Ensemble Models of Seabirds Abundance at-Sea

Martin Renner

John Piatt, Gary Drew, Kathy Kuletz, George Hunt, and all contributors to the NPPSD

NPPSD North Pacific Pelagic Seabird Database

USGS science for a changing world

NSF

PICES

Goals

Raw data:

densities in a 0.3 km × 3 km sample at a certain day and time

→

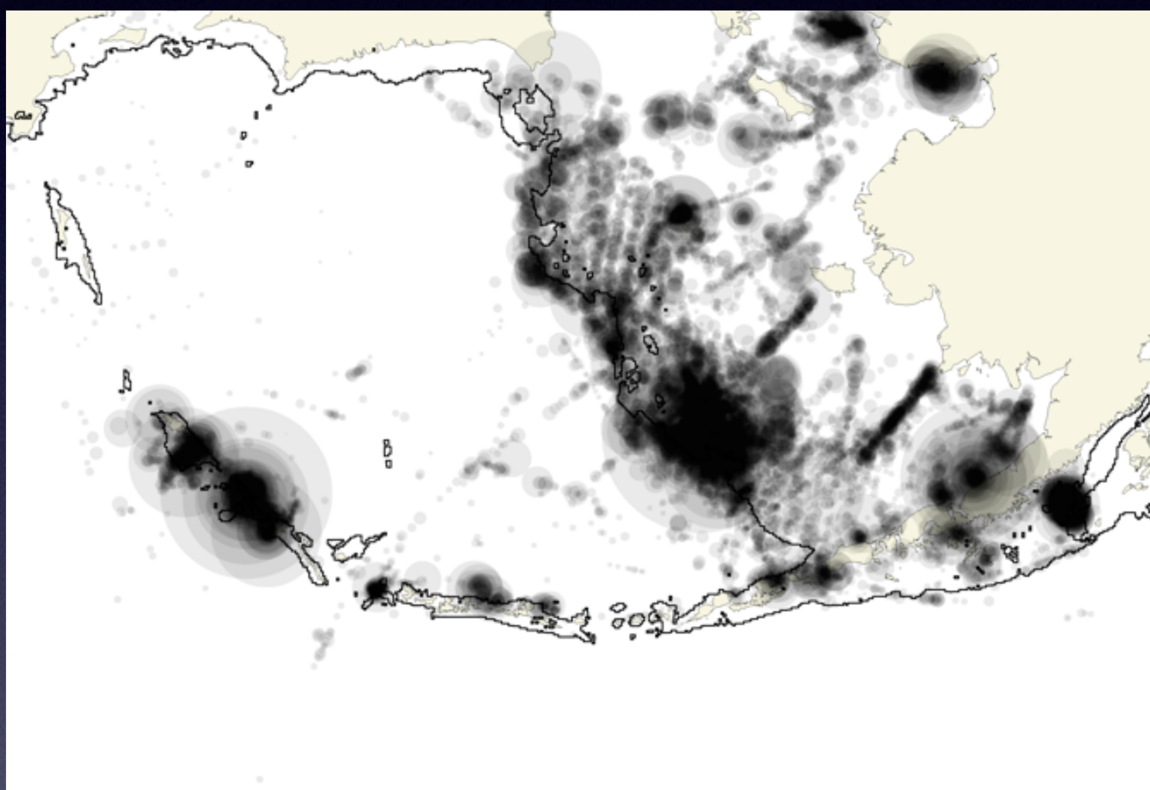
Expected means:

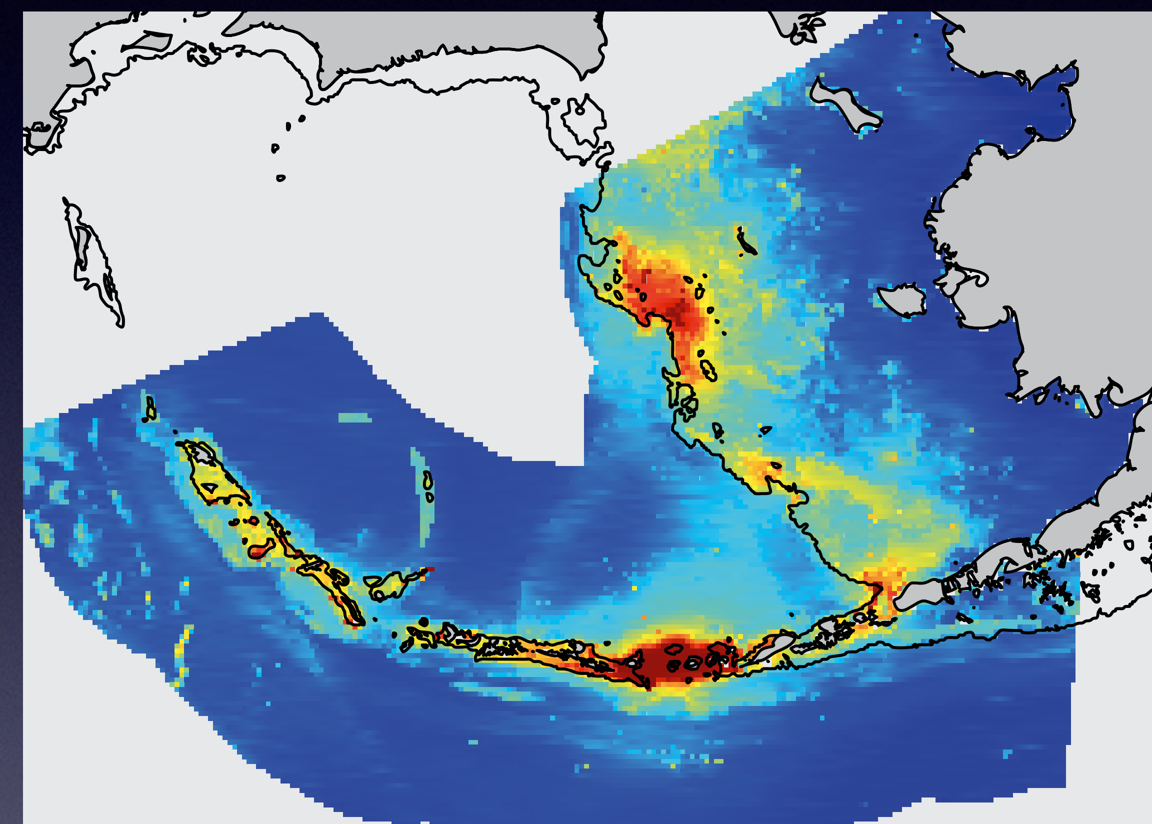
continuous densities at any given location and specific/average time

impossible to *know*

truth = impractical to know

Goals





+ temporal dimension

“All models are wrong, but some are useful”

George Box

Outline

- Why Ensemble models
- Classification of models
- Compare model performance
  - example species
  - summary of 24 seabird species
- Why not Ensemble models
- Future

Guessing on decisions

- Modeling algorithms, Variable selection
- Linear or curvilinear, error-distribution
- Spatial scale (grid size)
- Degrees of interactions
- Deal with spatial autocorrelations
- Continuous vs. categorical (e.g. season)

Is there a best method we should all use?

Work flow

settings

fit model

examine prediction

optimize settings
to reduce artifacts

prohibitive for
large suite of species

Importance

- Raw data is of limited use
- Population trends
- MPAs, IBA, offshore wind energy, oil spills, hotspots, energy consumption, fisheries, shipping, etc.
- Averaging over a grid is a (simple) model

Ship or aerial surveys

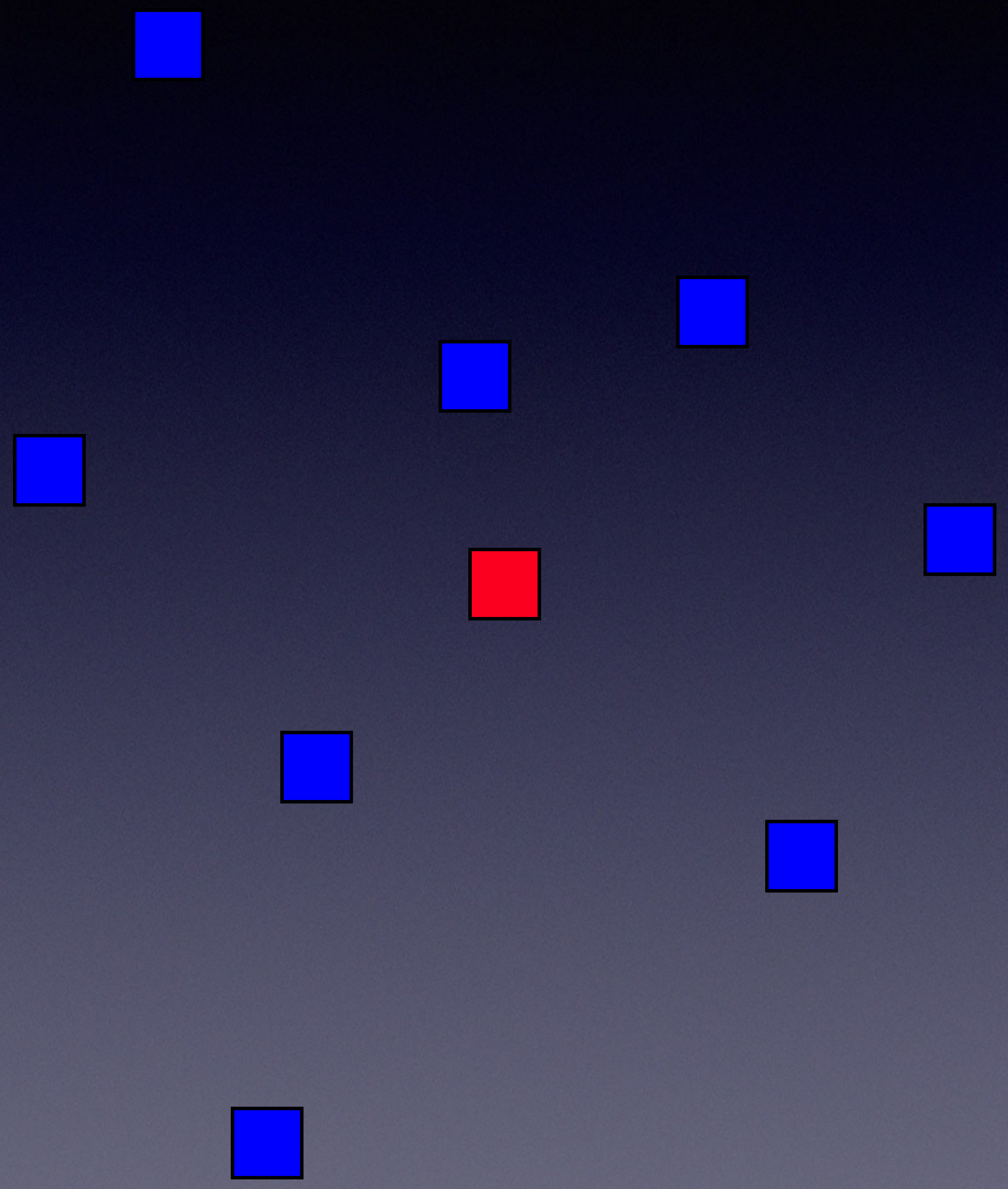
- does not mix well with tracking data
- abundance, not presence/absence

Meet the Candidates

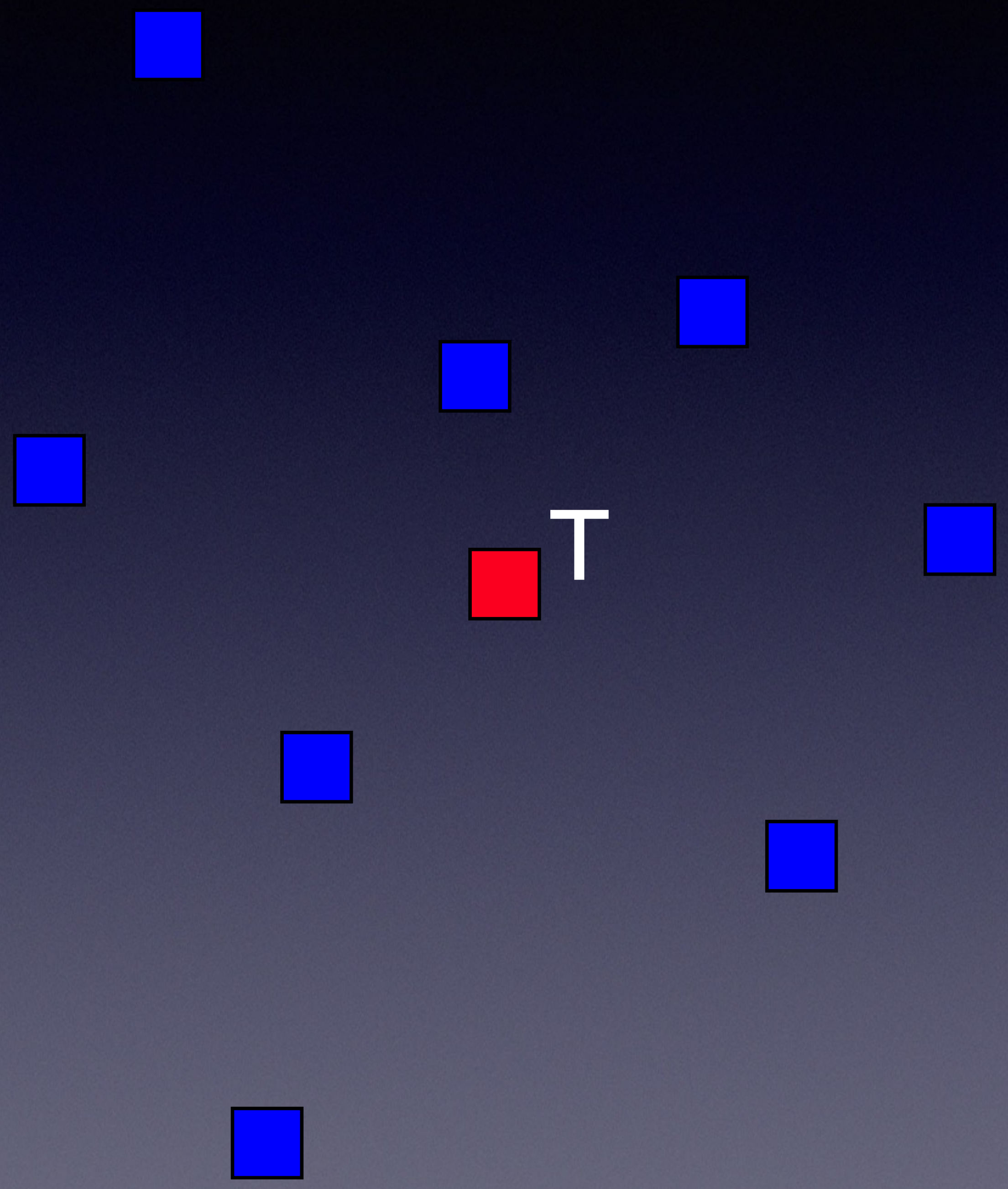
grid-based	classic	data mining	spatial interpolation
GLMM	GLM	MARS	ordinary kriging
	GAM shrinkage	Random Forest	universal kriging

