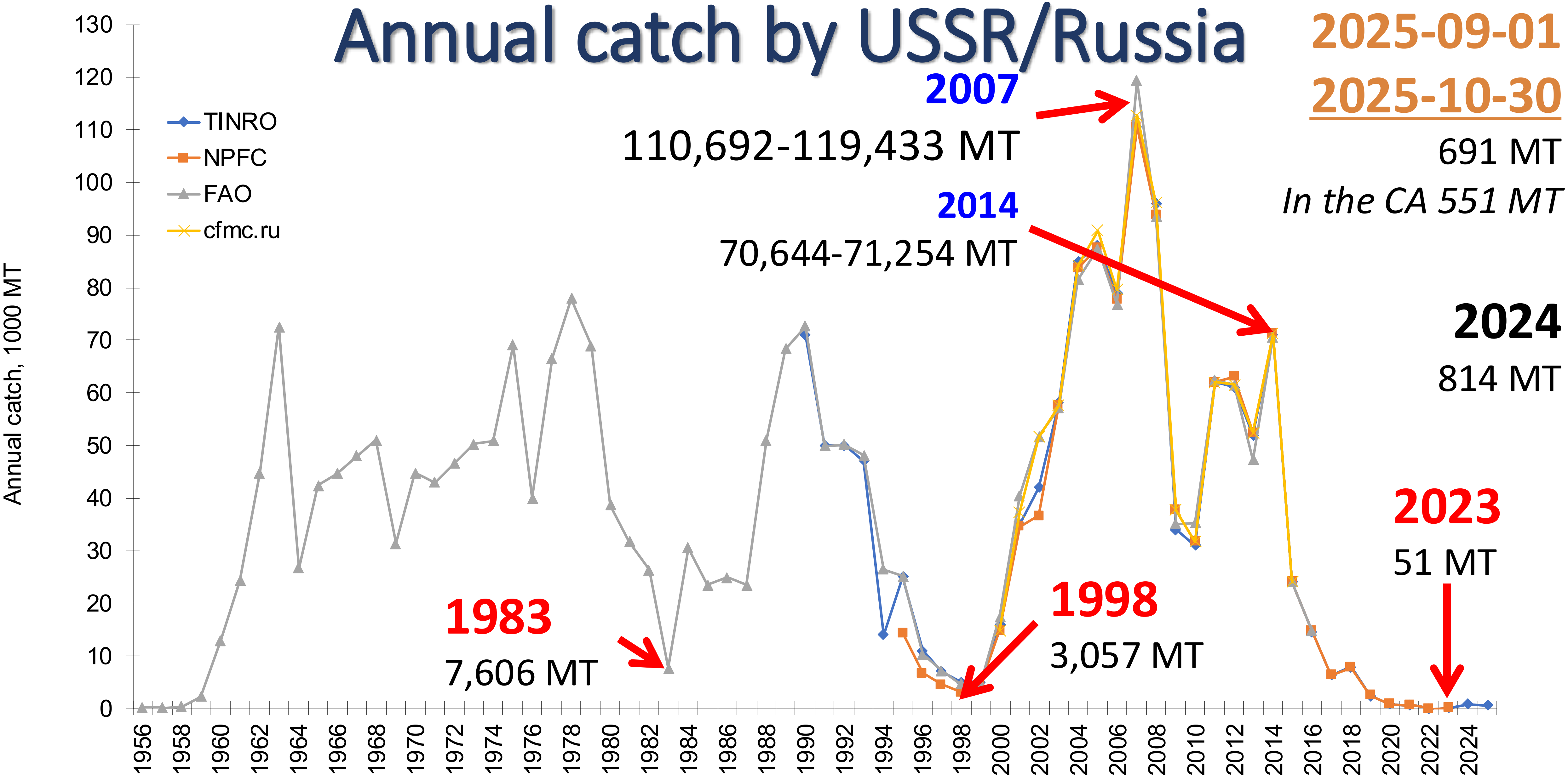
Pacific branch («TINRO») of Russian Federal
«Research Institute of Fisheries and Oceanography»
(«VNIRO»)

Significance of water properties for prediction of the Pacific saury occurrence in Russian catches

Vladimir Kulik[1], Aleksei Baitaliuk[1,2], Oleg Katugin[1], Maxim Budyansky[3] and Michael Uleysky[3]

[1]Pacific branch of "VNIRO" ("TINRO"), Vladivostok, Russia
E-mail: vladimir.kulik@tinro.vniro.ru
[2]Russian Federal Research Institute of Fisheries and Oceanography "VNIRO", Moscow, Russia
[3]V.I. Il'ichev Pacific Oceanological Institute of the Far-Eastern Branch
of the Russian Academy of Sciences "POI FEBRAS"

Annual catch by USSR/Russia



- **The annual catch increased from 0 in 2022 to 51 MT (1 fishing vessel) in 2023 and then to 813.93 MT in 2024 (2 fishing vessels): 286.01 MT in the CA**

Earlier, for 2004–2018, we used the absolute values and gradients of speed of passive particles, and Lagrangian indicators for encounter probability prediction in Random Forest (RF). Area Under the receiver operating characteristic Curve (AUC) reached 0.85 in it.

