TS Relationships by Local Regression

W. C. Thacker
Atlantic Oceanographic and Meteorological Laboratory

TS relationship is important for assimilating XBT data

For reanalysis, XBT provide bulk of the data.

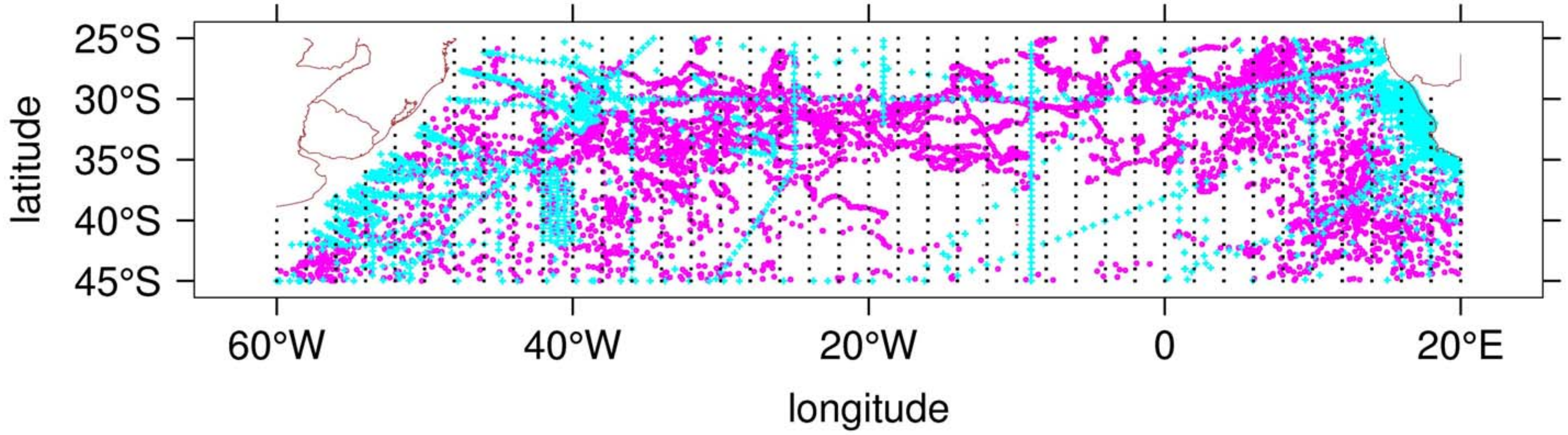
Fox, et al. (2002) used CTD data to get TS relationships.

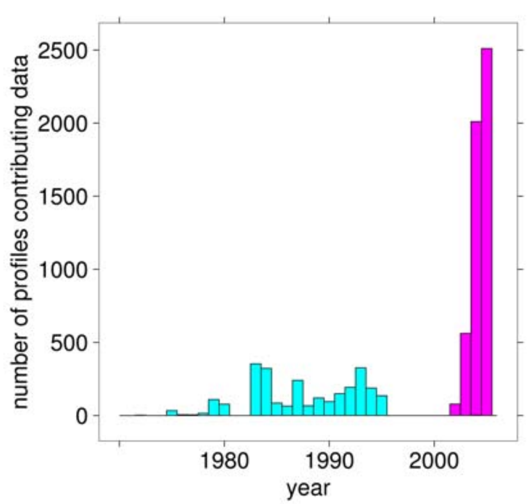
ARGO data now provide much better coverage and possibility of better TS relationships.

Local regression approach explored for South Atlantic.

- TS modeled on 3D grid: (2° lon, 1° lat, 25 dbar)
- Fit to data over larger area with distance-dependent weights.
- Models include quadratic terms for T, lon, lat.

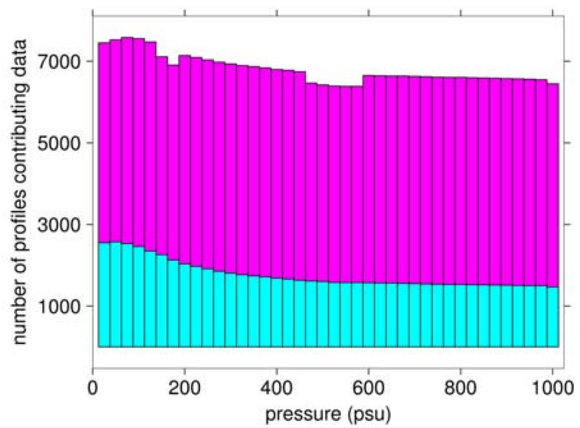
Available data for determining TS in South Atlantic





Cyan for CTD.

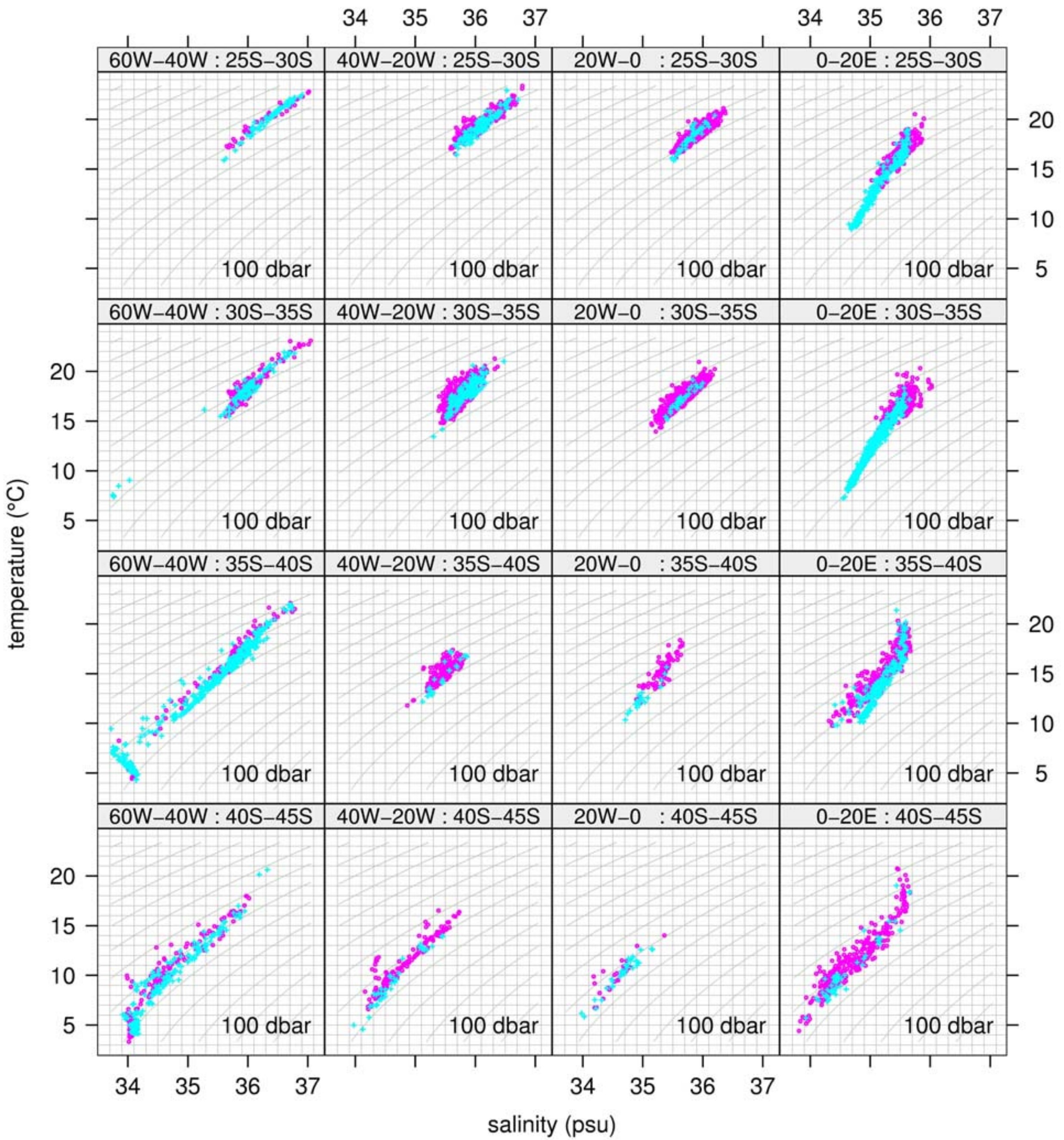
Magenta for ARGO.