Ensemble: weighted mean of GAM, MARS, RF

Ensemble Model



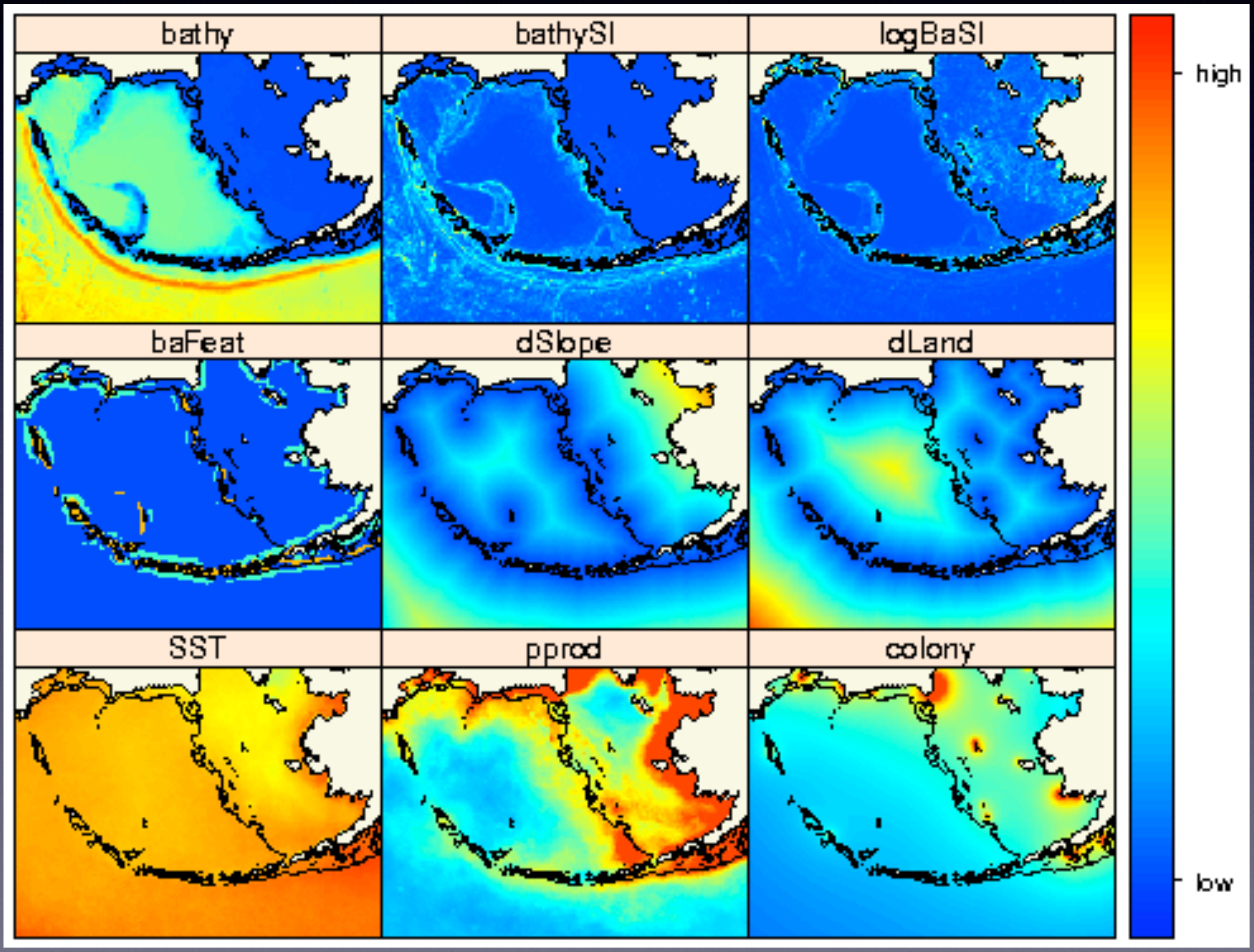
Ensemble Model



Model Performance

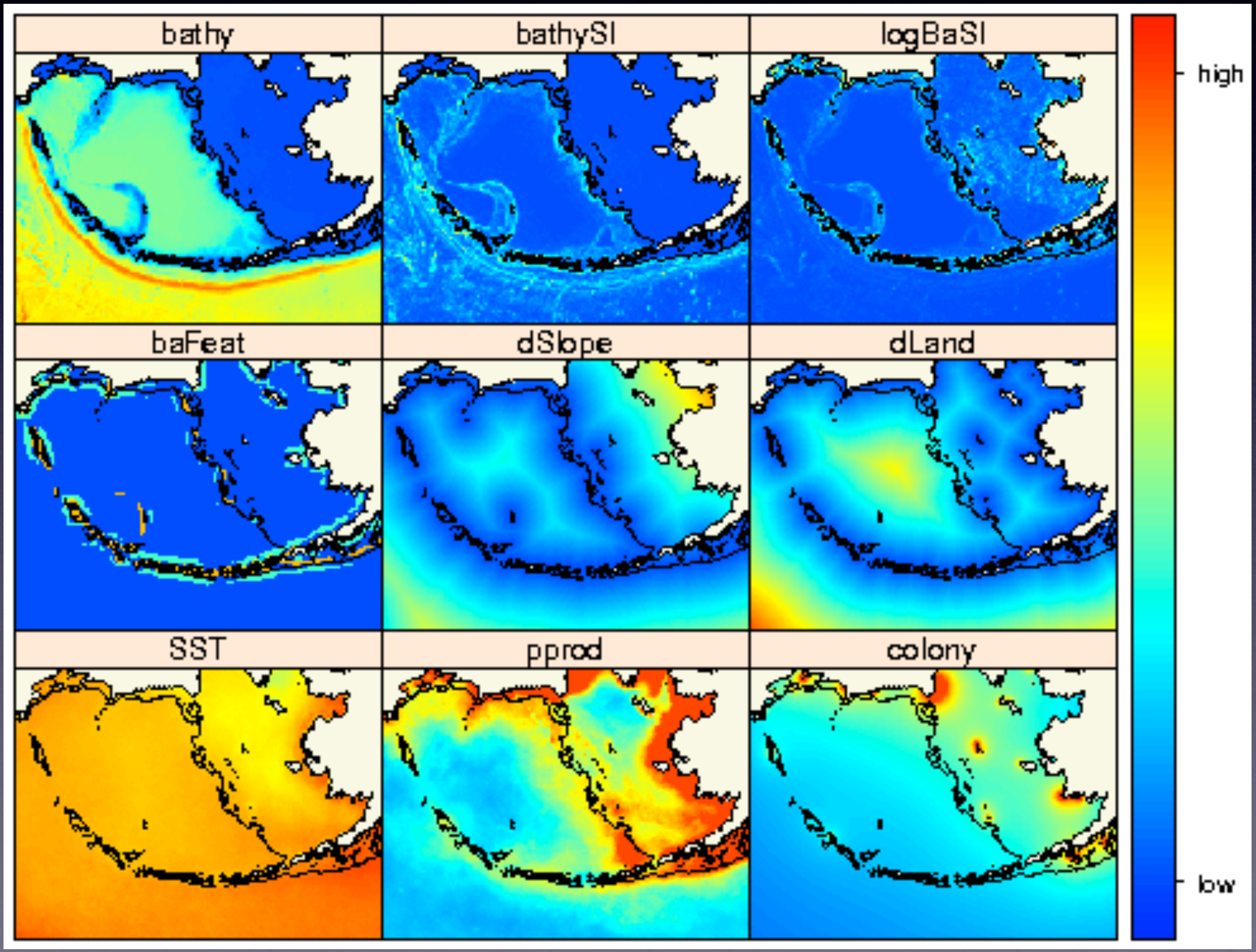
- external 10-fold cross-validation
- performance criterion: RMSE

