# On the precision of predicting fishing activities and locations

Angela Münch & Geret DePiper

NOAA Northeast Fisheries Science Center/SSB, Woods Hole (USA)

06/14/2016

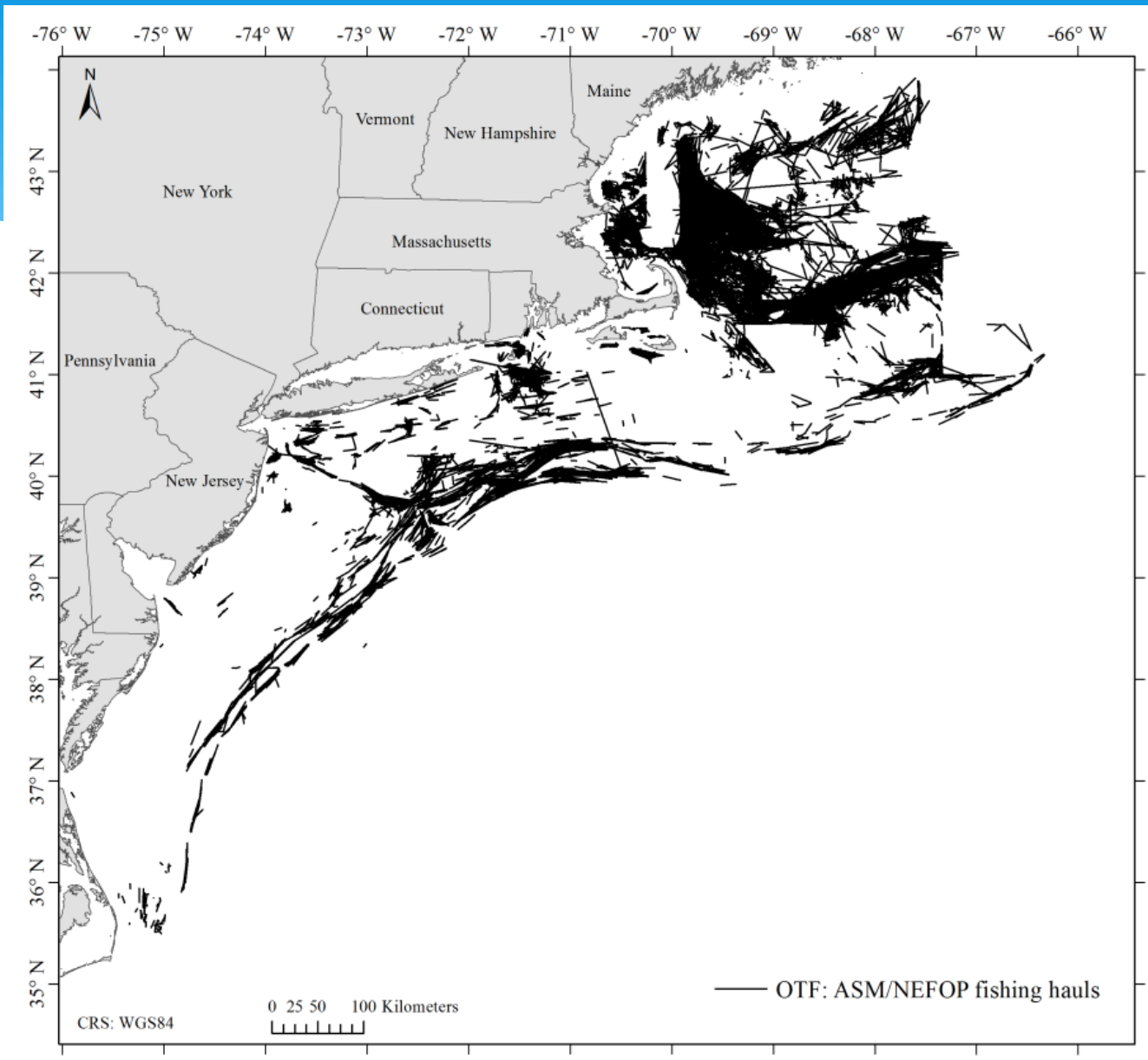
# Data on fishing activity

- Observer data (ASM/NEFOP) – ca. 10-20 percent of all commercial fishing trips
- Vessel monitoring system (VMS) – ca. 80 percent of the fleet in recent years carries a responder
- Logbook data (VTR) – requested for landings, but only one coordinate for the fishing activity