CTD (cyan)

ARGO (magenta)

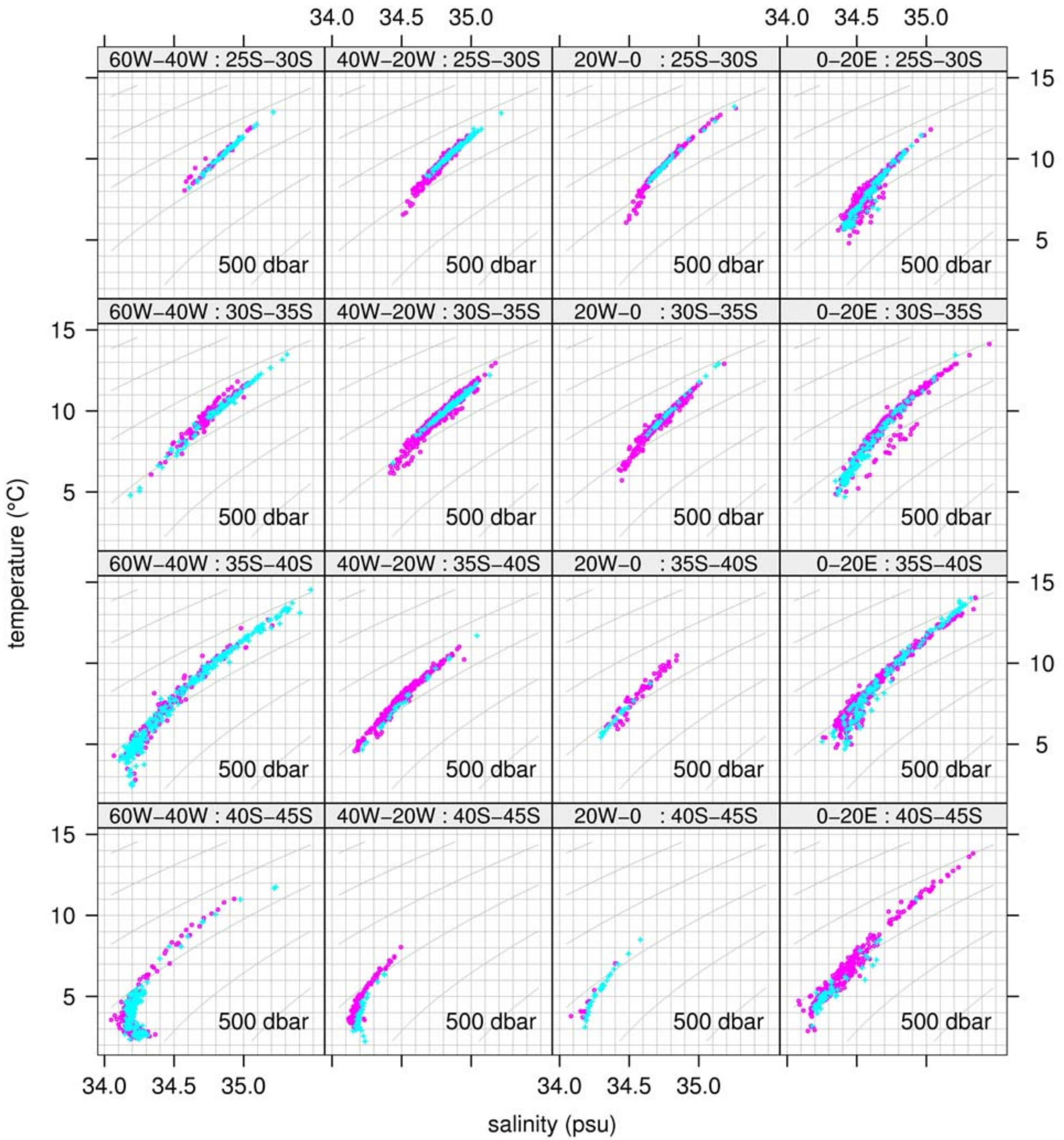
Panels show data for 20°×5° sub-regions at 100 dbar



CTD (cyan)

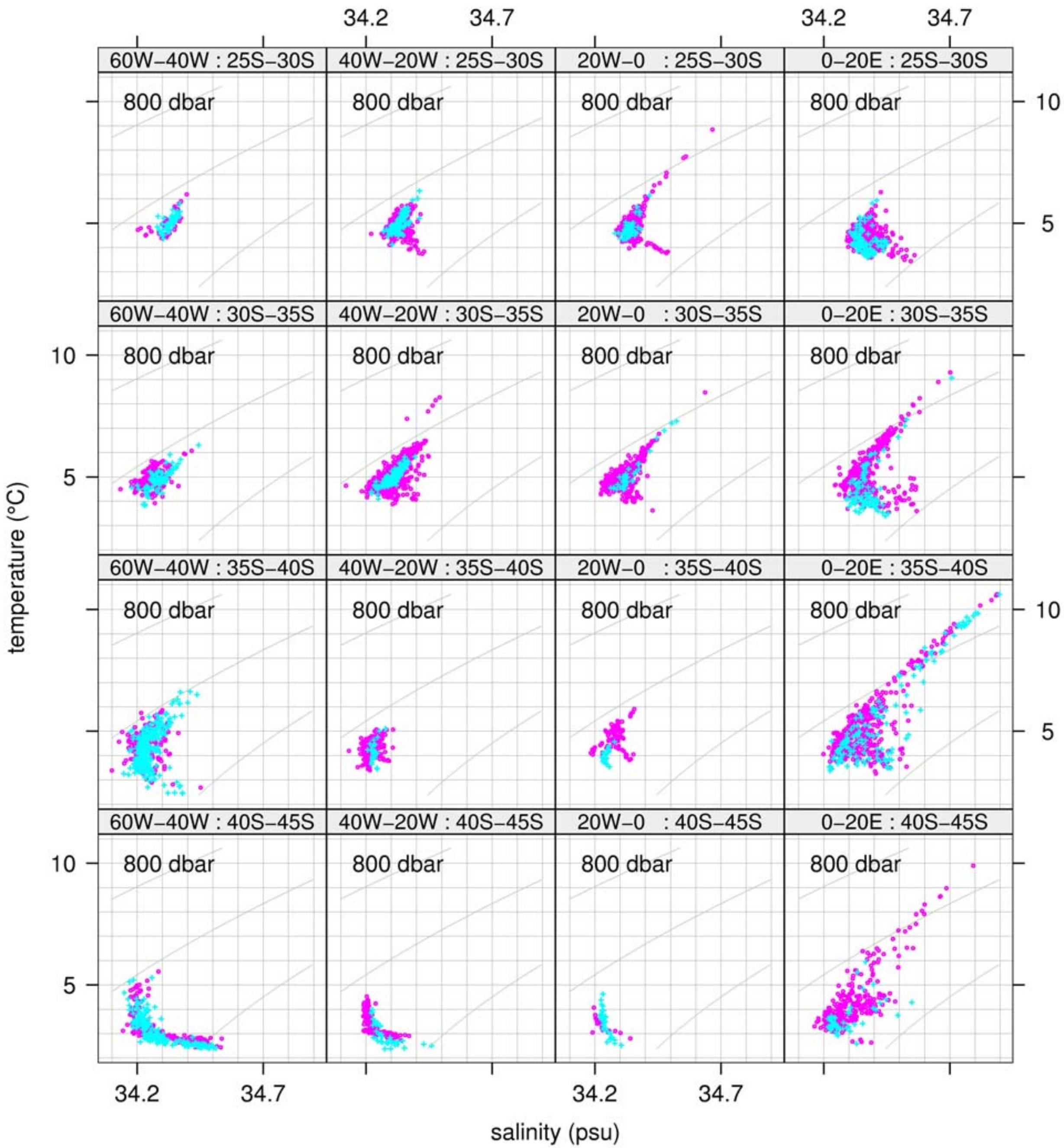
ARGO (magenta)