Environmental variables



climatologies

Environmental variables

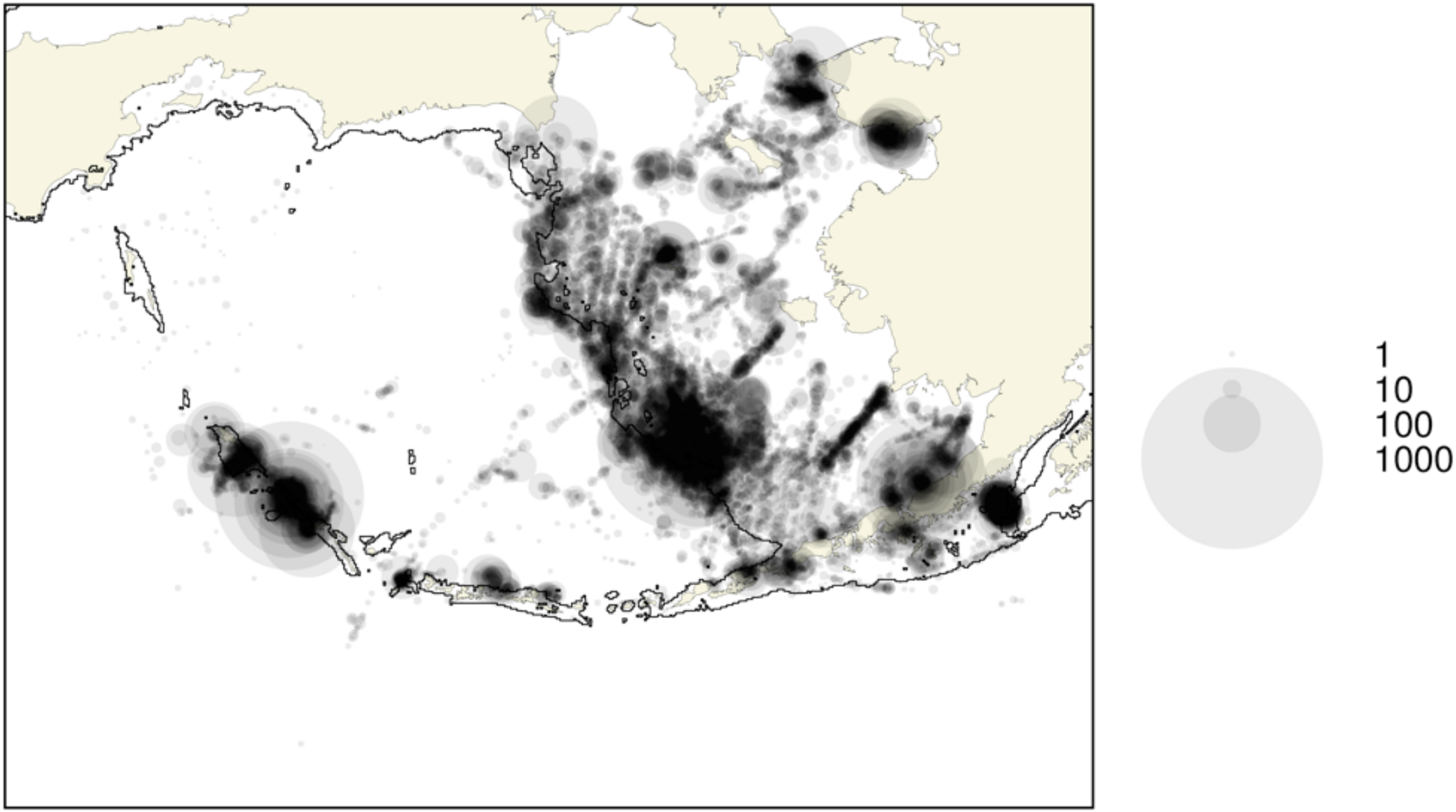


climatologies

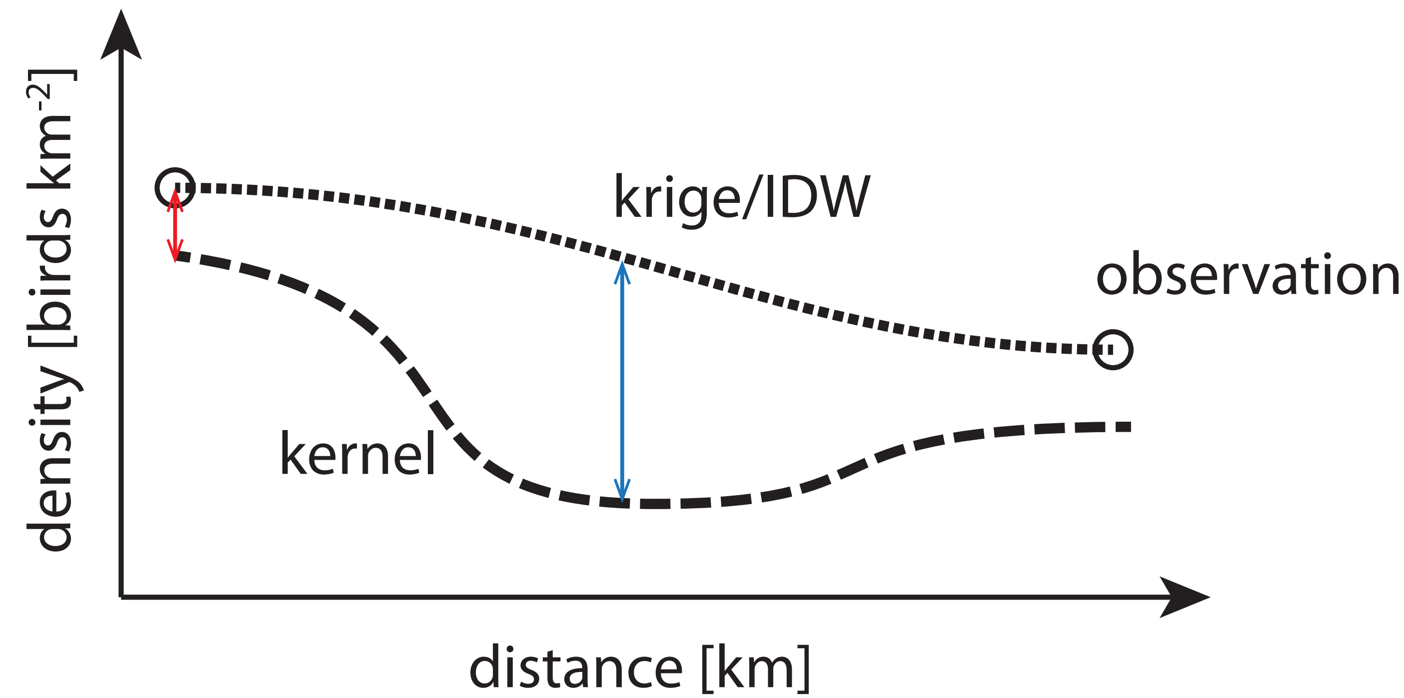
distance scaled by colony size

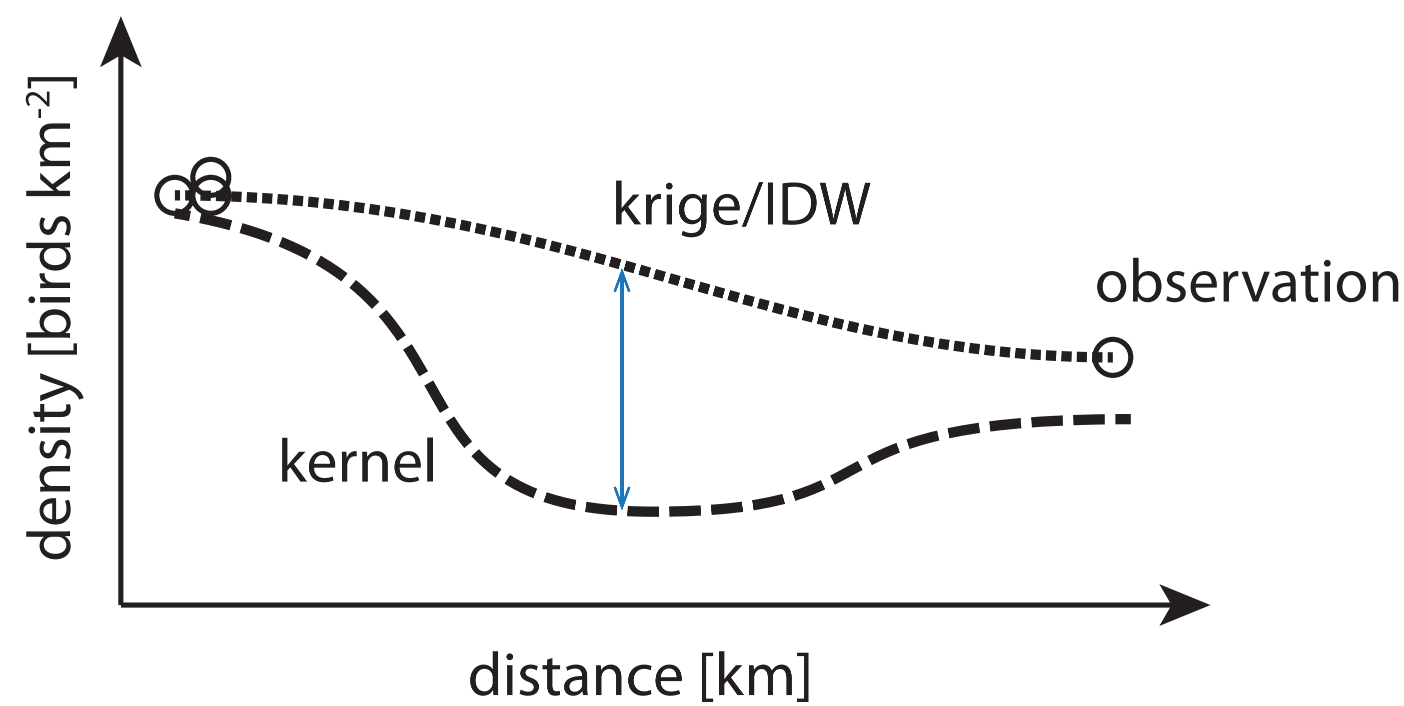
$$C_l = \sum_{i=1}^{i=n} \frac{s_i}{d_{i,l}}$$