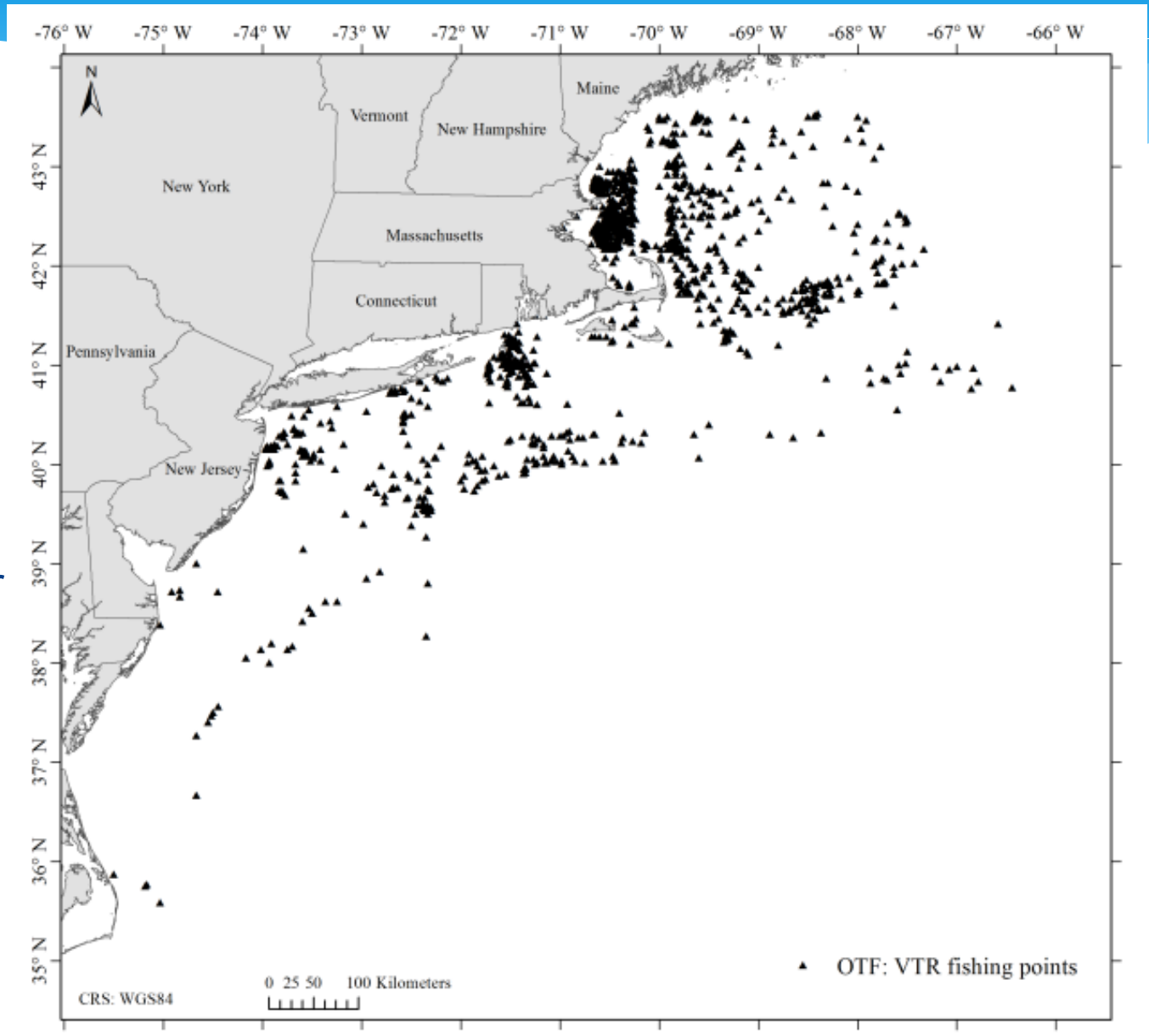
→ Who is fishing where (and what)?

- → Stock assessment
- → Impact of fishing activity on e.g. seabed
- → Reaction to regulations (e.g. closure)