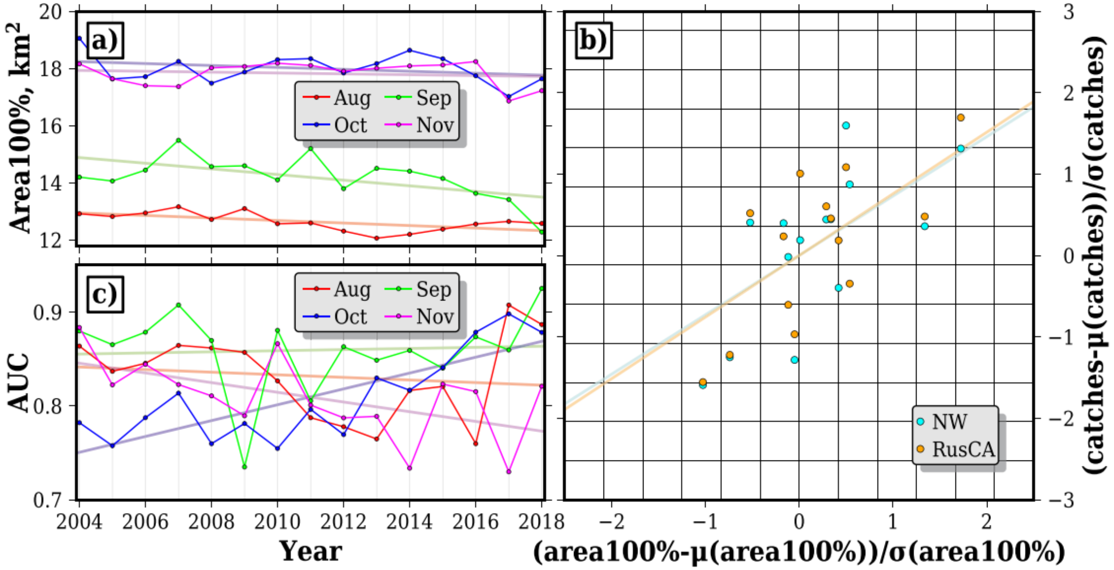


Fig. 3. Monthly averaged area (km^2) with 100% probability of saury catch per day per year with linear trends (a), scaled with the mean and *SD* catches of saury in the CA by Russia (RusCA) and totally in the national waters between Russia and Japan (NW) versus similarly scaled values of area (km^2) with 100% probability of catch of saury per day per year in September (b) and AUC values calculated for the test set sliced by month and year (c).

Fisheries Research 252 (2022) 106361

ELSEVIER

Contents lists available at ScienceDirect

Fisheries Research

journal homepage: www.elsevier.com/locate/fishres

Check for updates

Lagrangian characteristics in the western North Pacific help to explain variability in Pacific saury fishery

Vladimir V. Kulik [a,*], Sergey V. Prants [b], Michael Yu. Uleysky [b], Maxim V. Budyansky [b]

[a] Pacific branch (TINRO) of the Federal State Budget Scientific Institution "Russian Federal Research Institute of Fisheries and oceanography" (VNIRO), 4, Shevchenko Alley, Vladivostok 690091, Russia

[b] V.I. Il'ichev Pacific Oceanological Institute of the Far-Eastern Branch of the Russian Academy of Sciences (POI FEBRAS), 43, Baltiyskaya St., Vladivostok 690041, Russia

A R T I C L E I N F O

Keywords:
Saury
Species distribution modeling
Lagrangian indicator
Lyapunov exponent