Panels show data for 20°×5° sub-regions at 500 dbar



CTD (cyan)

ARGO (magenta)

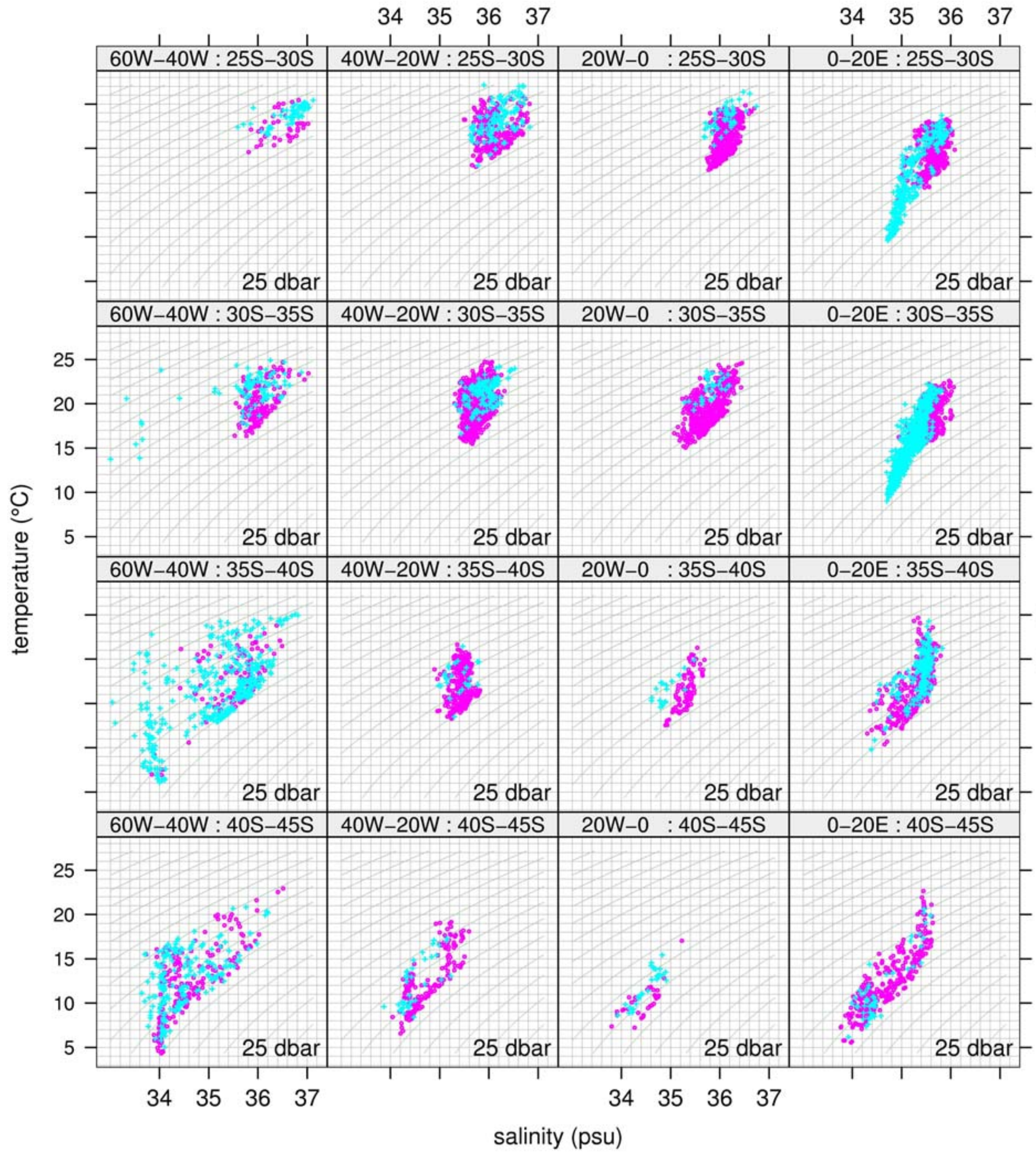


Panels show data for 20°×5° sub-regions

at 800 dbar

CTD (cyan)

ARGO (magenta)



Panels show data for 20°×5° sub-regions at 25 dbar

Local regression strategy.

Capture slow spatial change in TS at fixed depth by fitting models for each point on 2° lon × 2° lat × 25 dbar grid.

Fit to data from 100 closest stations.

Don't use data more than 7 cells away, even if fewer than 100.