Fishing activity according to observer programs (NEFOP/ ASM) in 2012

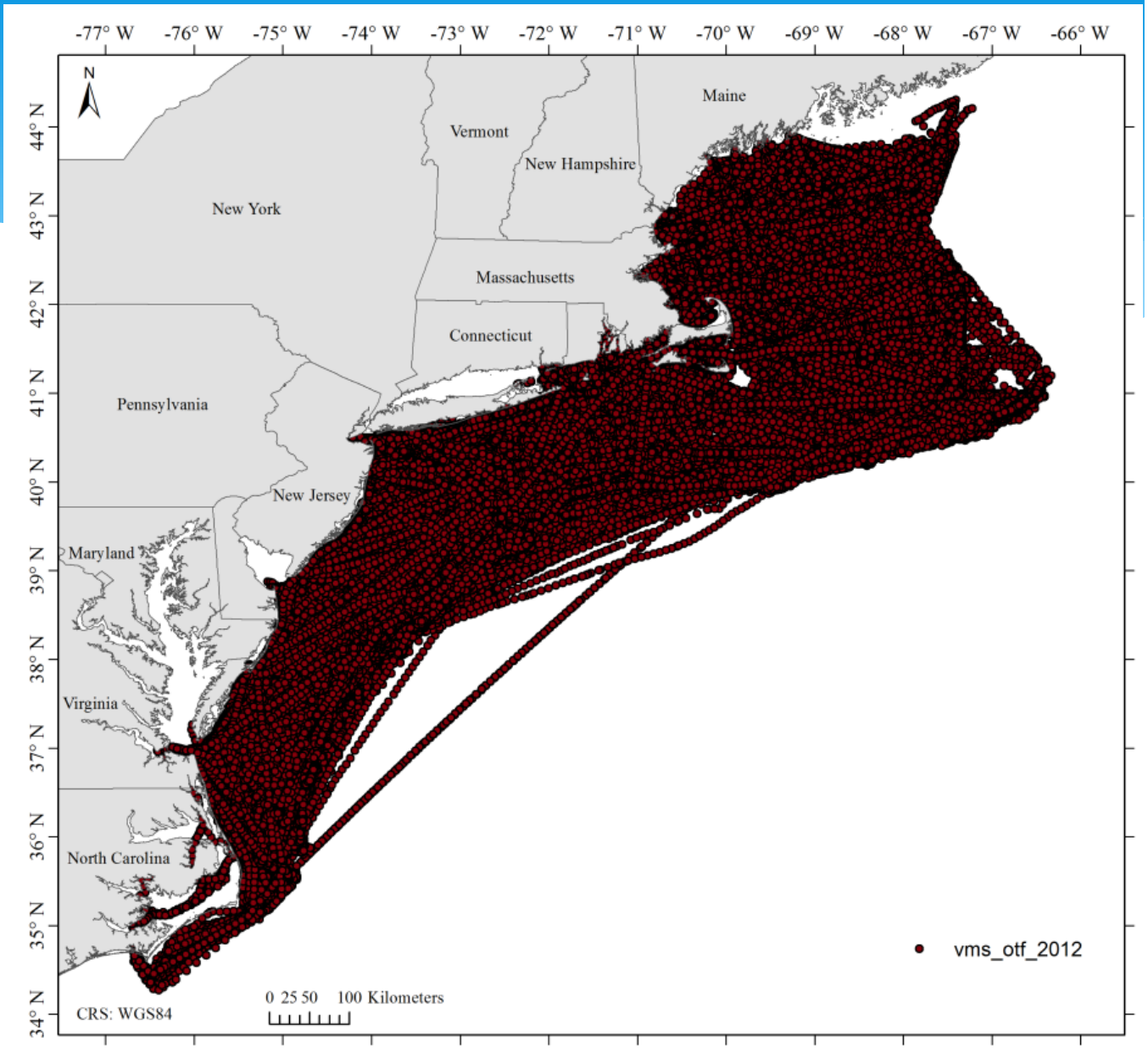


Fishing activity according to logbook in 2012 for bottom otter trawl



1. Research Question:

Which VMS poll is a fishing location?



# Fishing activity prediction based on VMS data

- Track interpolation vs. point summation based on speed rules
- Speed rules vs. probabilistic models
  - Palmer/Wigley 2009: a speed range of:
    - 3.7–7.4 km/h for bottom otter trawl
    - 4.6–11.1 km/h for scallop dredge, and
    - 0.2–2.4 km/h for sink gill net
  - Logit: $\Pr(\mathrm{Y}=1 \mid \mathrm{x}) = \mathrm{F}(\mathrm{x}, \beta) = \frac{e^{x\prime\beta}}{1+e^{x\prime\beta}}$
  - GAM: $\Pr(\mathrm{Y}=1 \mid \mathrm{x}) = \sum_1^p s_j(X_j) = s(X)$
    - X: speed, speed-speed$_{t-1}$, speed range, contour, contour-contour$_{t-1}$, angle, angle-angle$_{t-1}$, position, accumulated distance and distance$^2$, temporal dummies (for e.g. quarter, day of the week, and moon phase) and locational dummies (for e.g. Territorial Sea, Contiguous Zone)

# Prediction quality based on observed trips

| | Speed rule of Palmer & Wigley (2009) | VMS-GLM | VMS-GAM |
|---|---|---|---|
| Gear | Scallop Dredge | | |
| | (N=361, n=133,581) | (N=188, n=64,789) | |
| Positive-positive | 61.8 | **86.4** | **87.8** |
| False-positive | 26.8 | **43.2** | **43.2** |
| $Diff_G$ | 35.0 | 43.2 | 44.6 |
| | | | |
| Gear | Sink Gillnet | | |
| | (N=1,659, n=37,148) | (N=831, n=18,553) | |
| Positive-positive | 61.9 | **89.5** | **90.0** |
| False-positive | 22.1 | **29.3** | **29.2** |
| $Diff_G$ | 39.8 | 60.3 | 60.8 |
| | | | |
| Gear | Bottom Trawl | | |
| Rule | (N=1,768, n=196,077) | (N=891, n=105.424) | |
| Positive-positive | **82.6** | **92.7** | **92.3** |
| False-positive | **18.8** | **29.3** | **28.5** |
| $Diff_G$ | 63.8 | 63.3 | 63.8 |
| Overall Diff | 138.6 | 166.8 | 169.2 |

