Small Area Estimation and GSI

What is Small Area Estimation

Direct Estimation:

- Each area only uses information collected in that area
- Problem:
    * many small areas,
    * data available is small
    * estimates are highly variable with poor precision

- E.g. 100 salmon stocks with one or two CWT from each stock.

Small Area Estimation and GSI

Indirect, synthetic, small area estimation

- Areas close together "share" common experiences, e.g.
    * similar survival rates
    * similar exploitation patterns

- Use information from neighbouring areas to augment information from an area

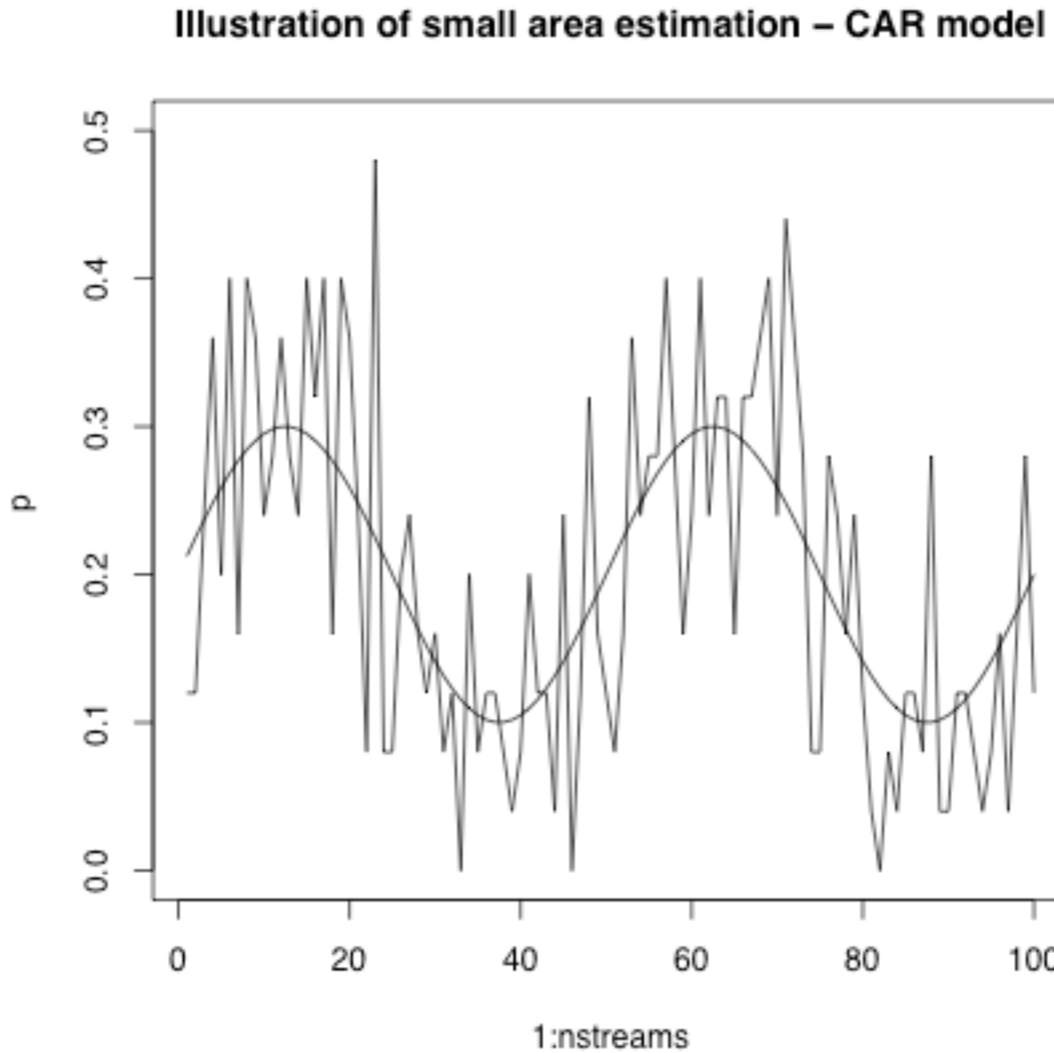
Small Area Estimation and GSI Example

Generate data from $Binomial(p_i; n = 25)$ where i=1…100 and p_i vary "smoothly" over the streams but averages about .2.