Raw data - Black-legged Kittiwake

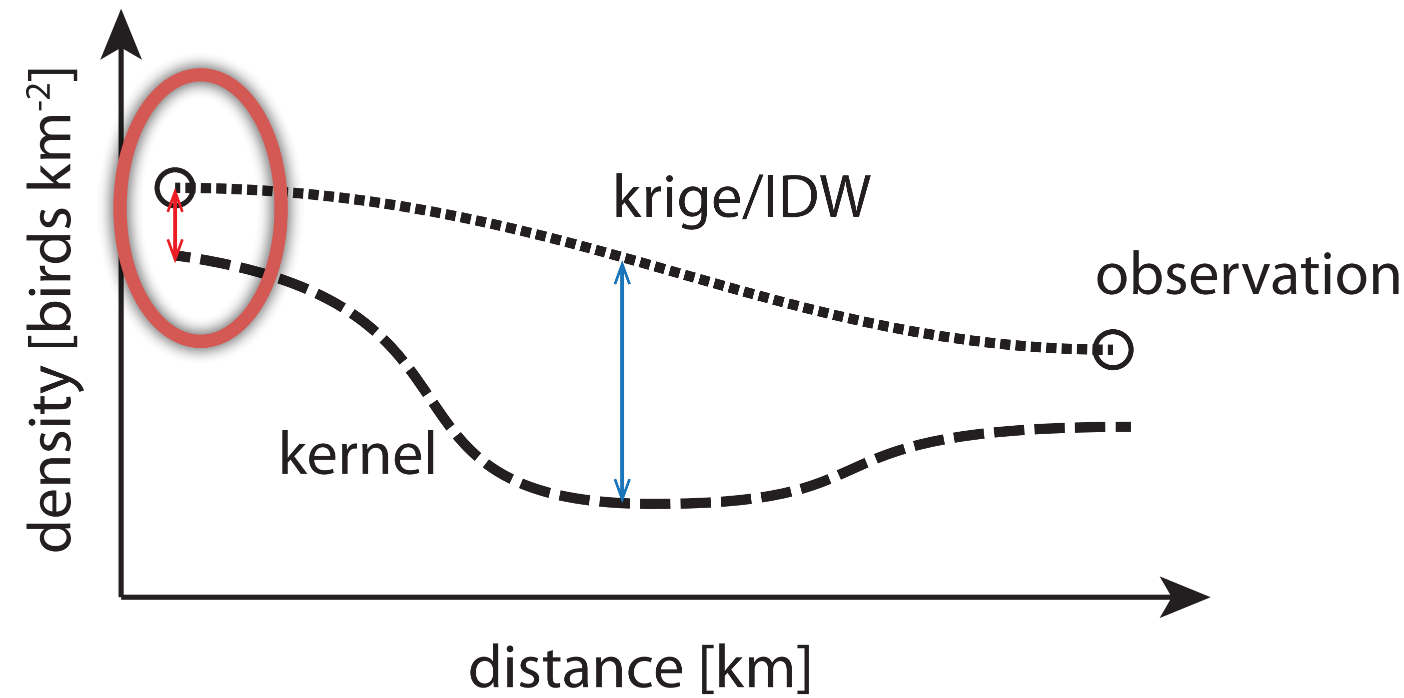


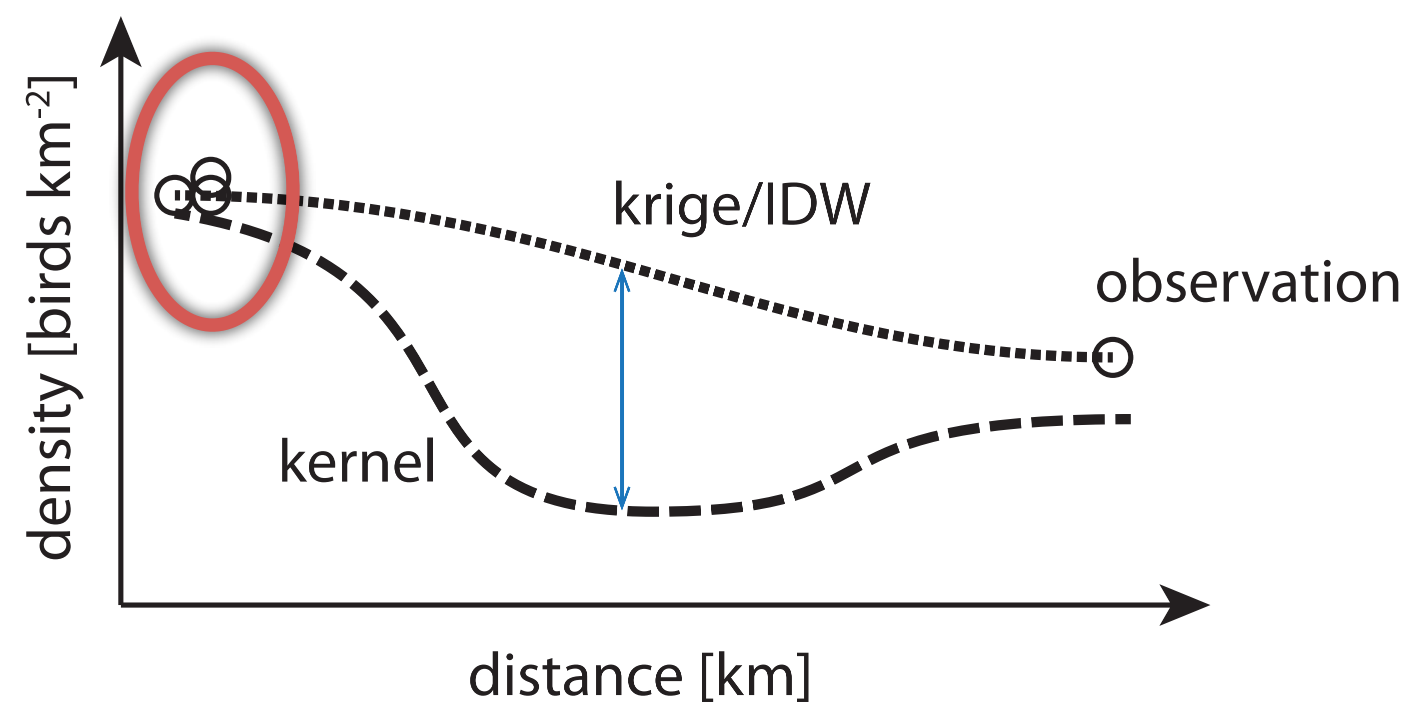
Kernel-densities





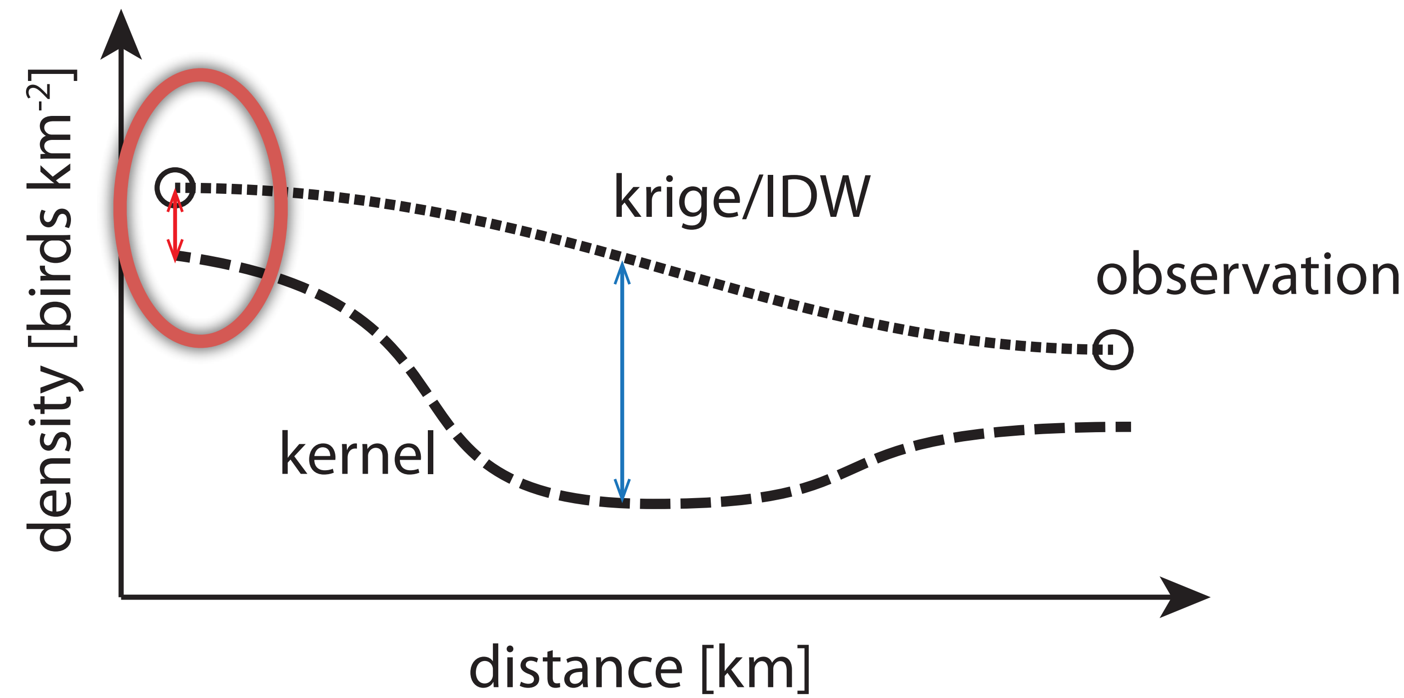
Kernel-densities

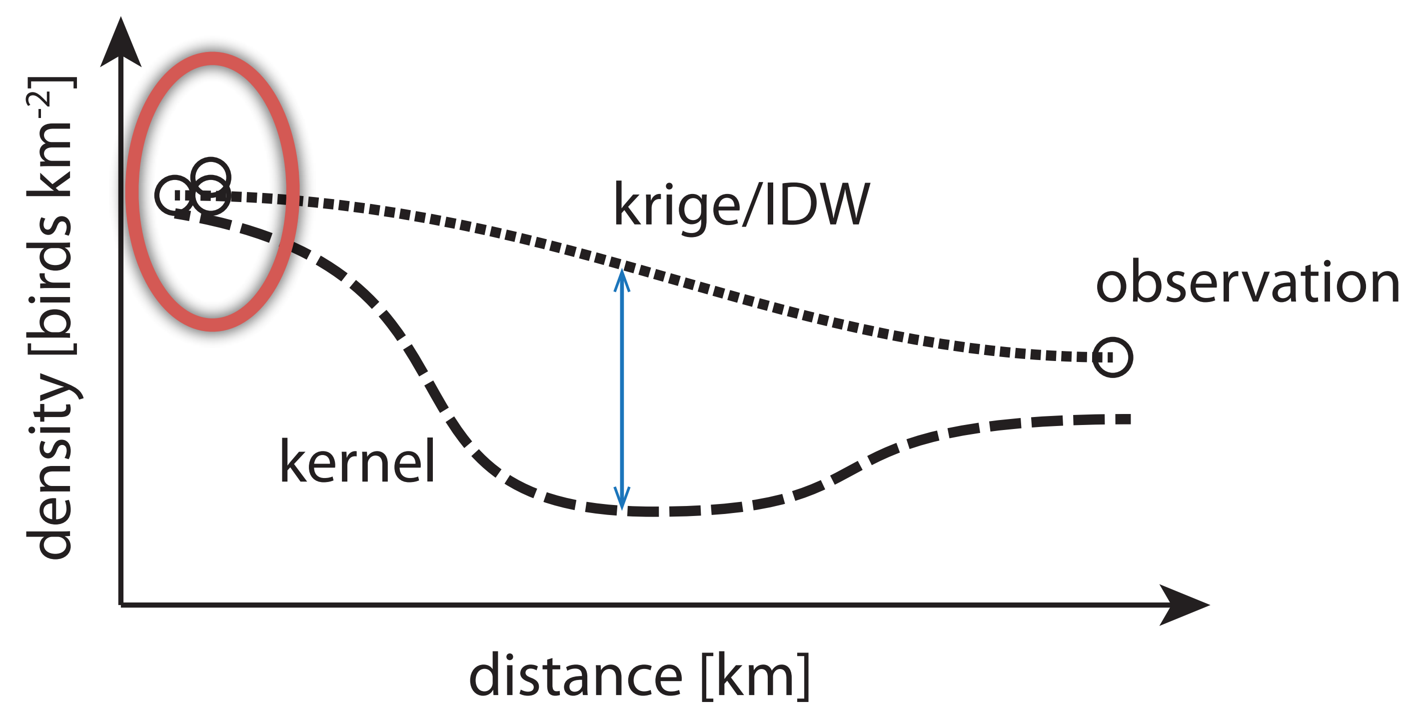


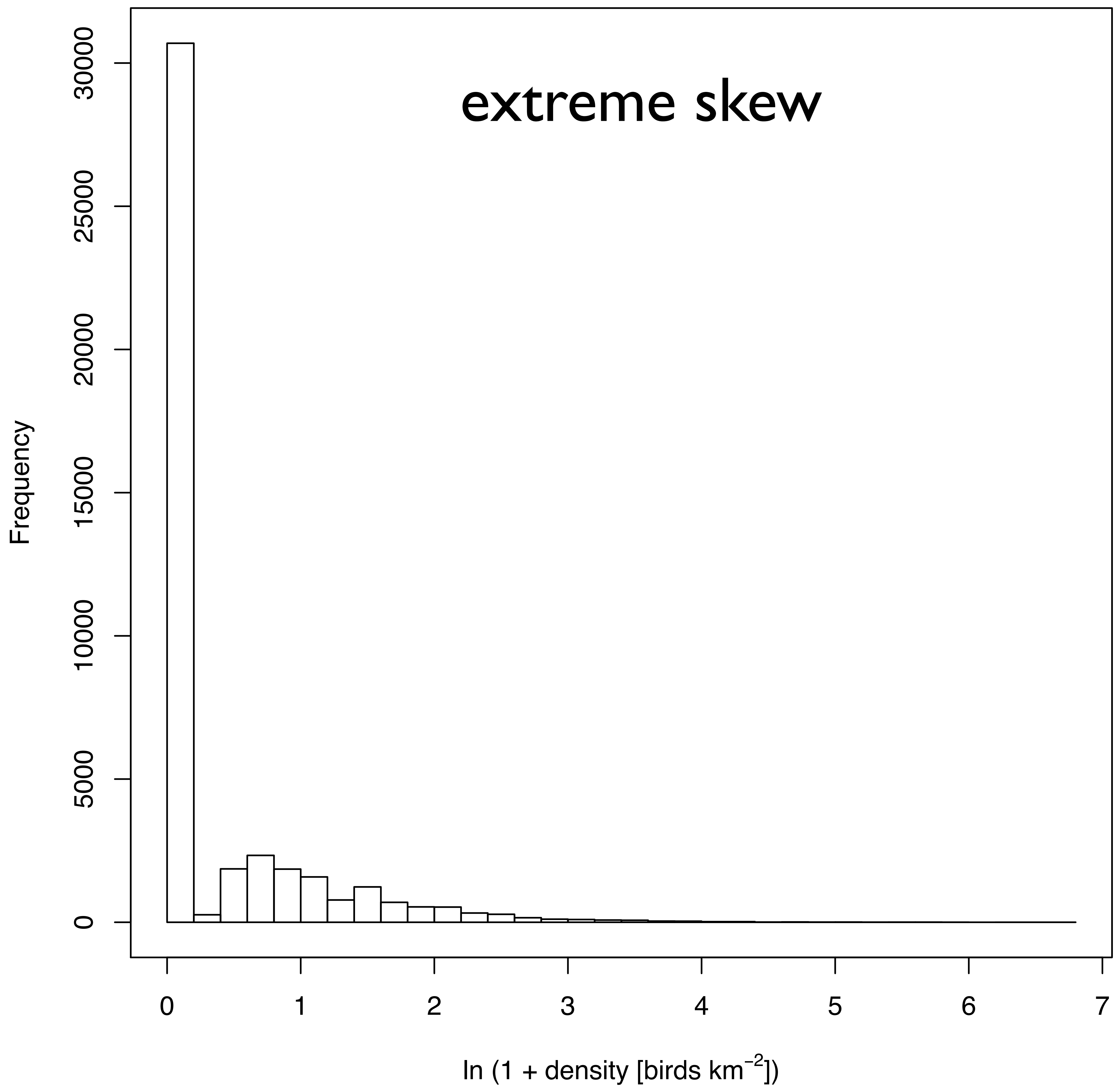


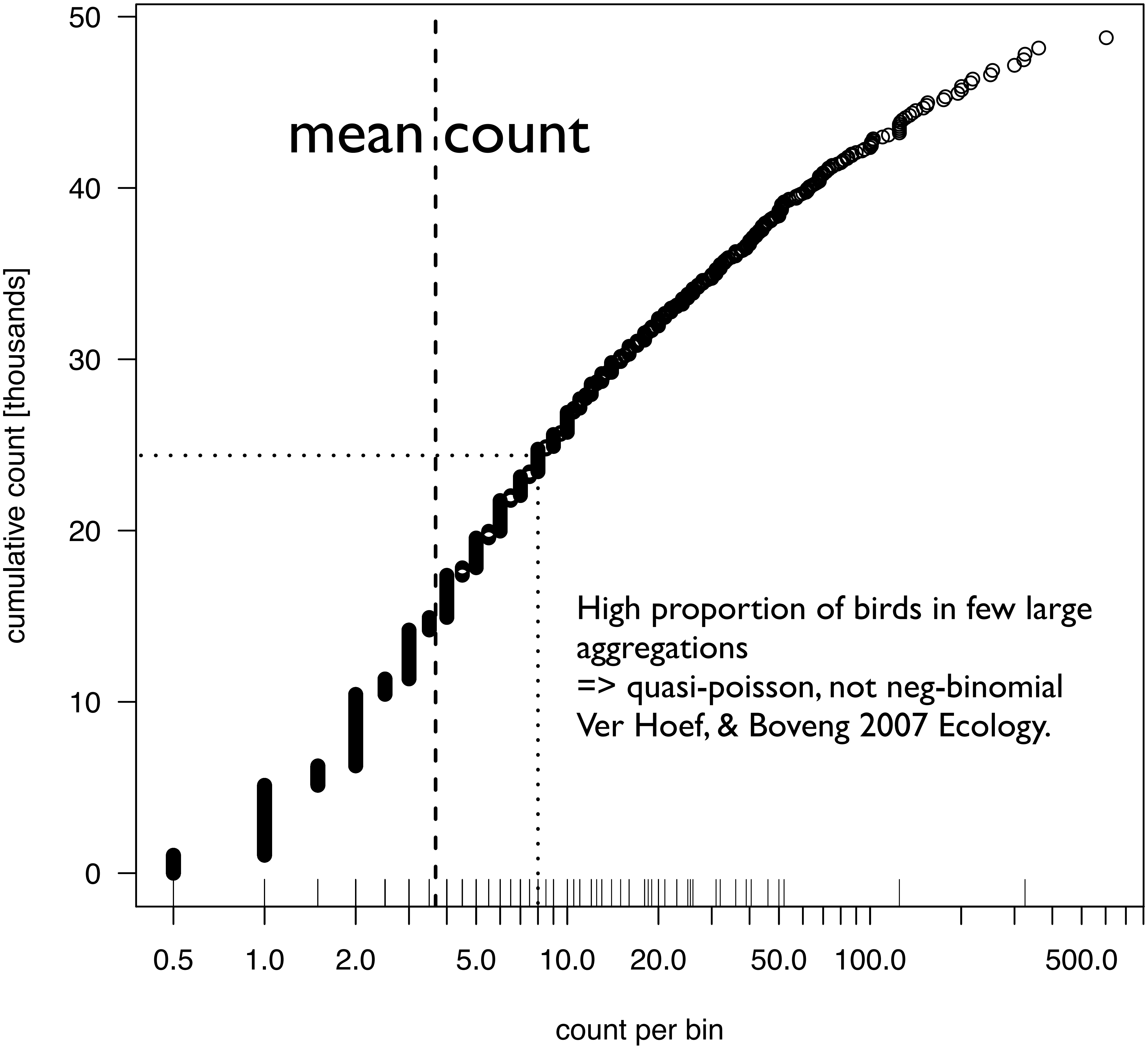
Kernel-densities

biased
effort-dependent
poor interpolation





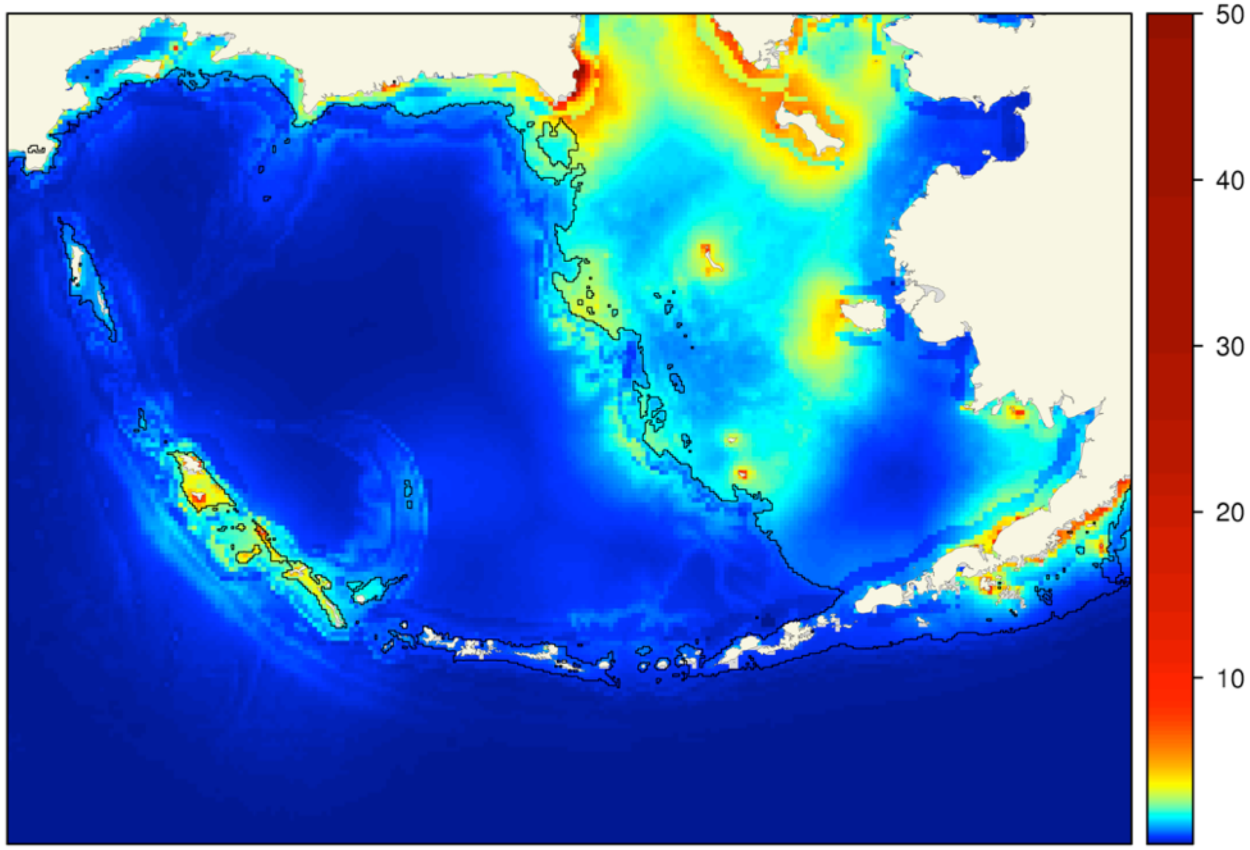

extreme skew
Frequency
0
5000
10000
15000
20000
25000
30000
0
1
2
3
4
5
6
7
ln (1 + density [birds km−2])



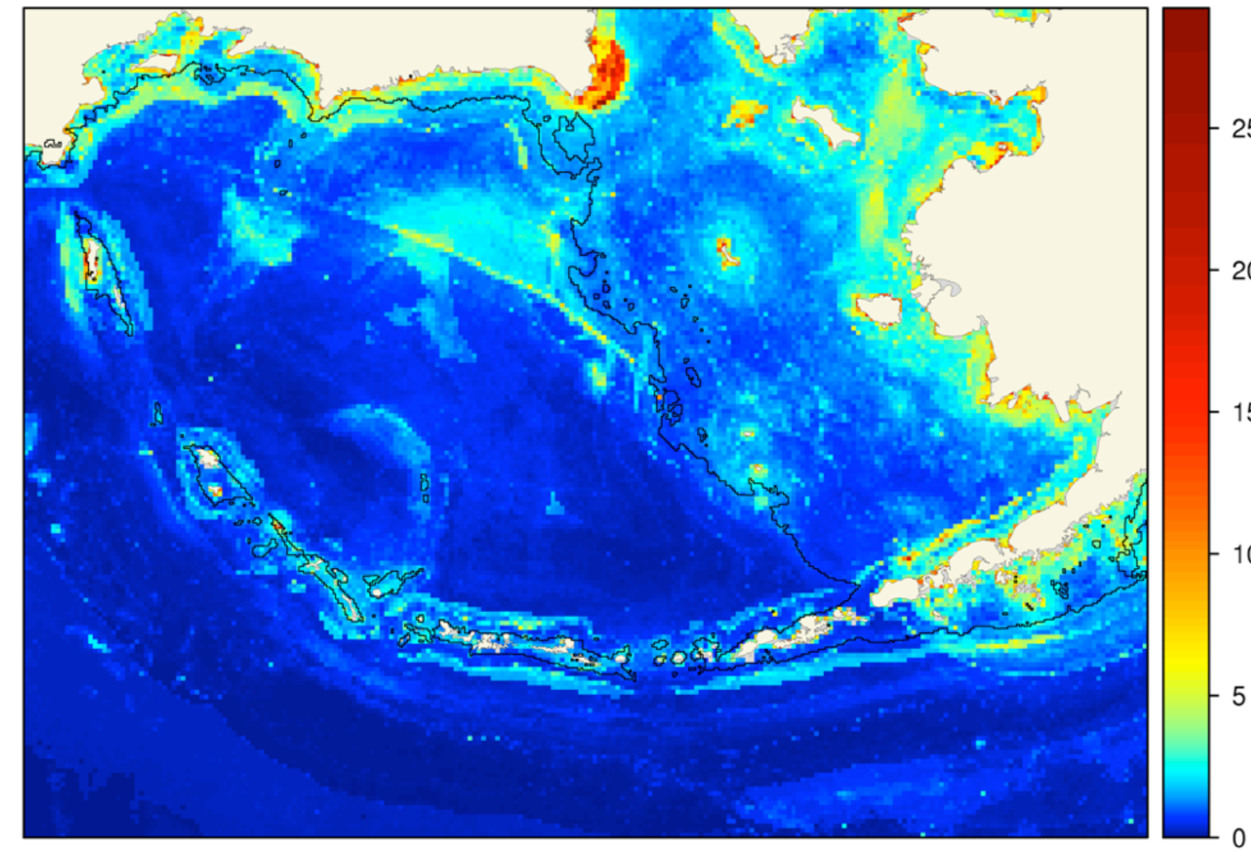
mean count
High proportion of birds in few large aggregations
=> quasi-poisson, not neg-binomial
Ver Hoef, & Boveng 2007 Ecology.
cumulative count [thousands]
0
10
20
30
40
50
0.5
1.0
2.0
5.0
10.0
20.0
50.0
100.0
500.0
count per bin