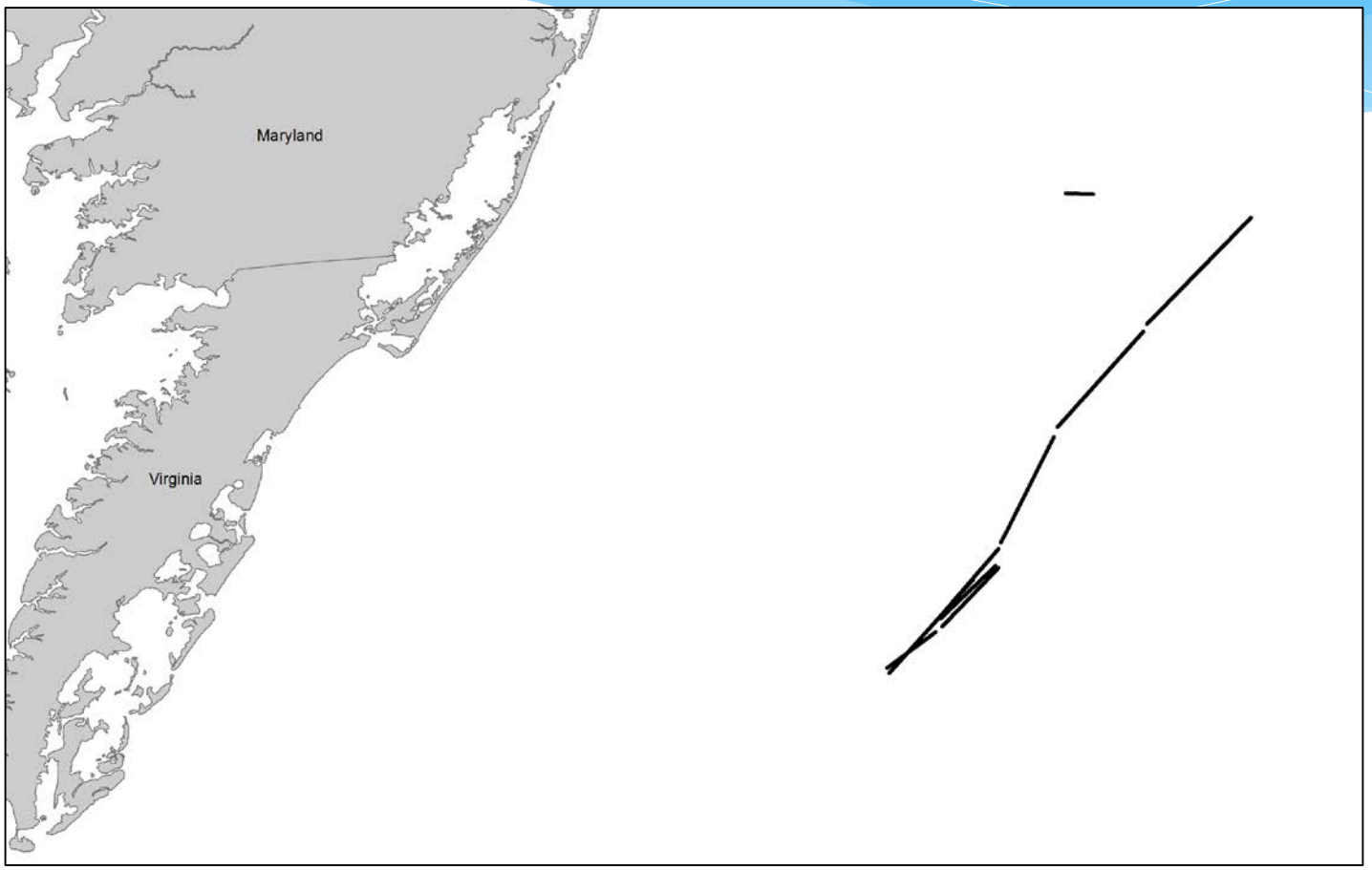
# Prediction of fishing location

- Research question2: How to transfer point data into areas?