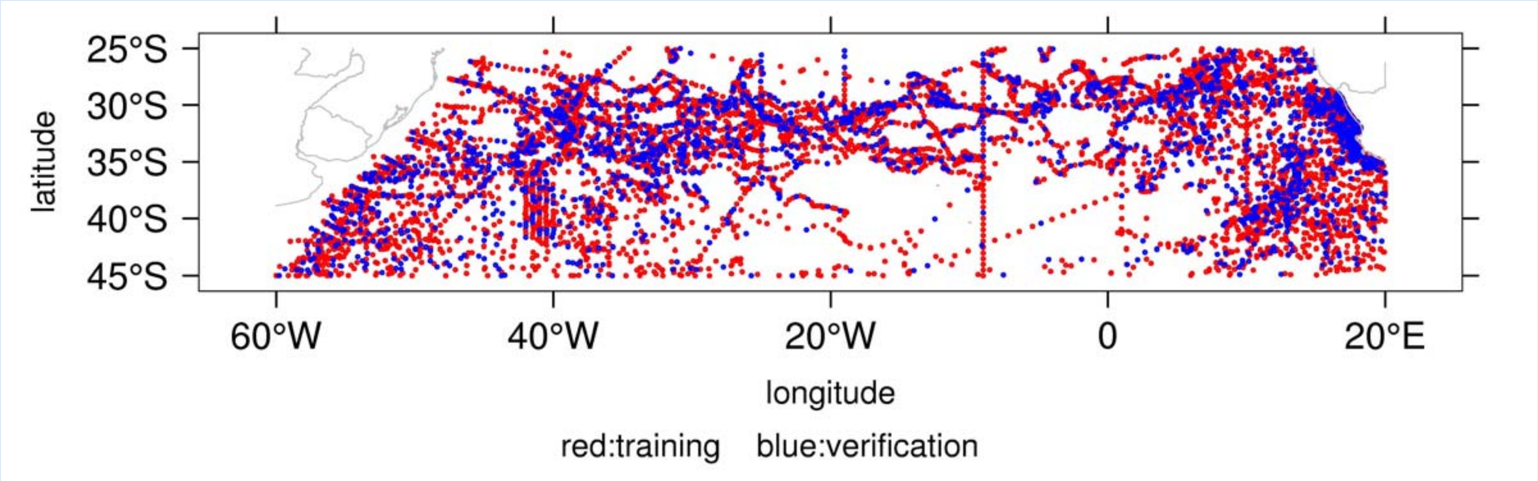
Use all data within a 1-cell radius, even if more than 100.

Diminish influence of distant data using variable weights:

$$w = \left(1 - \left(\frac{d}{D}\right)^3\right)^3$$

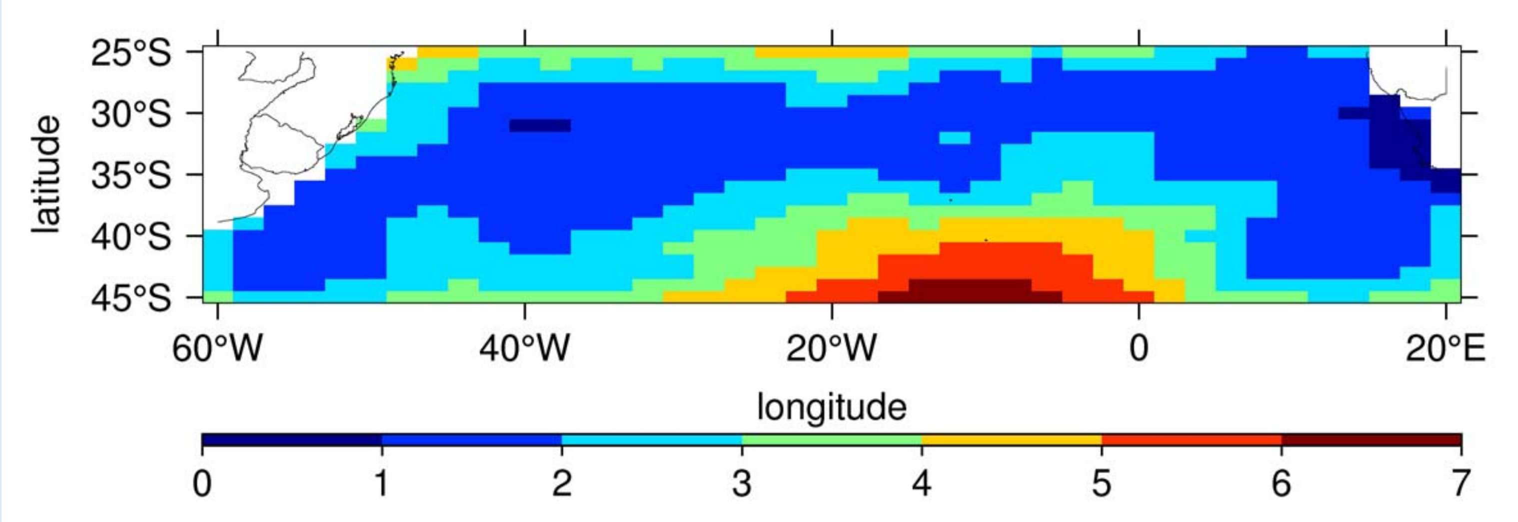
$D =$ distance to most remote station

Training and verification stations

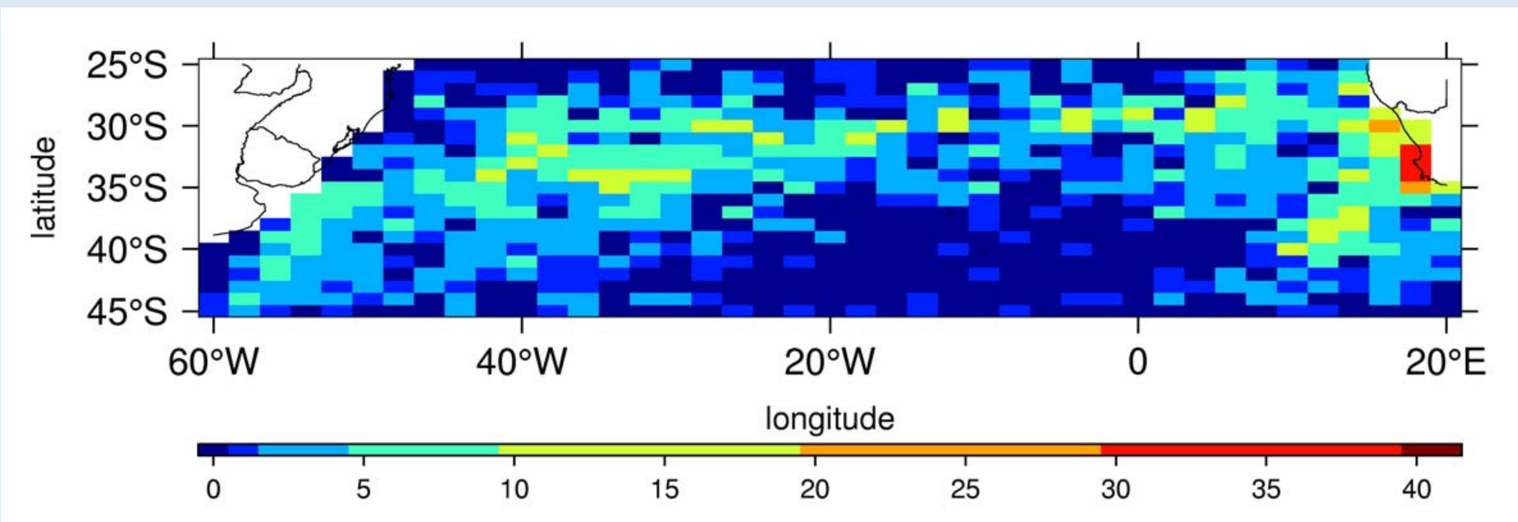


When there are at least 100 stations within a 7-cell radius, randomly choose 1/3 of the stations in the local cell to set aside for independent verification.

Each cell’s model was verified using only that cell’s data.



Elliptical distance to most distant point used for fitting.
(1 unit corresponds to 2° longitude and 1° latitude)



Number of local verification profiles available.