Black-legged Kittiwake

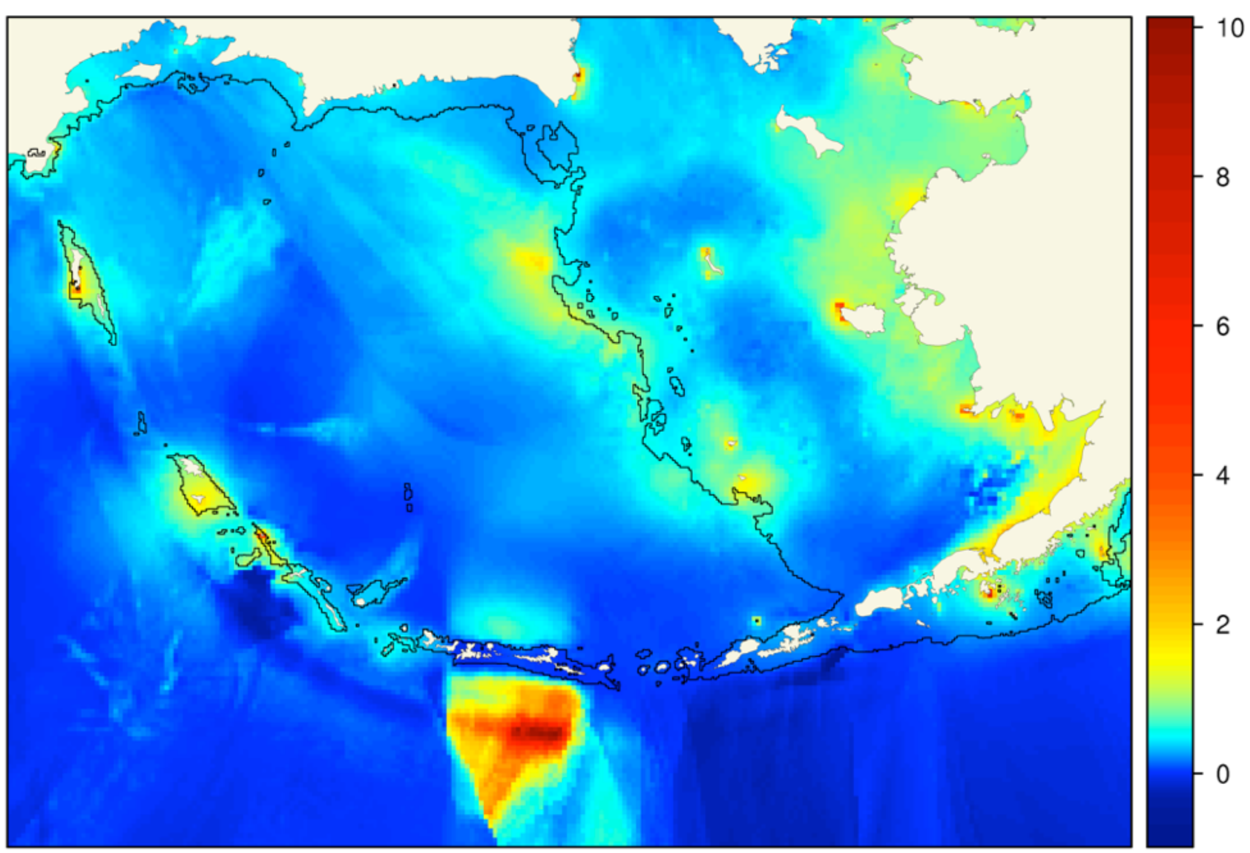
gam2



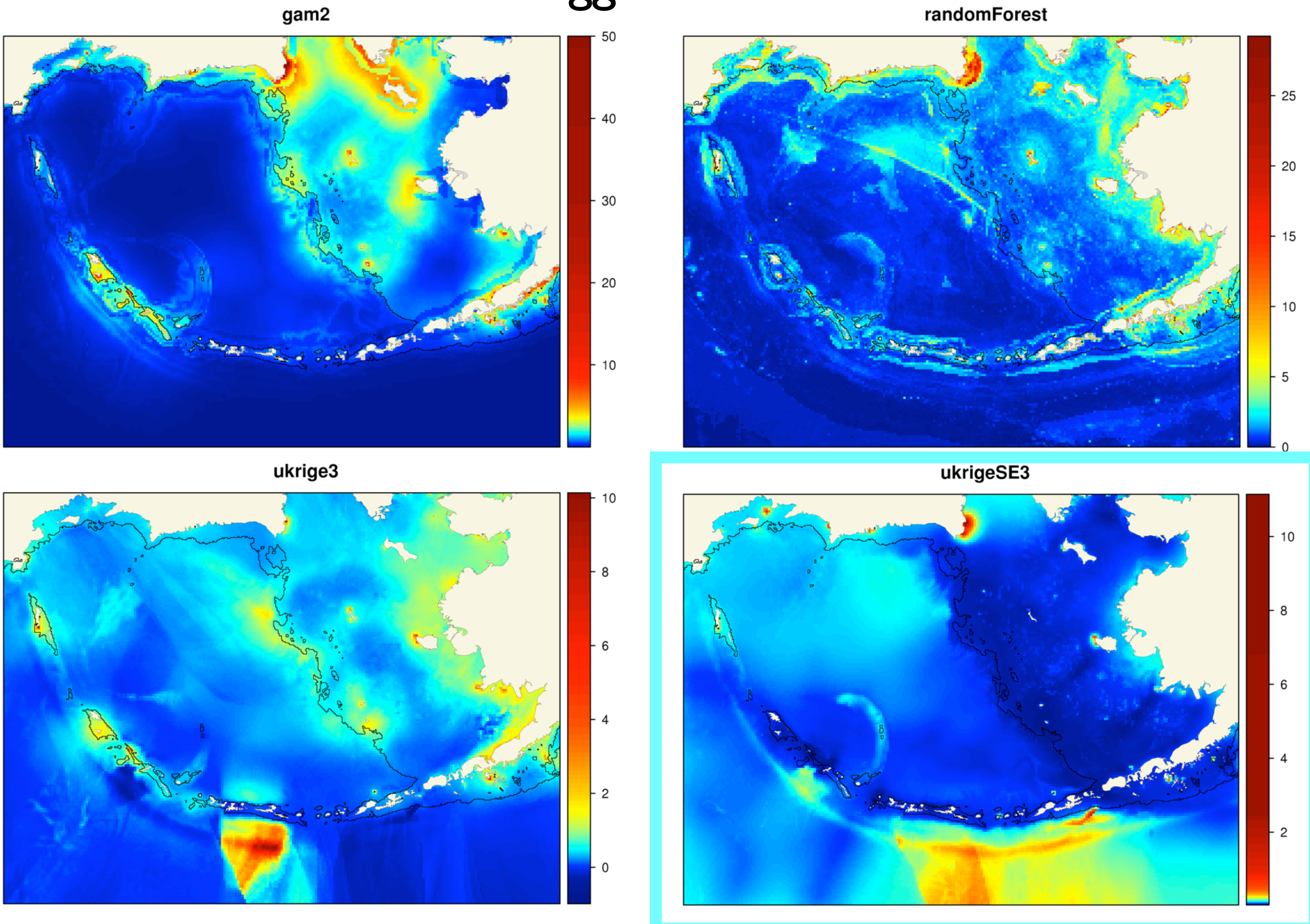
randomForest



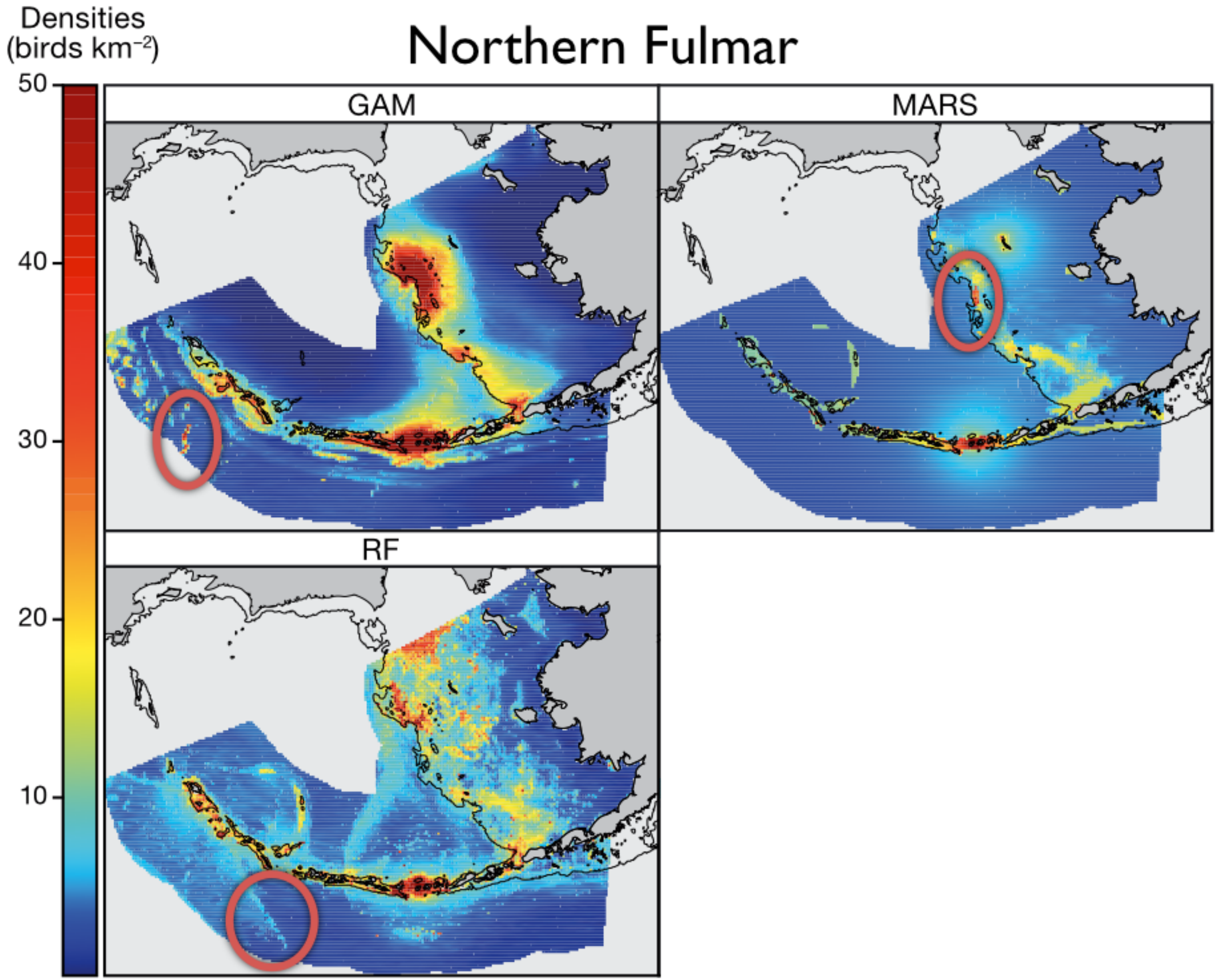
ukrige3



Black-legged Kittiwake



Northern Fulmar



Renner et al 2013 MEPS

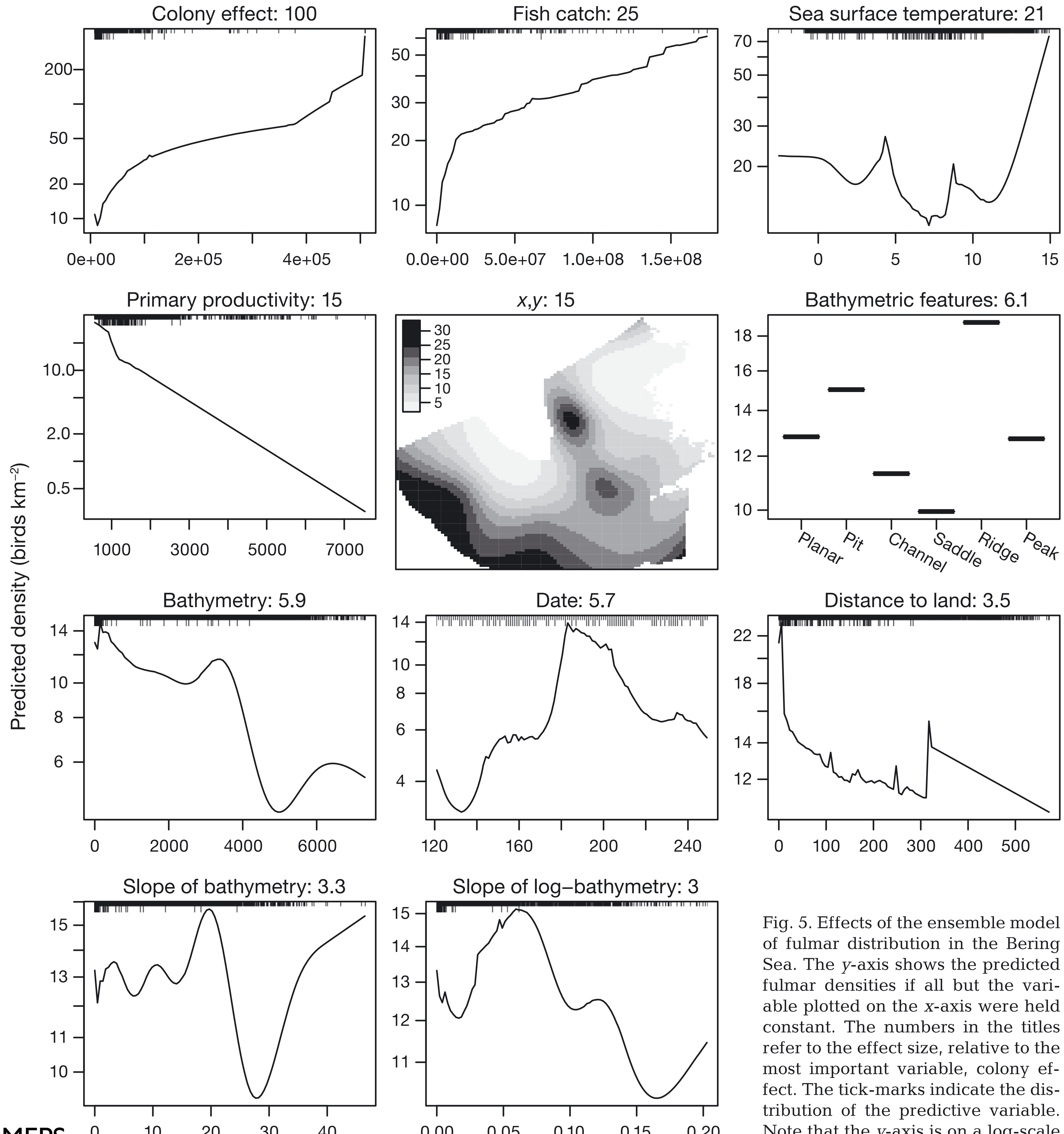
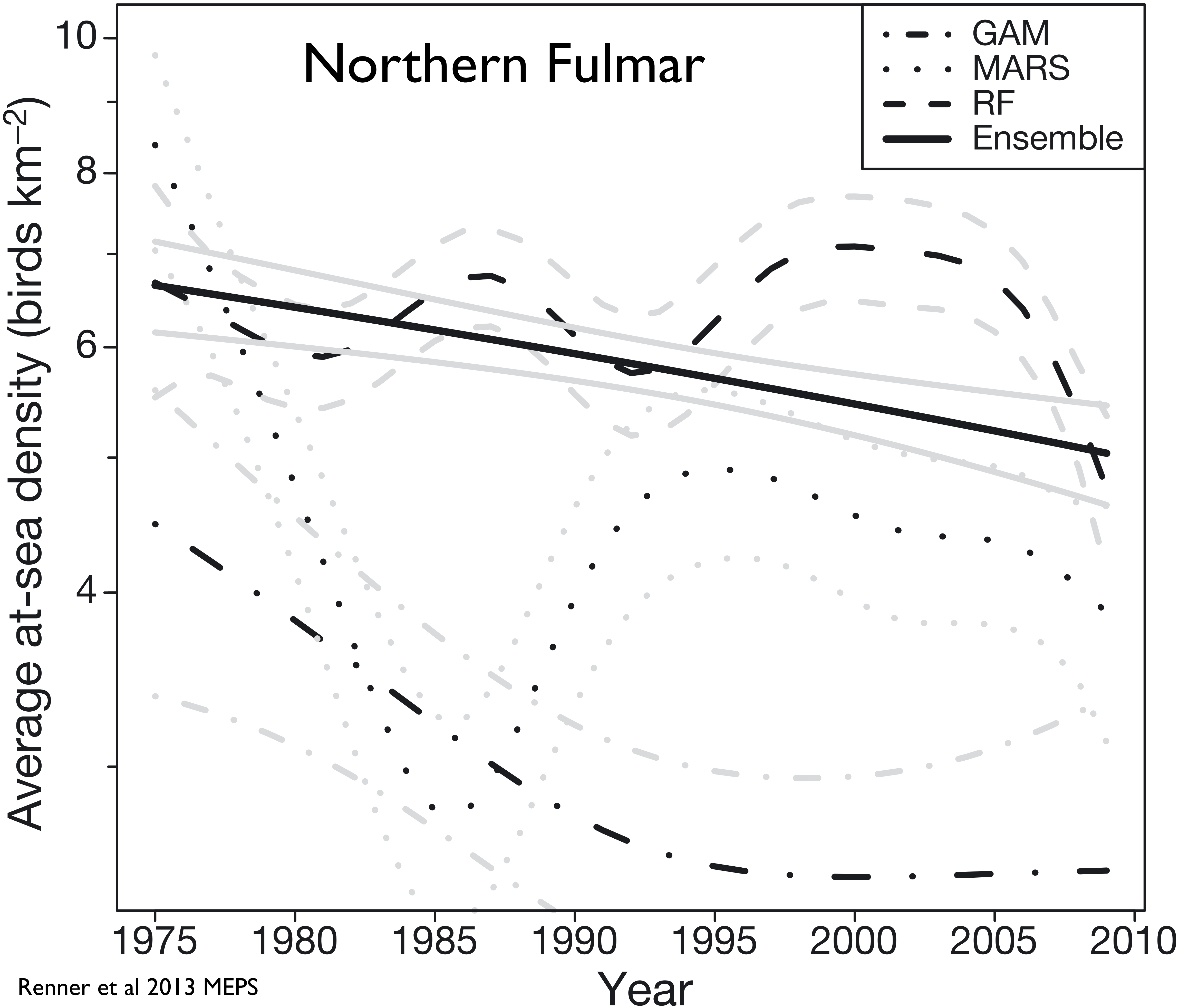


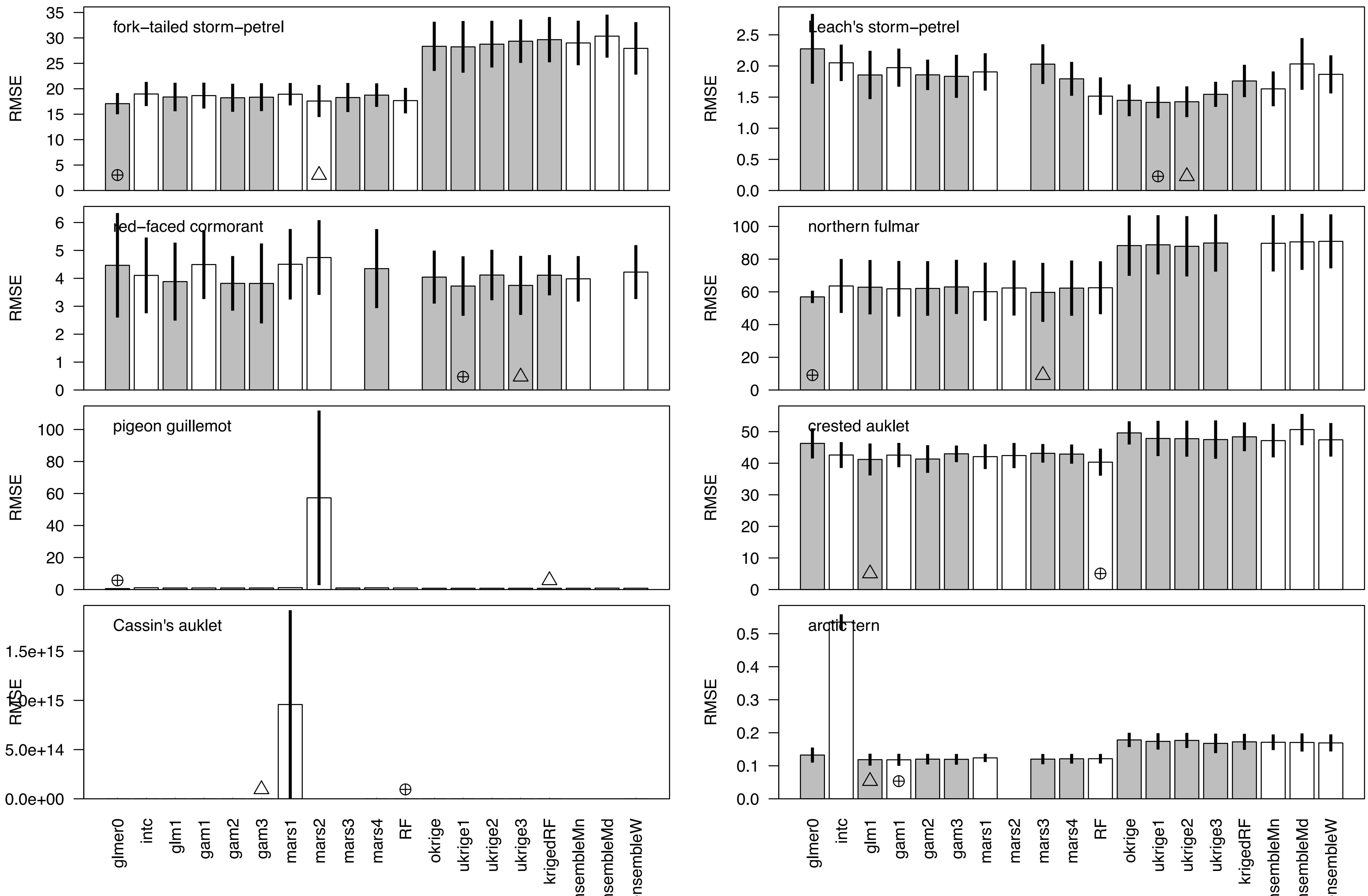
Fig. 5. Effects of the ensemble model of fulmar distribution in the Bering Sea. The *y*-axis shows the predicted fulmar densities if all but the variable plotted on the *x*-axis were held constant. The numbers in the titles refer