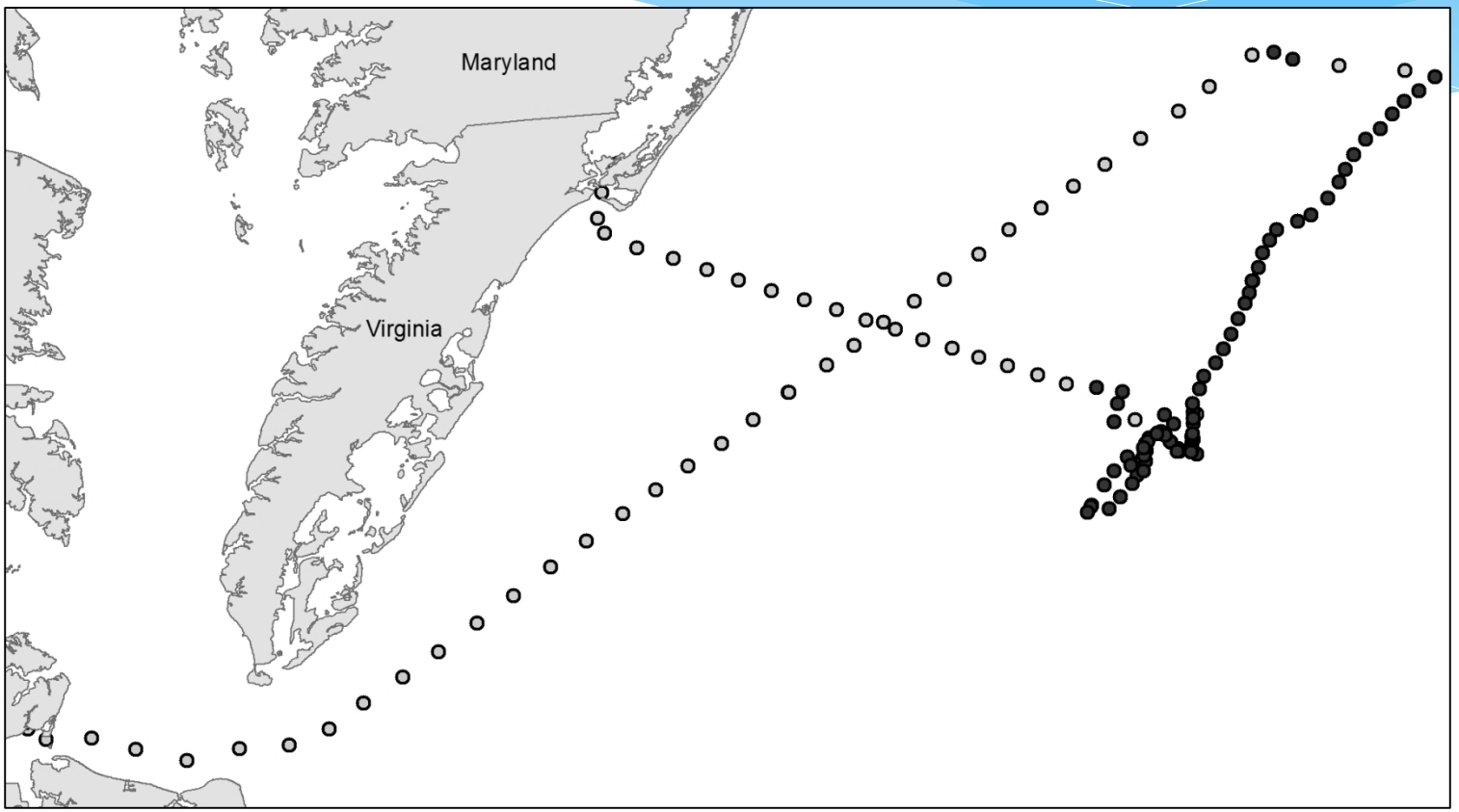
- Data:
  - ASM/NEFOP: Start- and end coordinates of fishing hauls
  - VMS: GLM/GAM predicted VMS polls
  - VTR: Fishing location – one coordinate

- Method:
  - Aggregation into a grid
  - Determine an area (buffer/ellipse) before gridding

# Example – trip A
## - observed fishing hauls -



# Example – trip A
## - VMS data -

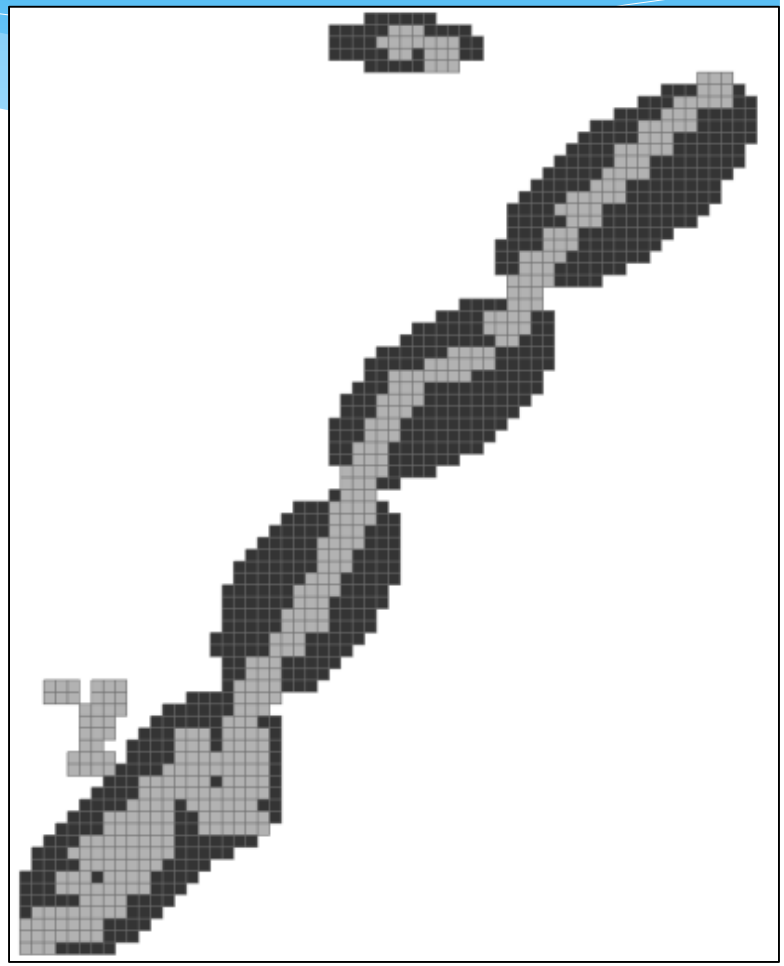



VMS-GAM predicted fishing (dark grey) and non-fishing (light grey).

