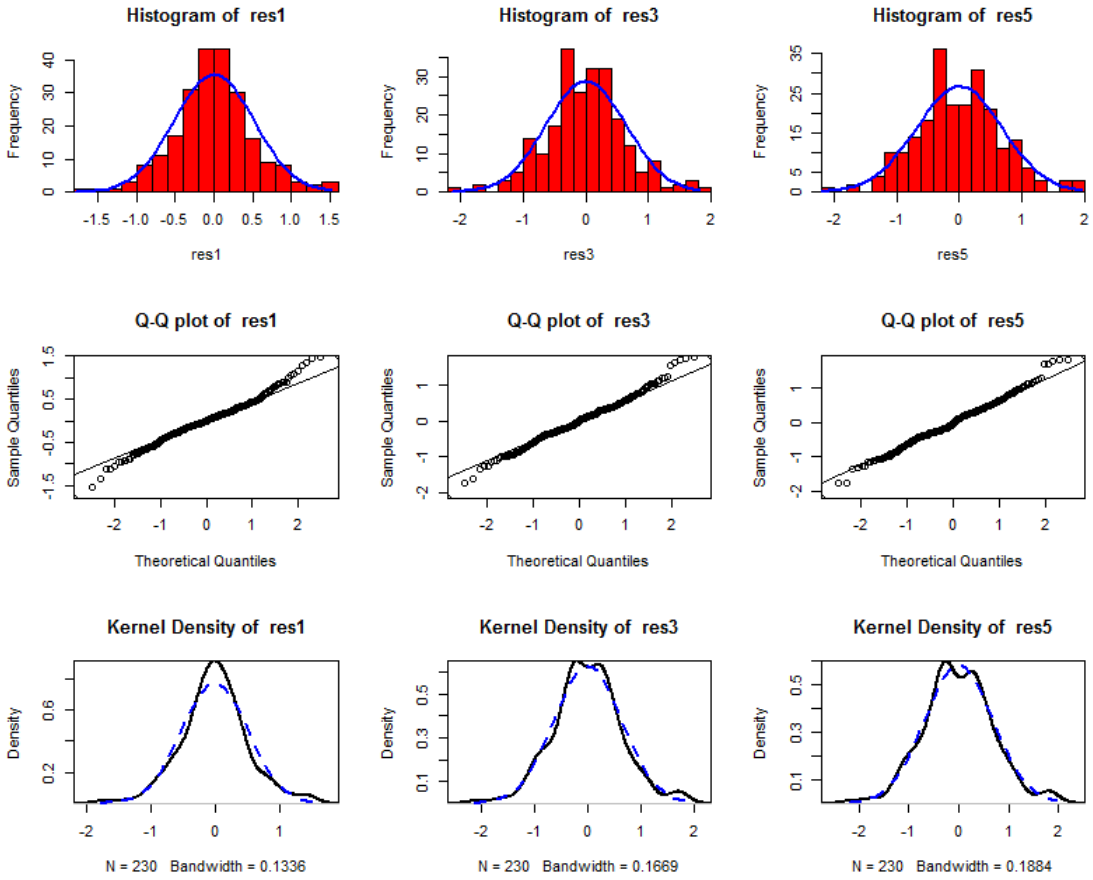






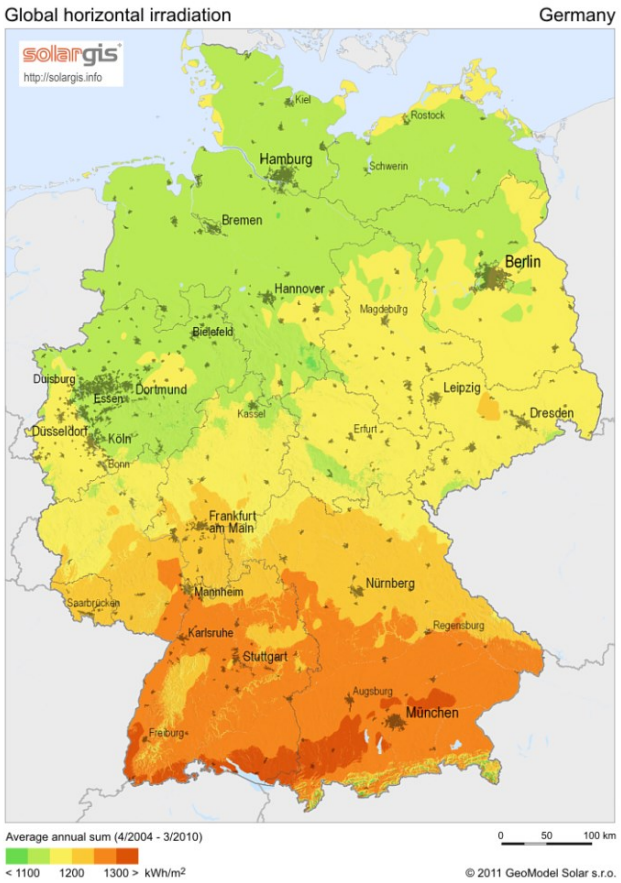
Assumptions contd.

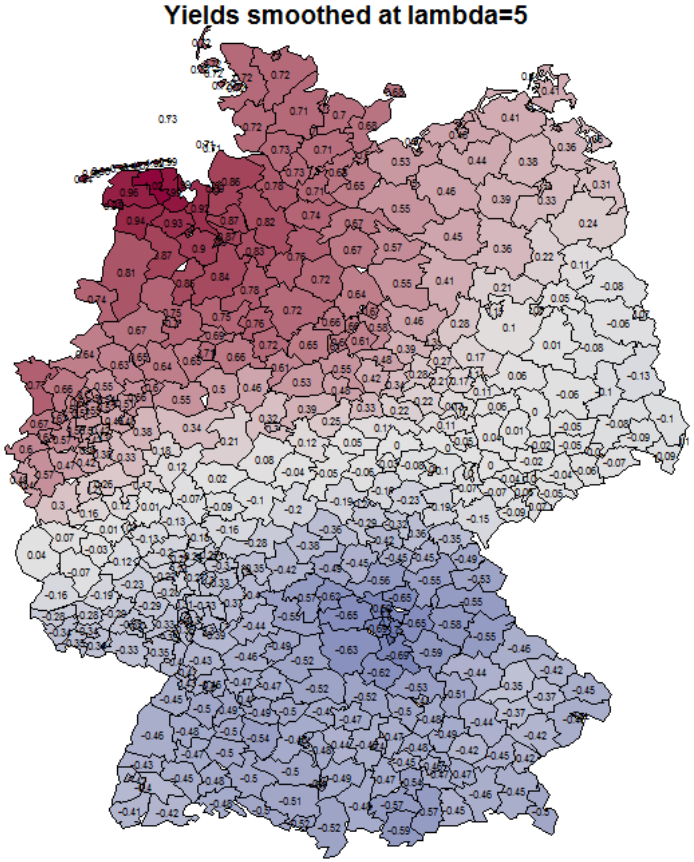
- ~ normally distributed residuals



Reason

- Solar potential? (no, probably other gainful products)





Source: Wikipedia, 2017

Thank you…

Caner Erdem

Sources

- Fahrmeir, L., Kneib, T., Lang, S., & Marx, B. (2013). Regression: Models, Methods and Applications. Berlin Heidelberg: Springer-Verlag.
- Fleet, D., & Jepson, A. Markov Random Fields. Retrieved July 08, 2017, from http://www.cs.toronto.edu/~kyros/courses/2503/Handouts/mrf.pdf
- Wikipedia. Solar power in Germany. Retrieved July 12, 2017, from https://en.wikipedia.org/wiki/Solar_power_in_Germany
- eBay. 8 Reasons Why Playing in the Sand Is Good for Kids. Retrieved July 12, 2017, from http://www.ebay.com/gds/8-Reasons-Why-Playing-in-the-Sand-Is-Good-for-Kids-/10000000177634049/g.html