to the effect size, relative to the most important variable, colony effect. The tick-marks indicate the distribution of the predictive variable. Note that the *y*-axis is on a log-scale

Renner et al 2013 MEPS

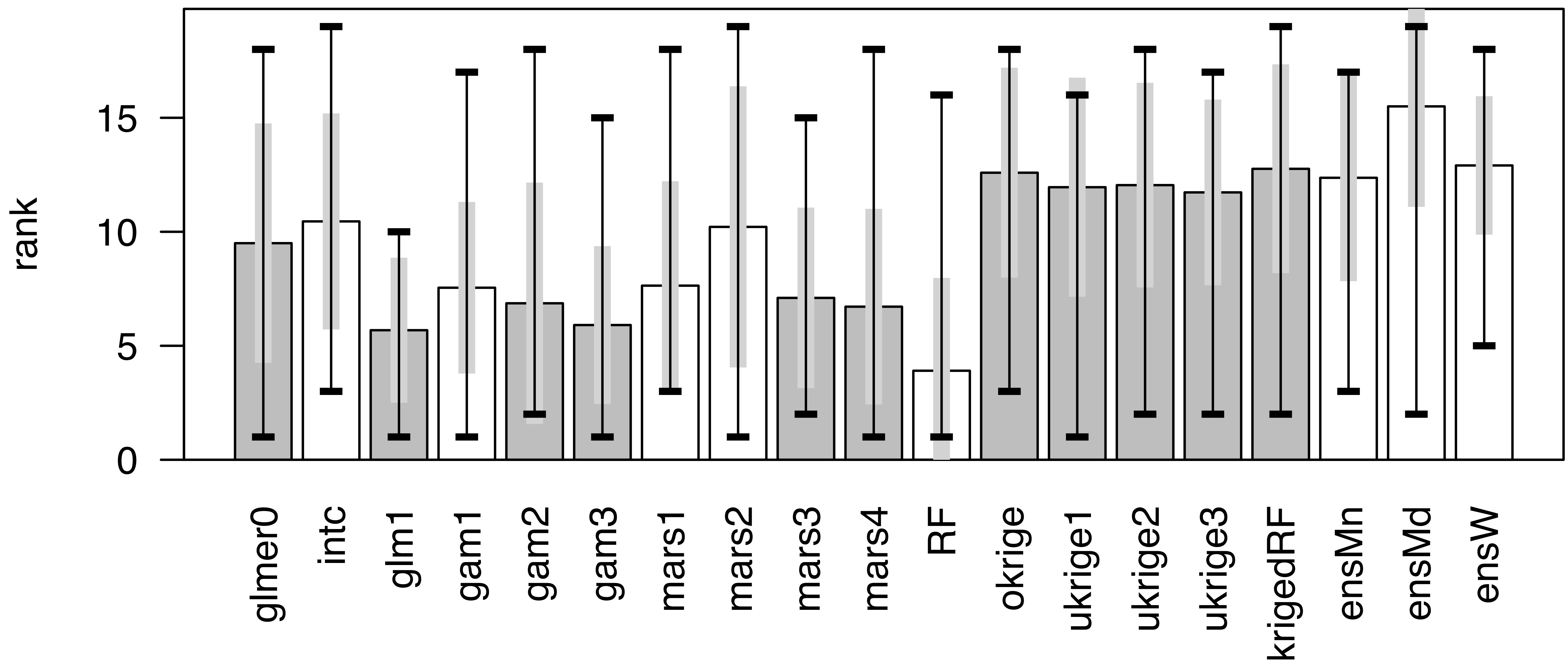


Renner et al 2013 MEPS


24 Breeders
photos: © M. Renner

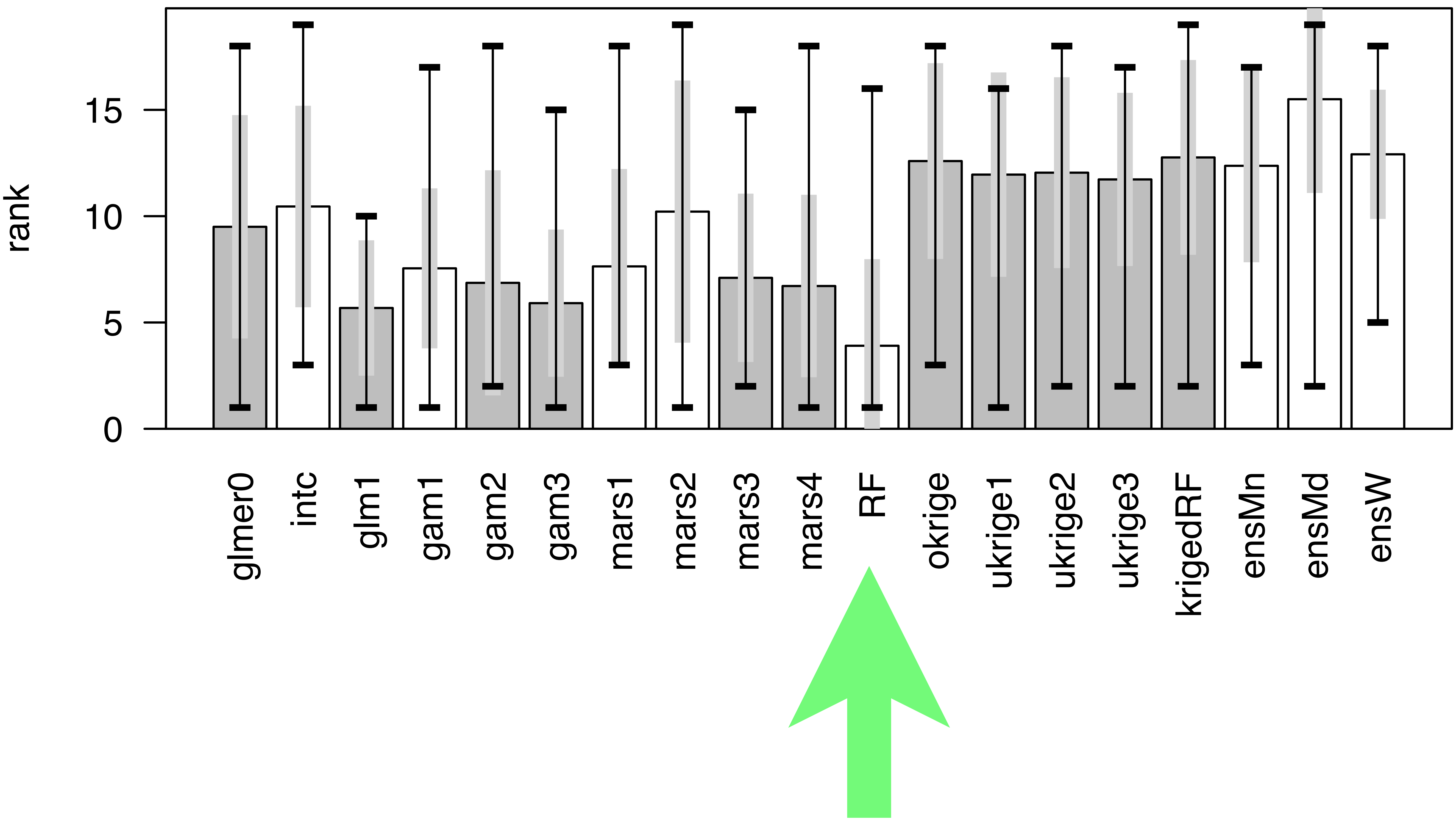

fork−tailed storm−petrel
Leach's storm−petrel
red−faced cormorant
northern fulmar
pigeon guillemot
crested auklet
Cassin's auklet
arctic tern
RMSE
glmer0
intc
glm1
gam1
gam2
gam3
mars1
mars2
mars3
mars4
RF
okrige
ukrige1
ukrige2
ukrige3
krigedRF