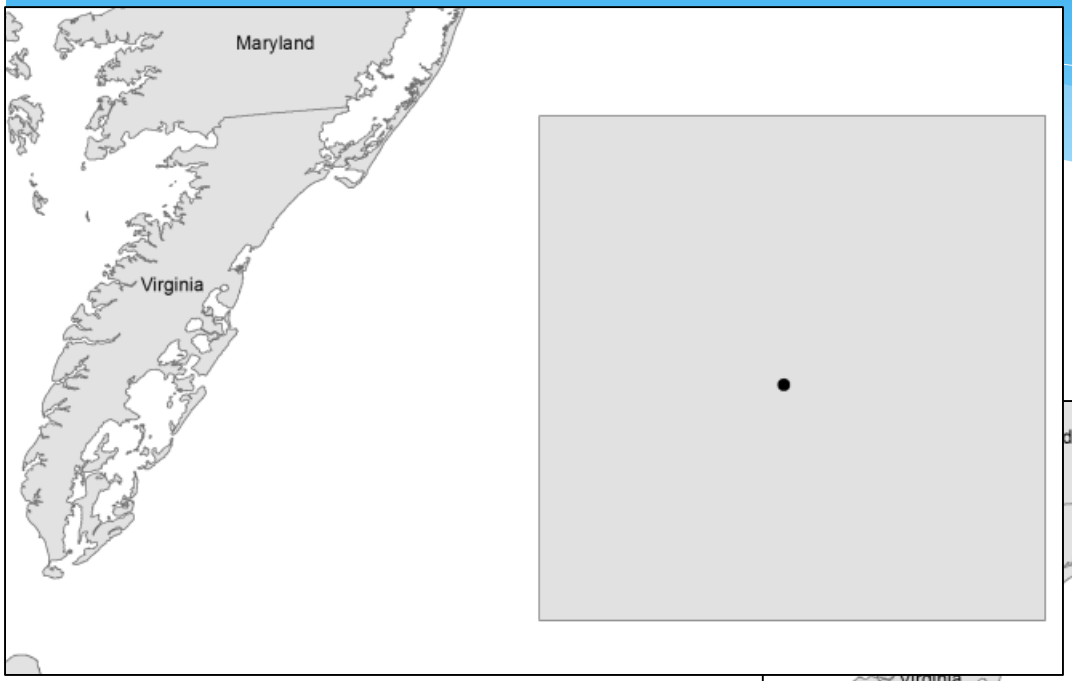
# Example – trip A

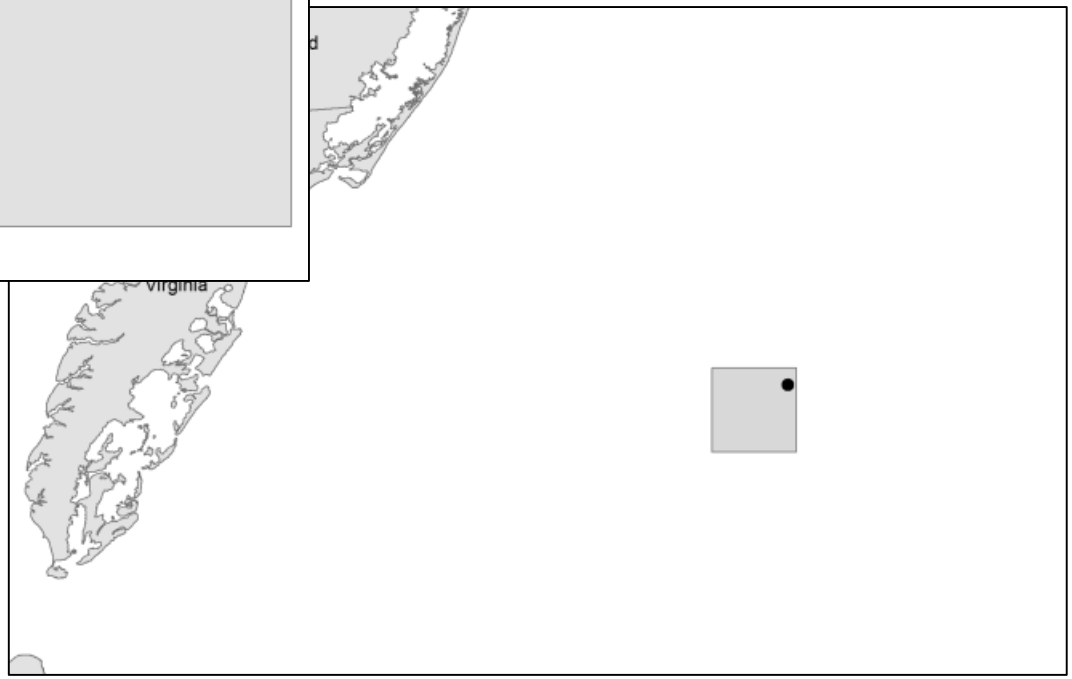
- VMS: Kernel smoothing (gear dependent bandwidth)
- ASM/NEFOP: create ellipse around hauls with width of 95$^{th}$ quantile deviation of VMS