Steps

1) Identify outliers on TS plots.
   - Might use residuals to a preliminary fit to data.
   - Discard bad data.
2) Interpolate to target levels.
3) Gather data in neighborhood of grid point.
4) Compute distance-weighted fit to local data.
   - Can explore a variety of models.
   - Get set of regression coefficients for that point and level.
5) Repeat 3) and 4) for other points and levels.

Local regression models

$$\hat{S} = a + bT \qquad\qquad + dx + ey \tag{1}$$

$$\hat{S} = a + bT + cT^2 + dx + ey \tag{2}$$

$$\hat{S} = a + bT \qquad\qquad + dx + ey + fx^2 + gy^2 + hxy \tag{3}$$

$$\hat{S} = a + bT + cT^2 + dx + ey + fx^2 + gy^2 + hxy \tag{4}$$

$\hat{S}$ denotes the estimate for salinity, and T, x, and y denote observed temperature, longitude, and latitude, respectively.

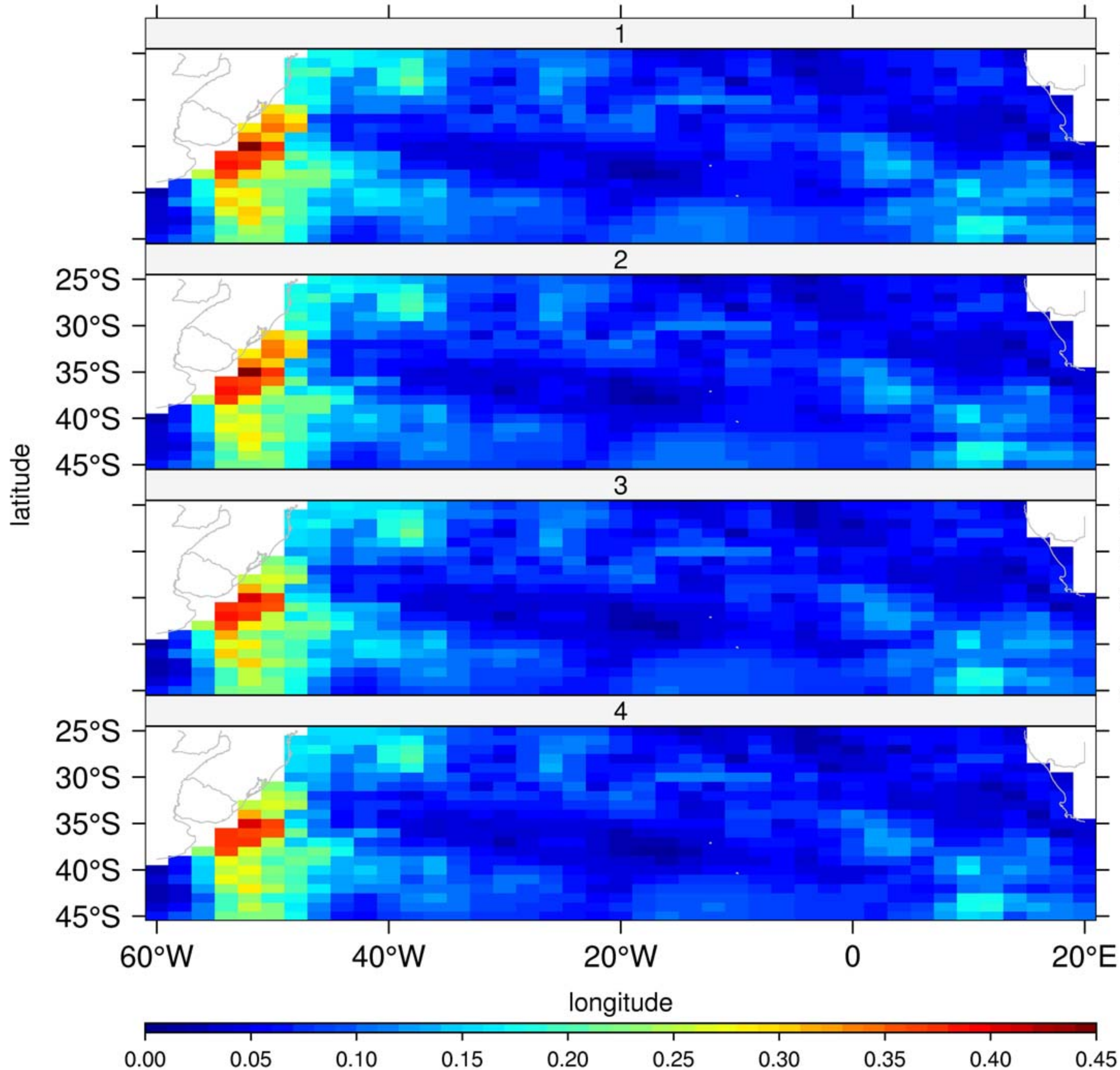
Coefficients $a, b, \ldots, h$ are determined for each model by fitting to the local training data.

25 dbar

Residual standard error (psu)