Mean ranks



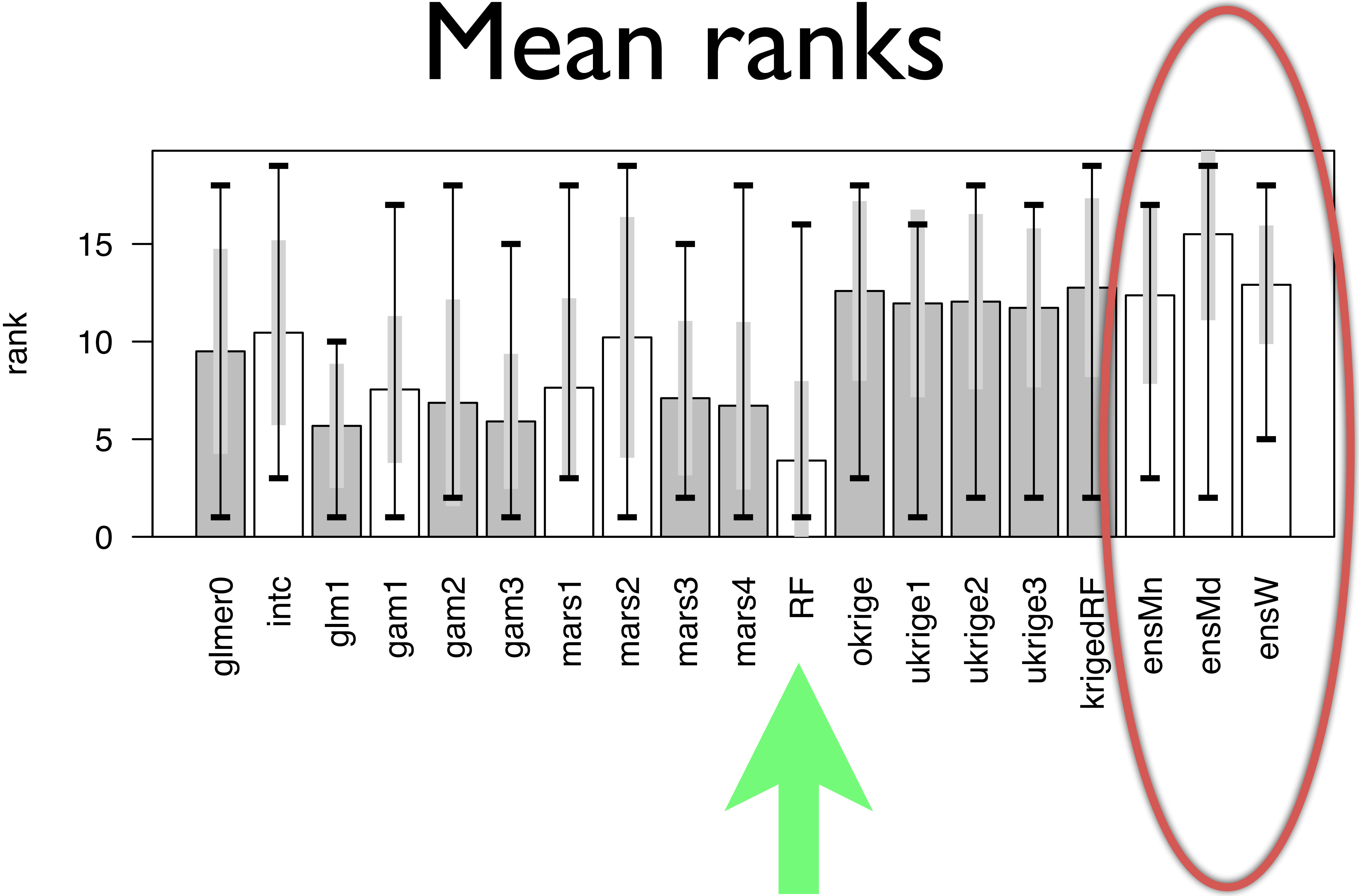
rank
0
5
10
15
glmer0
intc
glm1
gam1
gam2
gam3
mars1
mars2
mars3
mars4
RF
okrige
ukrige1
ukrige2
ukrige3
krigedRF
ensMn
ensMd
ensW

Mean ranks



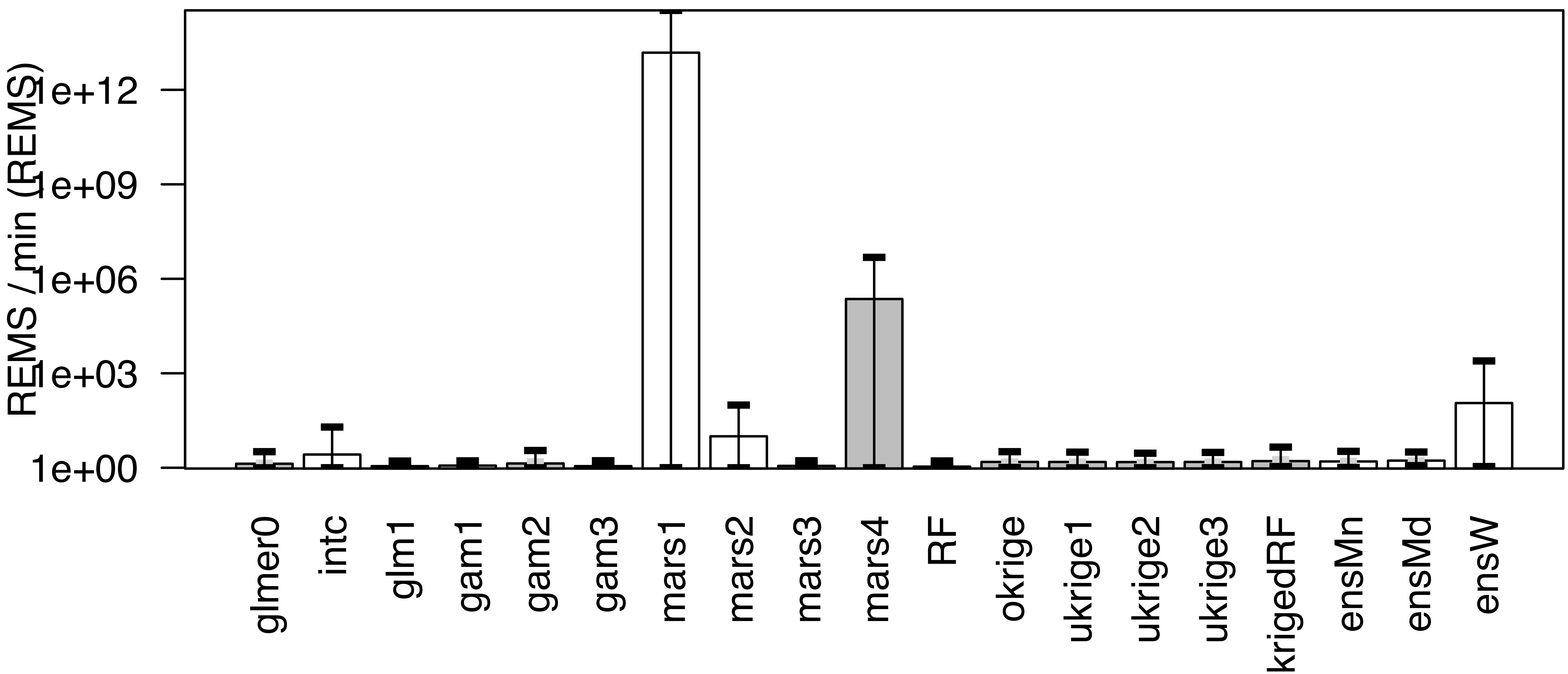
rank
0
5
10
15
glmer0
intc
glm1
gam1
gam2
gam3
mars1
mars2
mars3
mars4
RF
okrige
ukrige1
ukrige2
ukrige3
krigedRF
ensMn
ensMd
ensW

Mean ranks

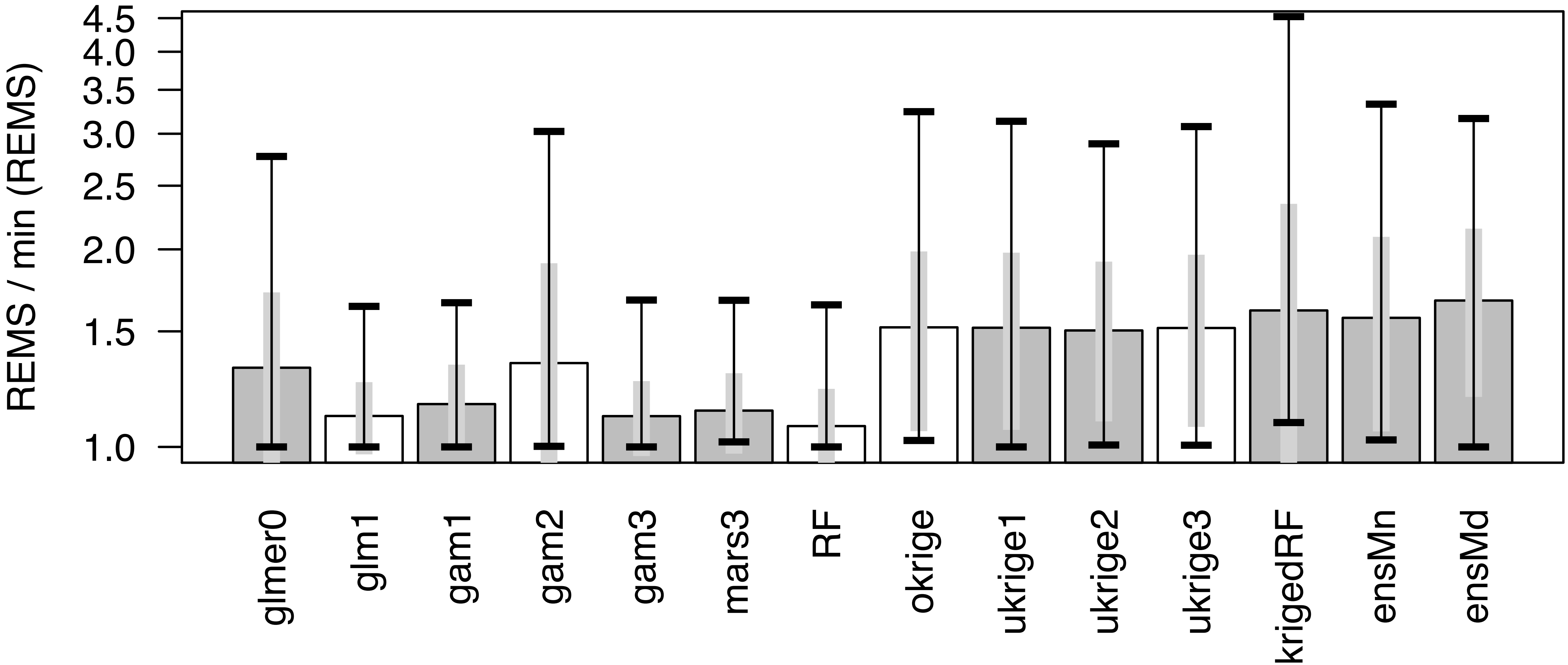


rank
0
5
10
15
glmer0
intc
glm1
gam1
gam2
gam3
mars1
mars2
mars3
mars4
RF
okrige
ukrige1
ukrige2
ukrige3
krigedRF
ensMn
ensMd
ensW