Each stream will have about 5 recoveries, and

$$rse \approx 45\%$$

Small Area Estimation and GSI Example



Small Area Estimation and GSI Example

A very simple conditional first order autoregressive moving average model (CAR(1))

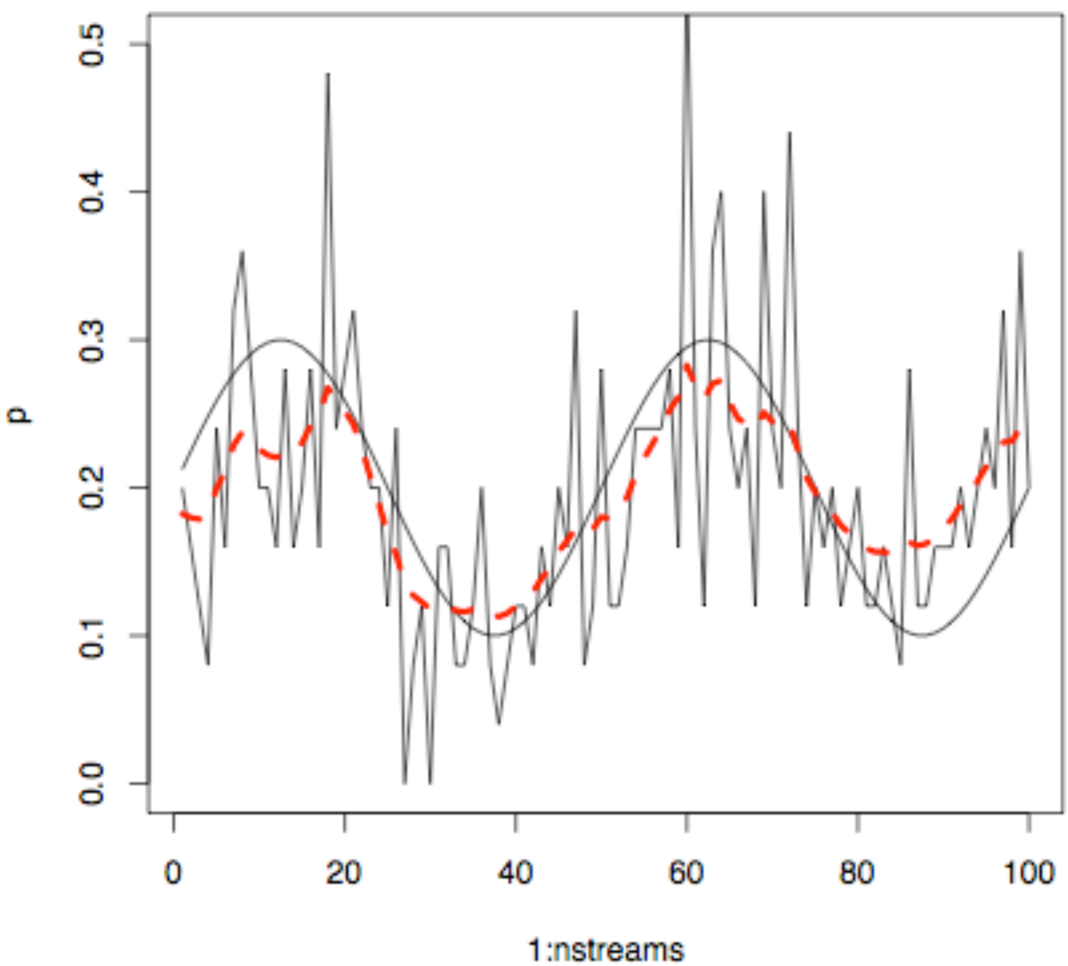
$$Y_i \sim Binomial(25, p_i)$$

$$\text{logit}(p_i) = \mu + \varepsilon_i$$

$$\varepsilon_{\text{i}} | \varepsilon_{\text{-i}} \sim \begin{cases} Normal(\varepsilon_{i+1}, \sigma^2) & i = 1 \\ Normal(\frac{\varepsilon_{i-1} + \varepsilon_{i+1}}{2}, \frac{\sigma^2}{2}) & i = 2, \ldots, N \\ Normal(\varepsilon_{i-1}, \sigma^2) & i = N \end{cases}$$

Small Area Estimation and GSI Example

Figure 1: Illustration of small area estimation – CAR model



Small Area Estimation and GSI
Example
Gain in precision

- Direct estimates have $rse \approx 45\%$45%.

- Small area estimates have $\overline{rse} \approx 16\%$

- Direct estimtes would require about a $(45/16)^2$=8 times increase in effort to achieve same precision as small area estimates.

Small Area Estimation and GSI
Summary

• Small area estimation is form of "localized" pooling.

• Pooling distance can be set a prior or can be estimates with higher weights given to streams that are "closer" to the stream of interest or weights can also be functions of distance.

• Pooling can be done over time and space

• Substantial gains of precision are possible with relative simple models.

• Key to approach is commonality of close sources.