→ Grid (1 km$^2$) – presence/absence comparison & weighted comparison



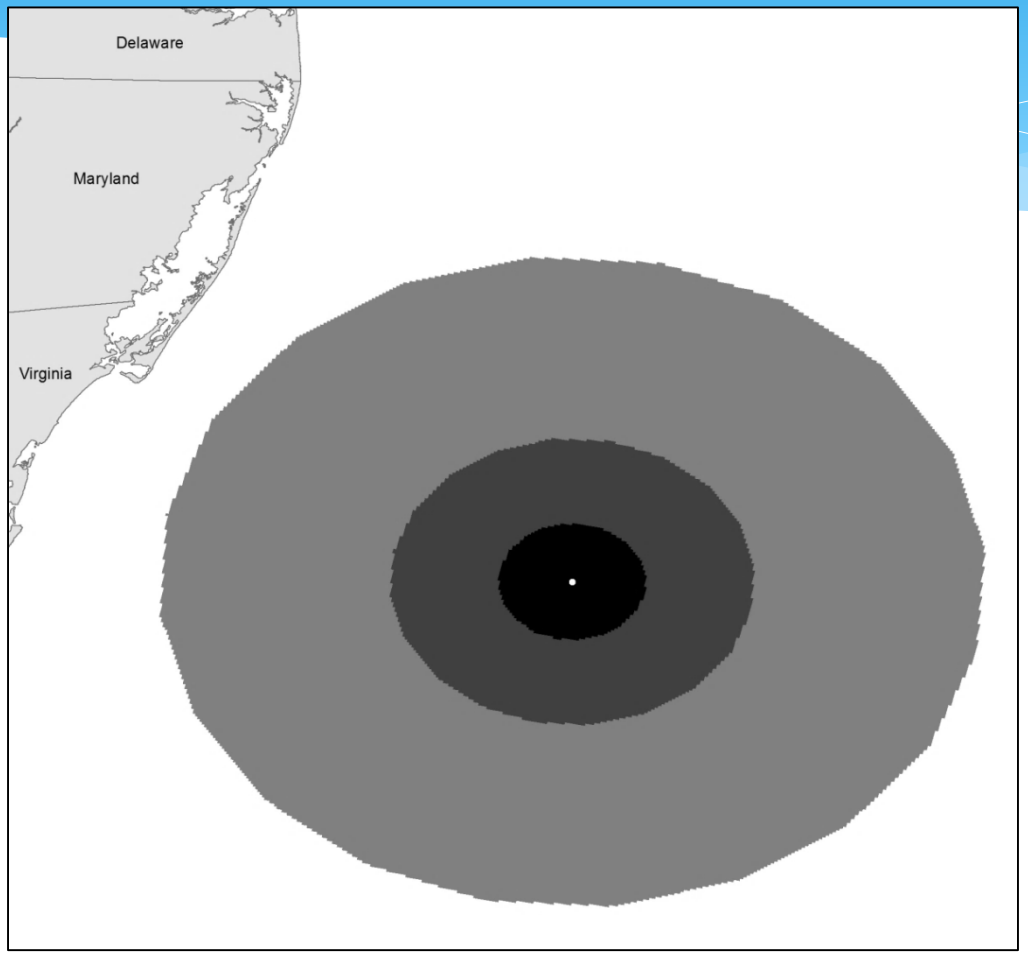
# Example – trip A
## - logbook data -




Aggregating VTR-coordinates into Statarea or TMSQ

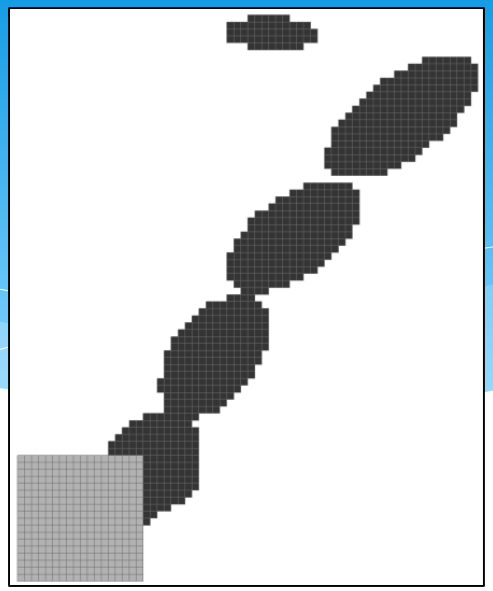
# Example – trip A
## - logbook data -

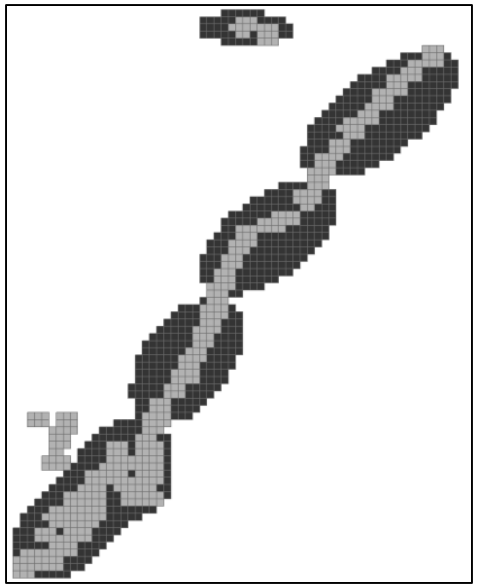


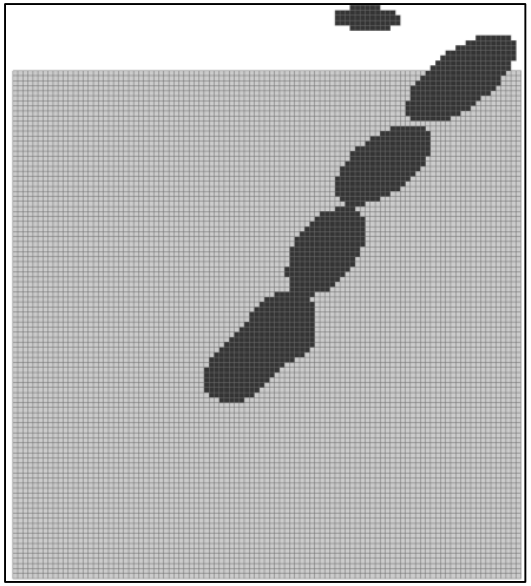

Predicting spatial extent of VTR-coordinates based on trip characteristic (e.g. gear, or duration)

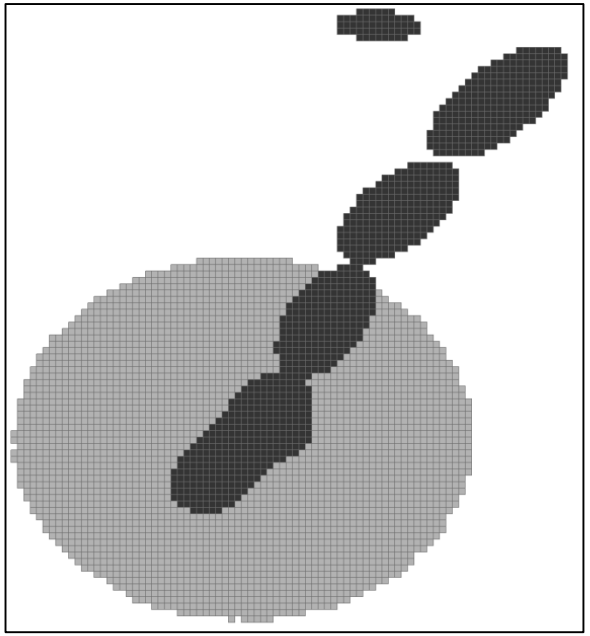
# Grid Overlay

Overlay of ASM/NEFOP with VMS data and VTR data in a 1x1 km grid









# Prediction quality
## - Presence/Absence -

| | VMS-GLM | VMS-GAM | VTR-Statarea | VTR-TMSQ | VTR-CDF |
|---|---|---|---|---|---|
| Gear | Scallop Dredge | | | | |
| Positive-positive | **87.0** | **86.9** | 99.5 | 68.7 | 88.5 |
| False-positive | **24.6** | **19.1** | 85.0 | 40.0 | 56.3 |
| Bandwidth (assumed) | 0.005 | 0.005 | | | |
| Gear | Sink Gillnet | | | | |
| Positive-positive | **95.0** | **95.0** | 99.5 | 91.3 | 92.9 |
| False-positive | **48.7** | **47.8** | 95.8 | 79.7 | 81.1 |
| Bandwidth (Gaussian) | 0.0037 | 0.0037 | | | |
| Gear | Bottom Trawl | | | | |
| Positive-positive | **93.1** | **92.7** | 98.7 | 69.8 | **89.2** |
| False-positive | **21.9** | **19.8** | 64.4 | 29.8 | **37.3** |
| Bandwidth (assumed) | 0.01 | 0.01 | | | |