Models 2 & 4 ~T^2

Models 3 & 4 ~$(\text{lon},\text{lat})^2$

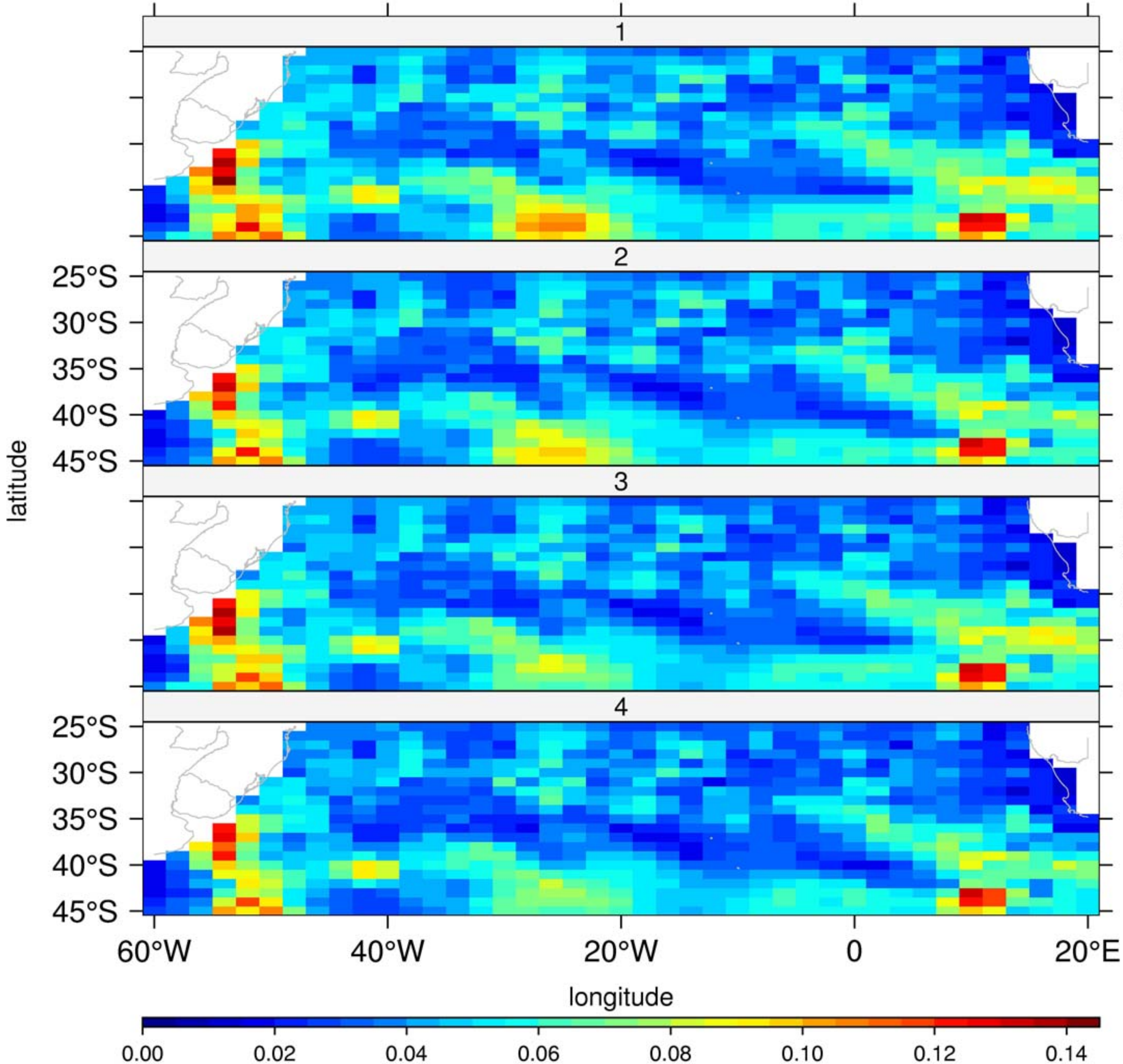


100 dbar

Residual standard error (psu)

Models 2 & 4 ~T^2

Models 3 & 4 ~$(lon,lat)^2$

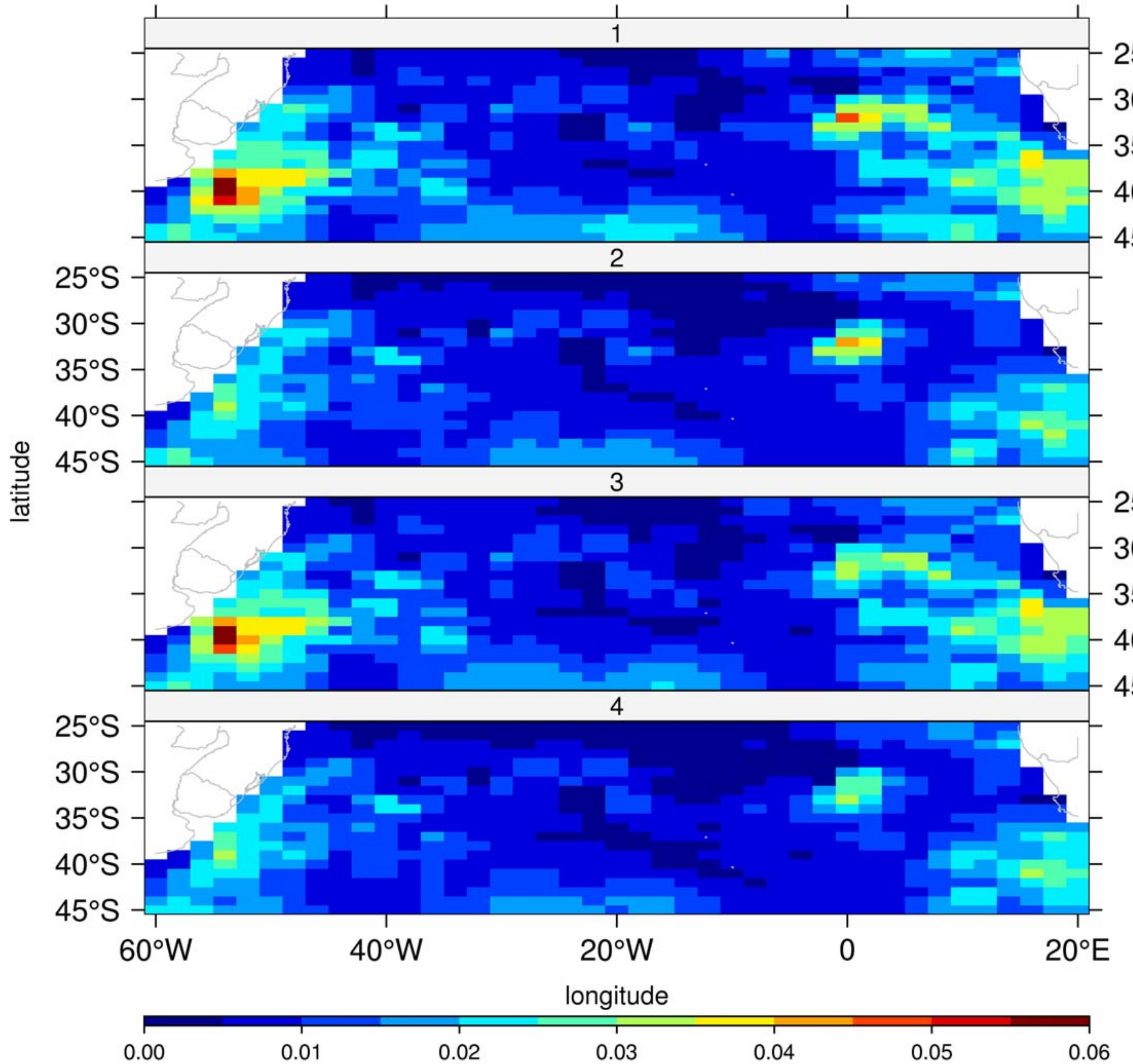


500 dbar

Residual standard error (psu)