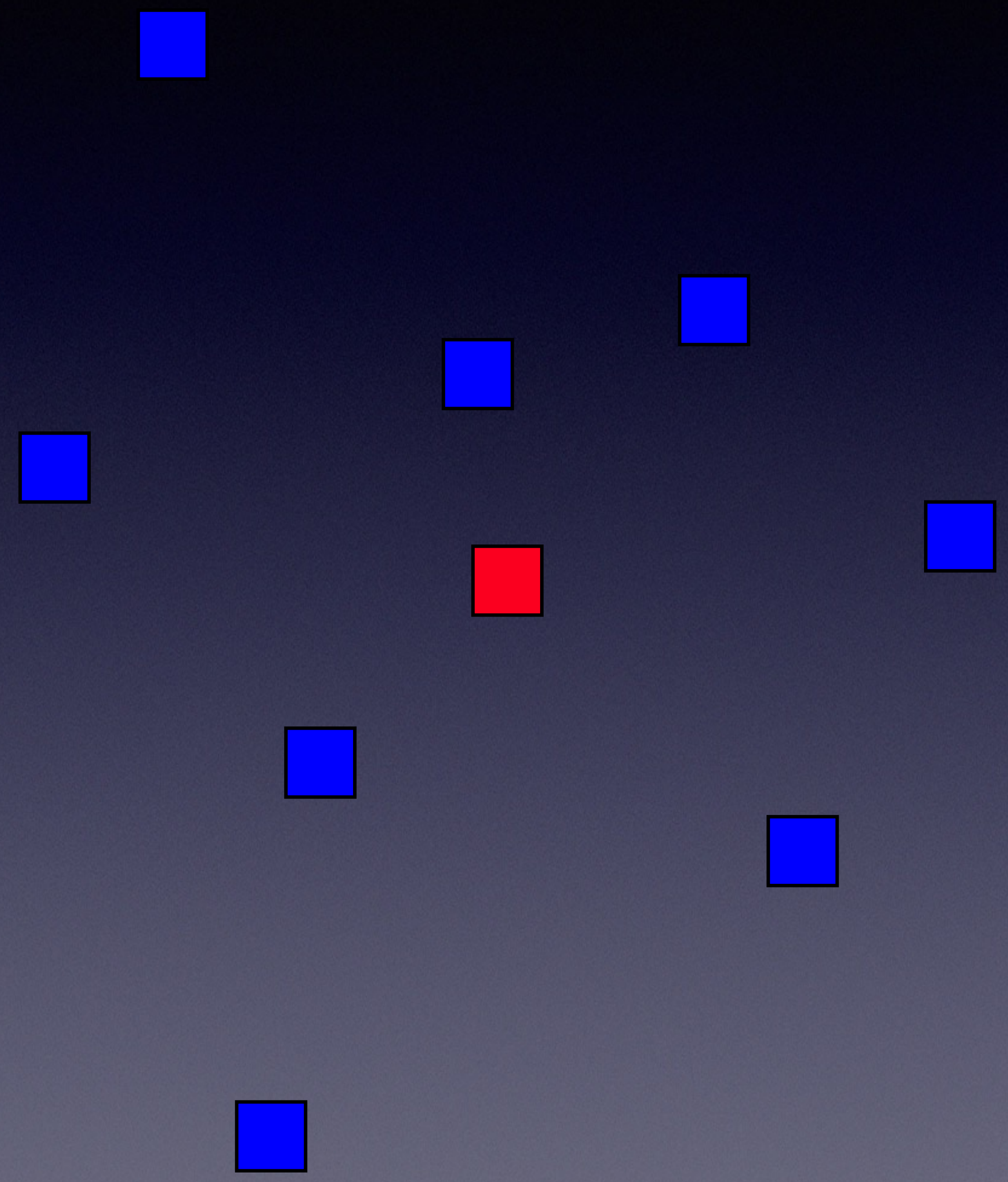
Mean REMS

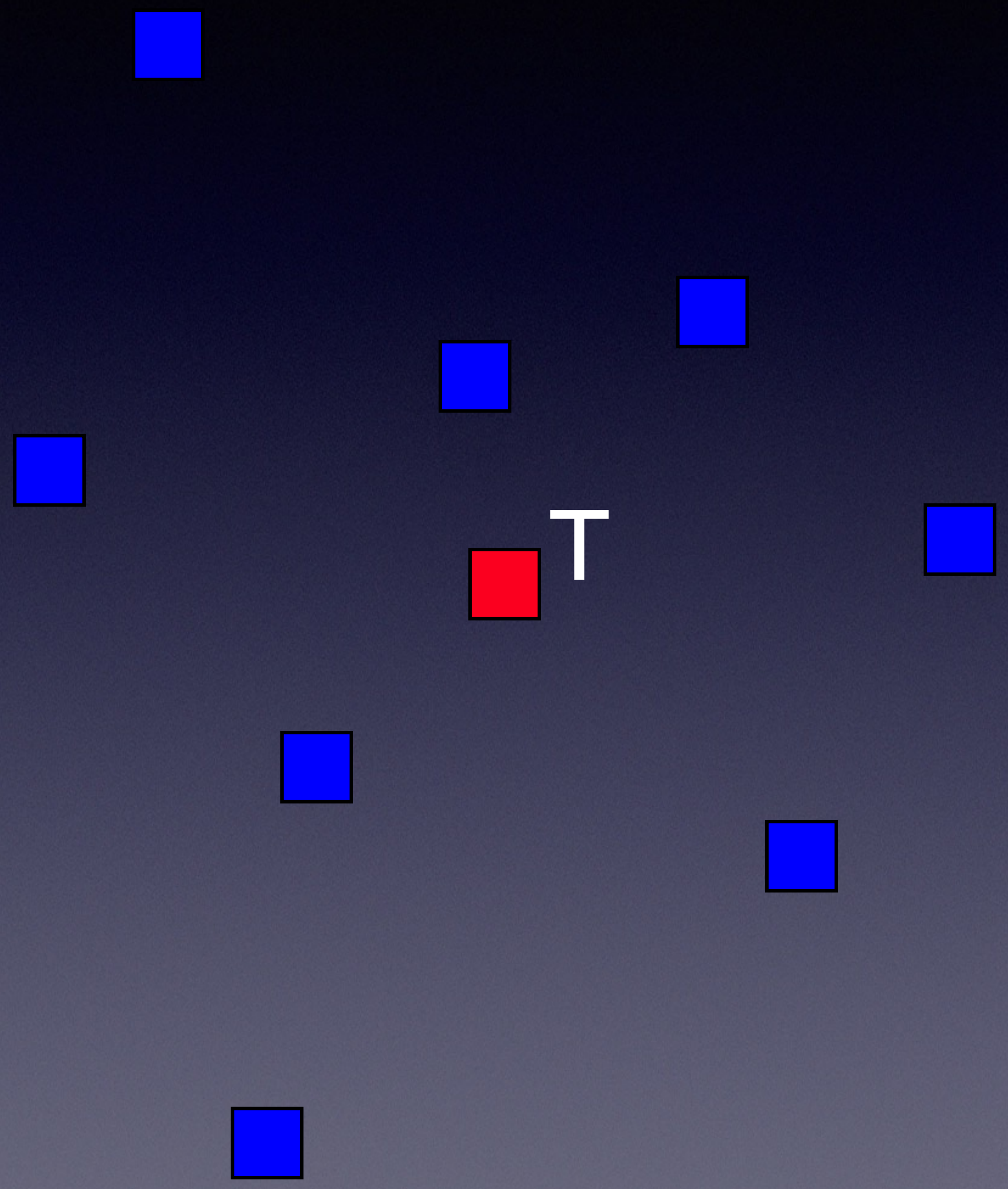

REMS / min (REMS)
1e+12
1e+09
1e+06
1e+03
1e+00
glmer0
intc
glm1
gam1
gam2
gam3
mars1
mars2
mars3
mars4
RF
okrige
ukrige1
ukrige2
ukrige3
krigedRF
ensMn
ensMd
ensW

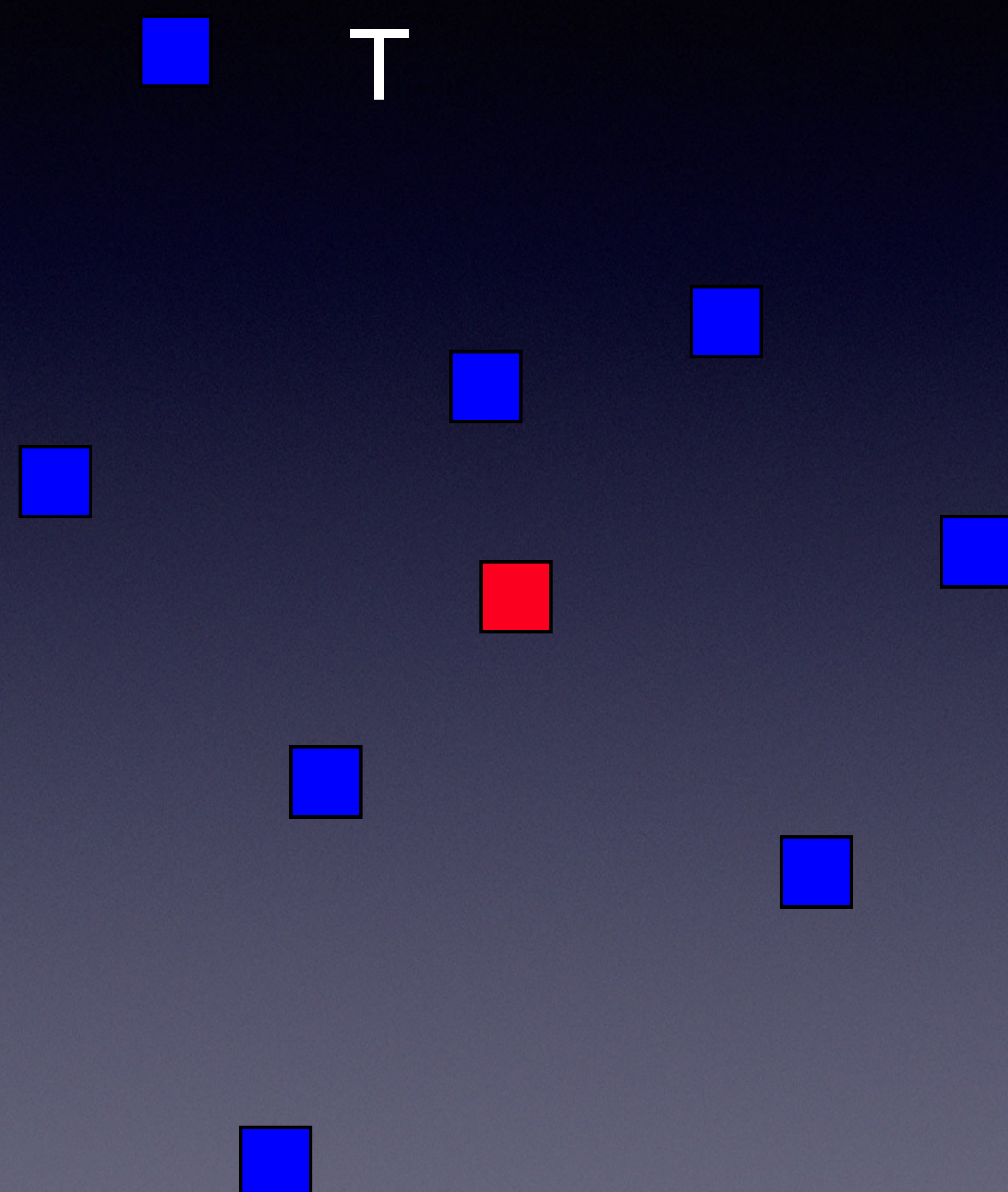
Mean REMS


REMS / min (REMS)
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
glmer0
glm1
gam1
gam2
gam3
mars3
RF
okrige
ukrige1
ukrige2
ukrige3
krigedRF
ensMn
ensMd

?




T


T

Ensembles: problems

- Error estimations: computation can be prohibitive
- Have to tune and fit multiple models
- Deal with fitting failures (non-convergence)
- Results dependent on sampling design?
- Return on investment = ?

Optimize for what?

- Prediction Accuracy
- Robustness
- Accuracy / effort

One size fits all?

One size fits all?

One size fits all?

photos: © M. Renner

One size fits all?

- In many cases, differences were small
- On average: RandomForest best non-spatial predictor, GAM best spatial
- Framework for comparison
- Method to fit question!

Open questions

- Algorithms: ZIP models, boosted trees, soap-film regression
- Spatial grid size
- Alternative cross-validations (by grid/year)
- Effect of survey design

Conclusions

- Use unbiased algorithms
- Extrapolation = Watchout
- Guard against Boundary Effects
- Computing cluster, EC2
- Wise allocation of $ and time = ?
- Get more samples and good predictors

Summary

- Models: spatial / non-spatial (habitat)

Summary

- Models: spatial / non-spatial (habitat)
- k-fold cross-validation framework

Summary

- Models: spatial / non-spatial (habitat)
- k-fold cross-validation framework
- Tested on 24 species

Summary

- Models: spatial / non-spatial (habitat)
- k-fold cross-validation framework
- Tested on 24 species
- RandomForests, spatial GAM

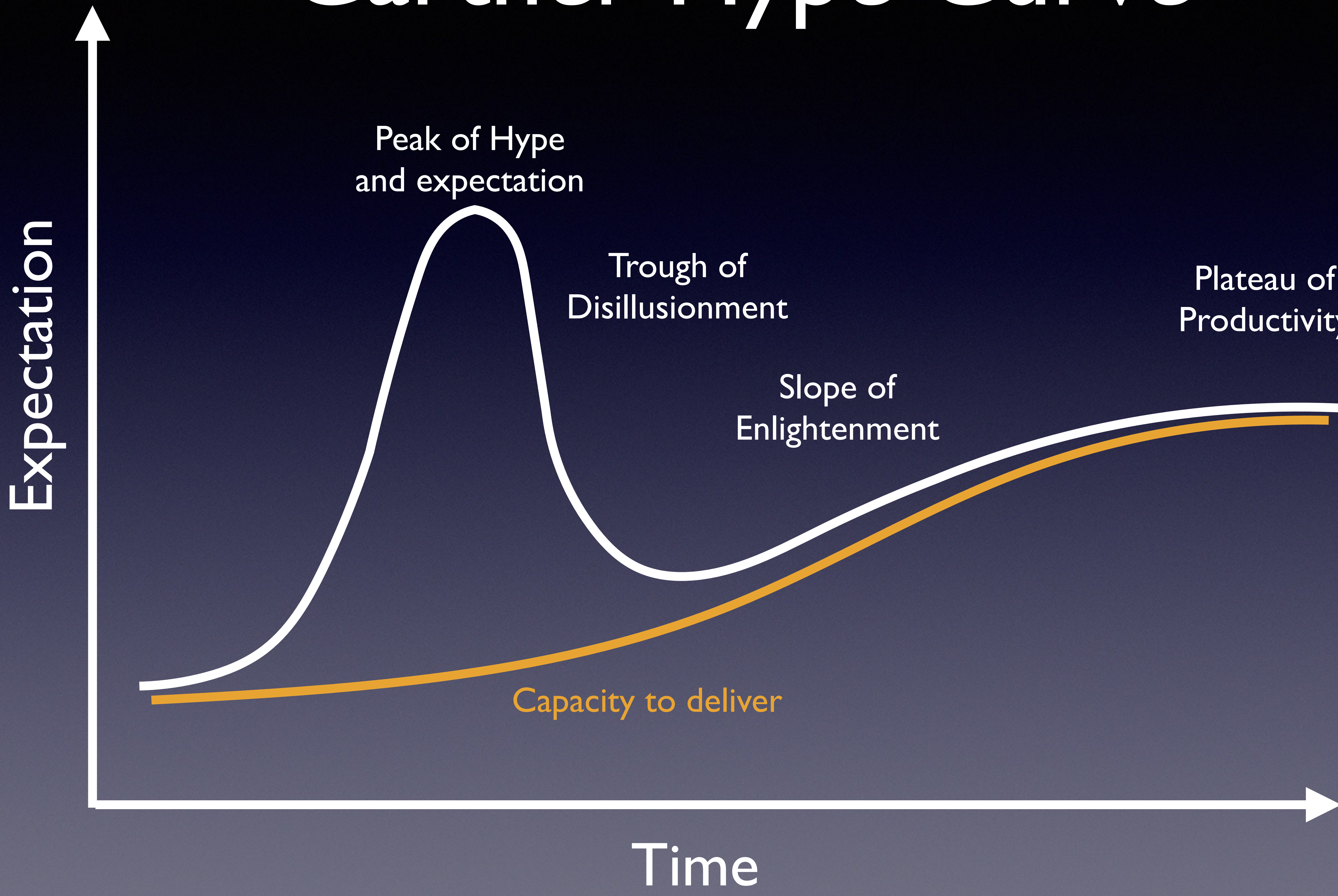
Summary

- Models: spatial / non-spatial (habitat)
- k-fold cross-validation framework
- Tested on 24 species
- RandomForests, spatial GAM
- Ensembles performed unexpectedly poorly, but are more robust against overfitting

Summary

- Models: spatial / non-spatial (habitat)
- k-fold cross-validation framework
- Tested on 24 species
- RandomForests, spatial GAM
- Ensembles performed unexpectedly poorly, but are more robust against overfitting
- There's no free lunch! There's a lot to learn!

Gartner Hype Curve



"All models are wrong, but some are useful"

George Box

“All models are wrong, but some are useful”

George Box

Use best fit for our *unique* study