# Prediction quality
## - weighted comparison -

| | VMS-GLM | VMS-GAM | VTR-Statarea | VTR-TMSQ | VTR-CDF |
|---|---|---|---|---|---|
| Gear | Bottom Trawl | | | | |
| Average weight of grid cell overlaps (Std.Dev.) [positive-positive] | 0.026 (0.058) | 0.025 (0.057) | 0.135 (0.218) | 0.025 (0.074) | 0.008 (0.029) |
| Average weights of grid cell non-overlaps (Std.Dev.) [false-positive] | 0.001 (0.002) | 0.001 (0.001) | 0.053 (0.141) | 0.008 (0.017) | 0.001 (0.003) |
| Kolmogorov-Smirnov test (Combined D) | 0.571*** | 0.572*** | 0.448*** | 0.304*** | 0.422*** |
| Percentage of hauls not predicted [positive-false] | 0.65 | 0.68 | 0.35 | 7.90 | 3.41 |

# Summary

- Proposed probability models showed best prediction quality by:
  - Disentangling VMS fishing from non-fishing polls
  - Predicting fishing location if a kernel smoothing is conducted
  - Without adding computational burden to the analysis
  - *BUT: decision between GAM and Logit depending on spatial bias*
- Logbook data prediction quality can be increased if areal extent is predicted instead of only aggregate the data (VTR-CDF)
- If speed rules are applied, constant updating necessary
- False-positive rate mainly in the prediction tail – careful with the parameters!

# Thanks for your attention!

- Question?
- Comments?
- Remarks?

angela.muench@noaa.gov or geret.depiper@noaa.gov