Models 2 & 4 ~T^2

Models 3 & 4 ~$(lon,lat)^2$

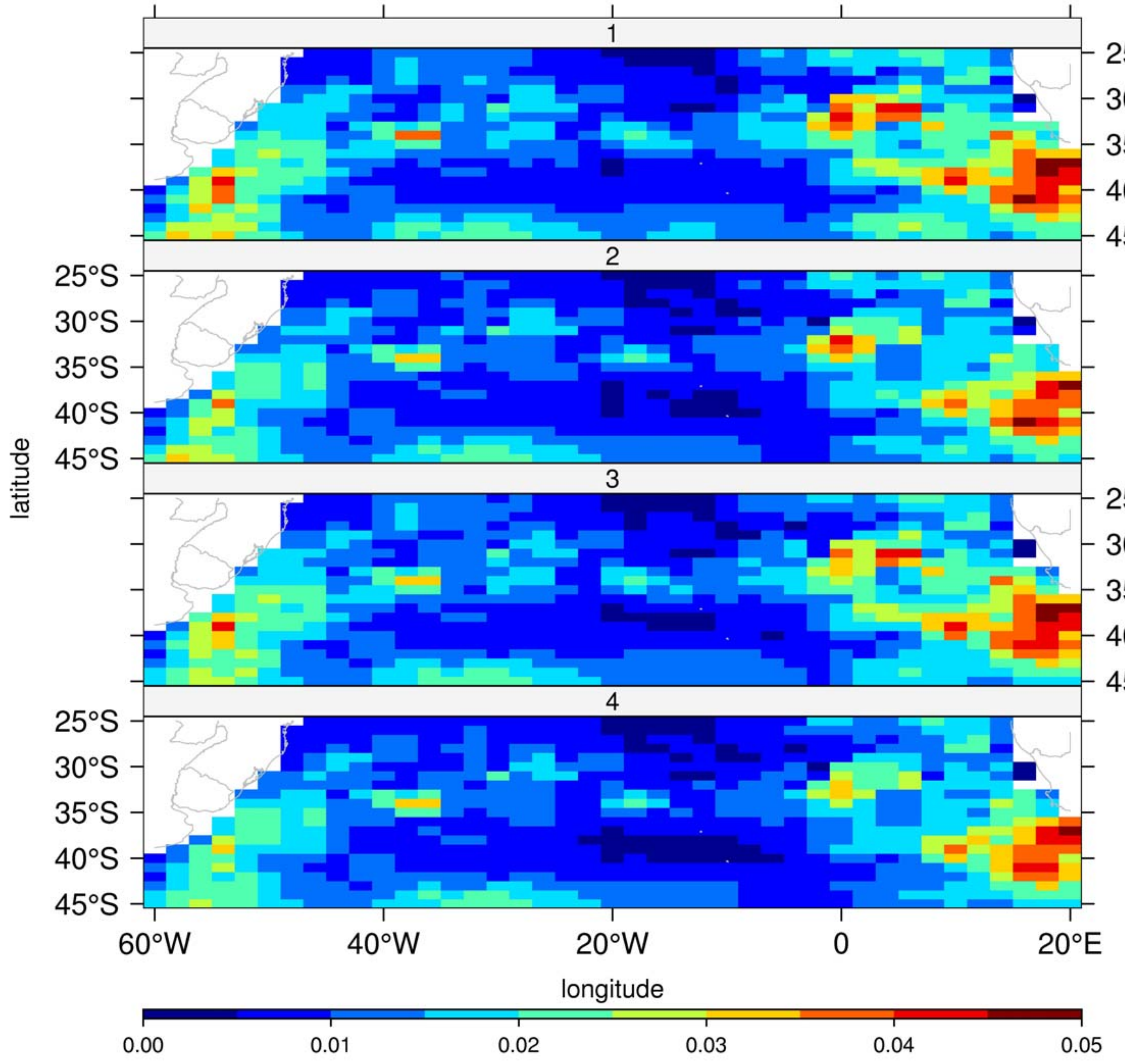


800 dbar

Residual standard error (psu)

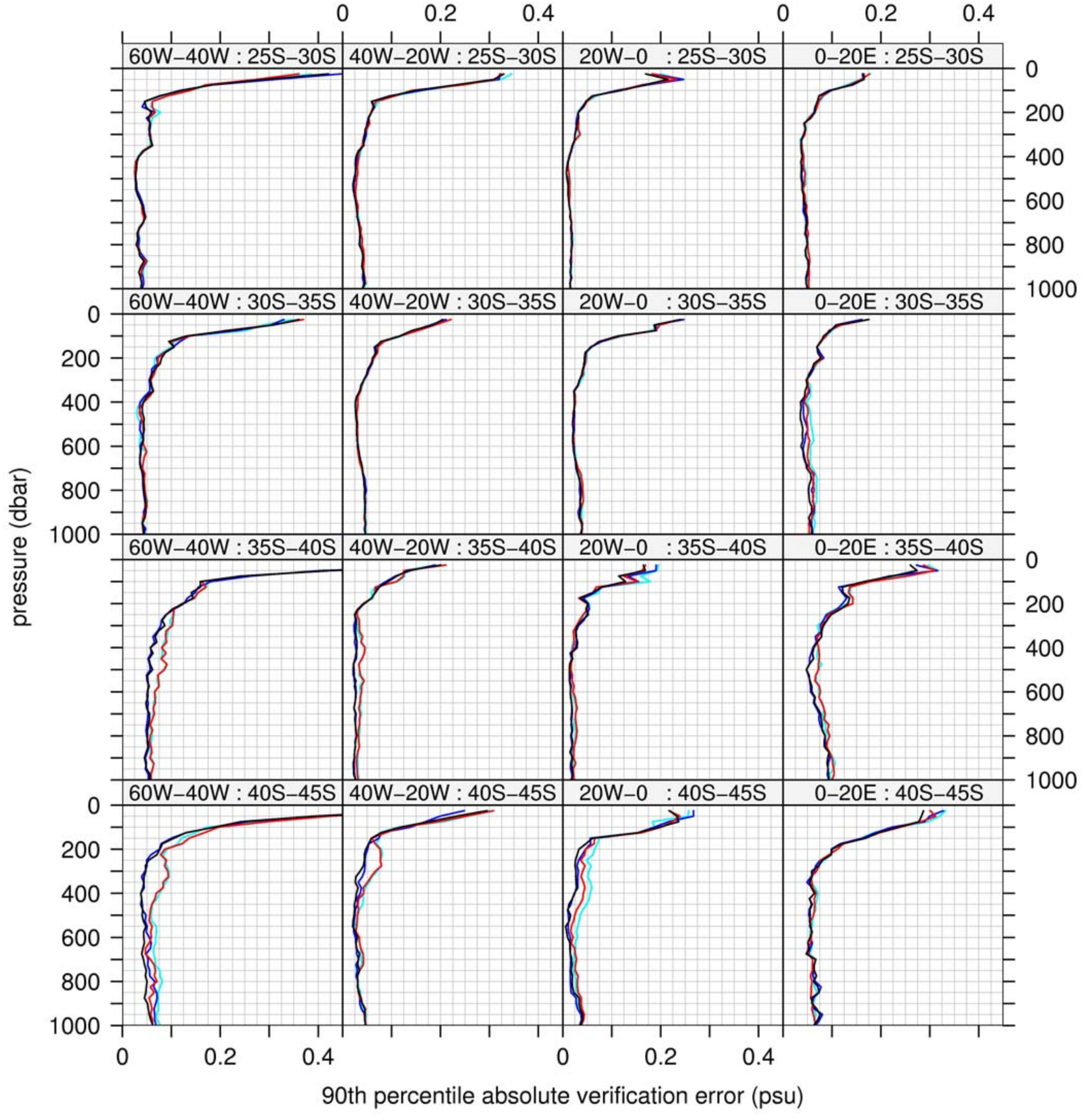
Models 2 & 4 ~T^2

Models 3 & 4 ~$(\mathrm{lon},\mathrm{lat})^2$



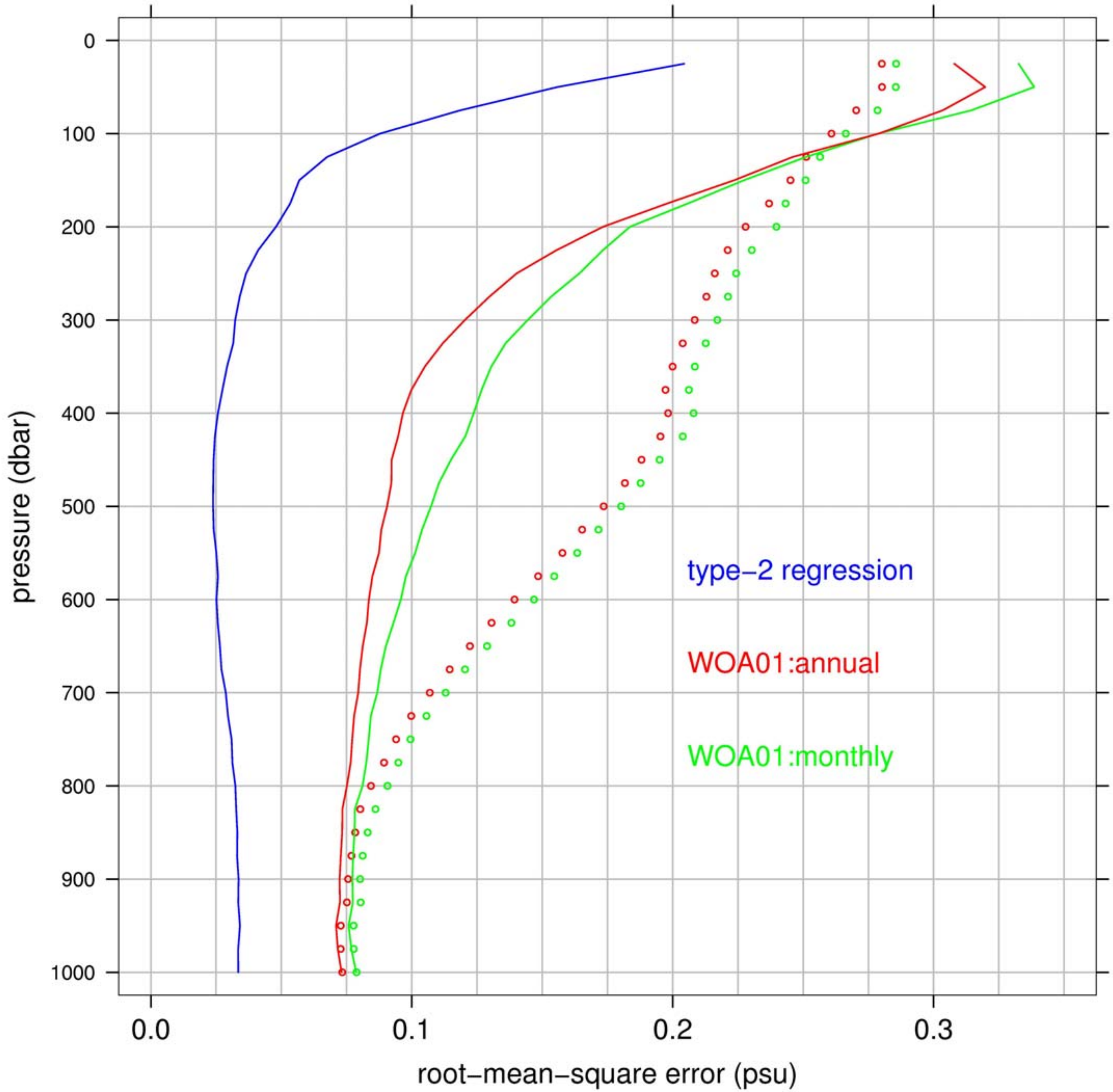
90th percentile absolute verification error (psu)

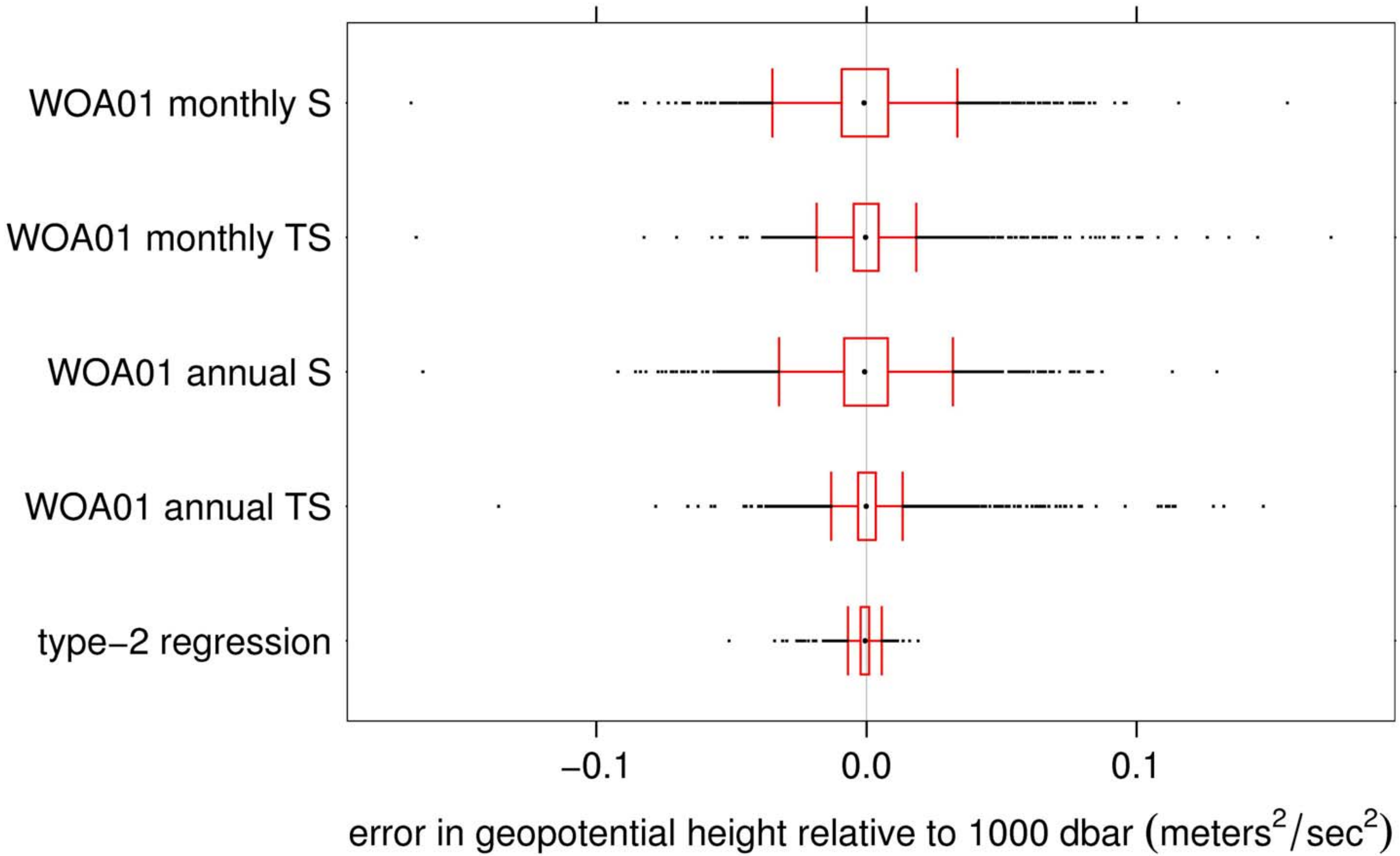
1-cyan
2-blue
3-red
4-black



Comparison of model 2 with estimates based on WOA01 climatology

Circles ignore TS relationship





0.1 m^2/sec^2 = 1 geodynamic centimeter

Conclusions

Local regression is easy to implement.

- Might be able to automate the process of excluding bad data.

Should work for most of the ocean.

- Can exploit ARGO's increased coverage.