A B S T R A C T

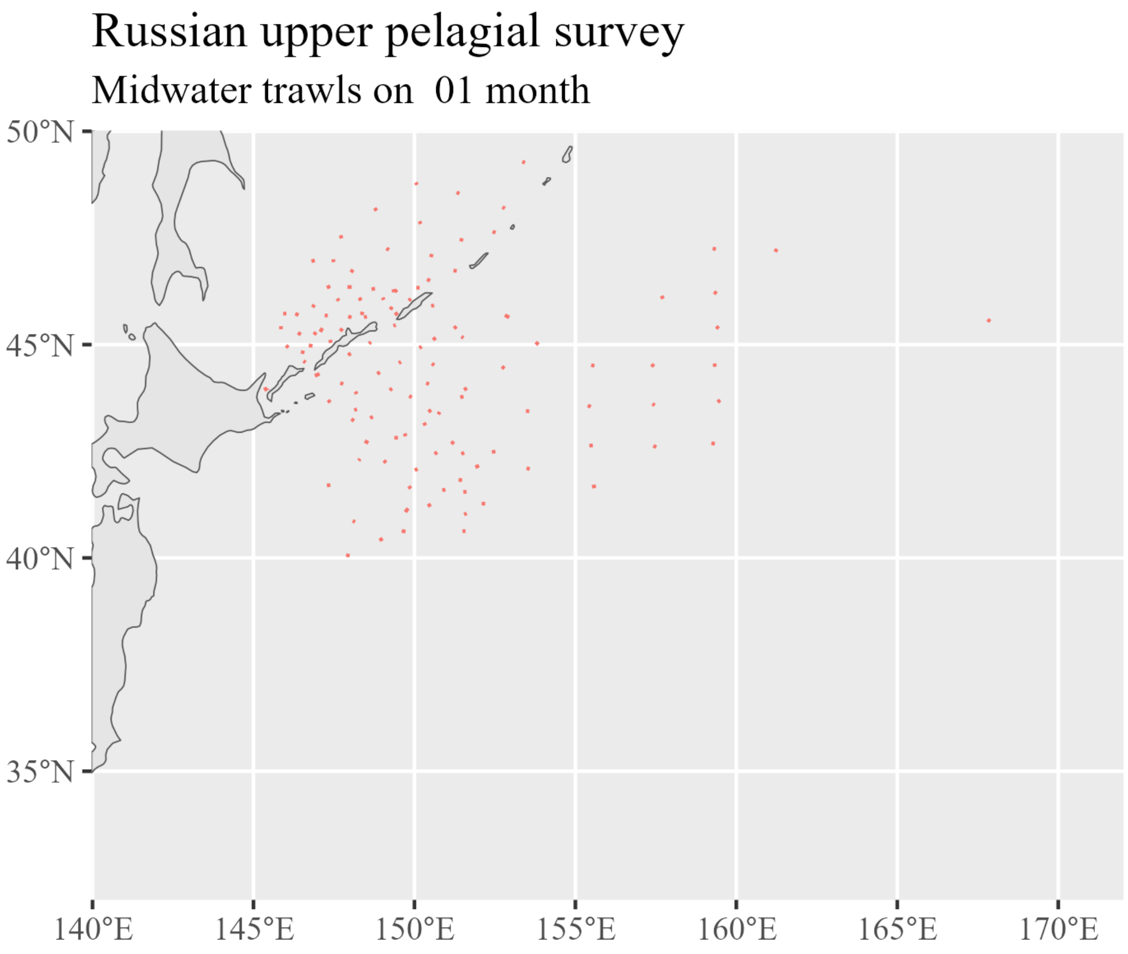
A new model for estimation of daily probability for the Pacific saury (*Cololabis saira*) encounter was proposed. The model performance was tested for the period of 2004–2018 (August–November) using the data from the Russian vessel monitoring system. The following physical oceanographic variables were used for encounter probability prediction: the absolute values and gradients (∇) of speed (V) of passive particles, imitating water parcels, and Lagrangian indicators. The positive effects on the encounter probability of saury were found for V, ∇V, and for the gradient of the finite-time Lyapunov exponent (∇Λ), while the effect of particle path length was negative. That means that saury preferred places close to the boundaries of the oceanographic features, where Lagrangian fronts are situated, but not inside the features themselves, because Λ is small in regular flows and large at Lagrangiam fronts. The model did not include information about years and volume of saury catches, but its monthly mean of catch probability in September had the highest correlation with Russian annual catches outside the national waters between Russia and Japan ($r = 0.76$, $p = 0.001$) and total annual catches there ($r = 0.73$, $p = 0.002$). Timeseries analysis of principle components (PC) from daily predictions of saury catch probabilities has also shown that the third PC correlated highly with the annual biomass of saury ($r \geq 0.8$, $p < 0.05$). The model seems to be useful to manage Russian fishery and may help to explain the reasons for the saury biomass decline. The latter is very important to take into account for development of the stock assessment models.

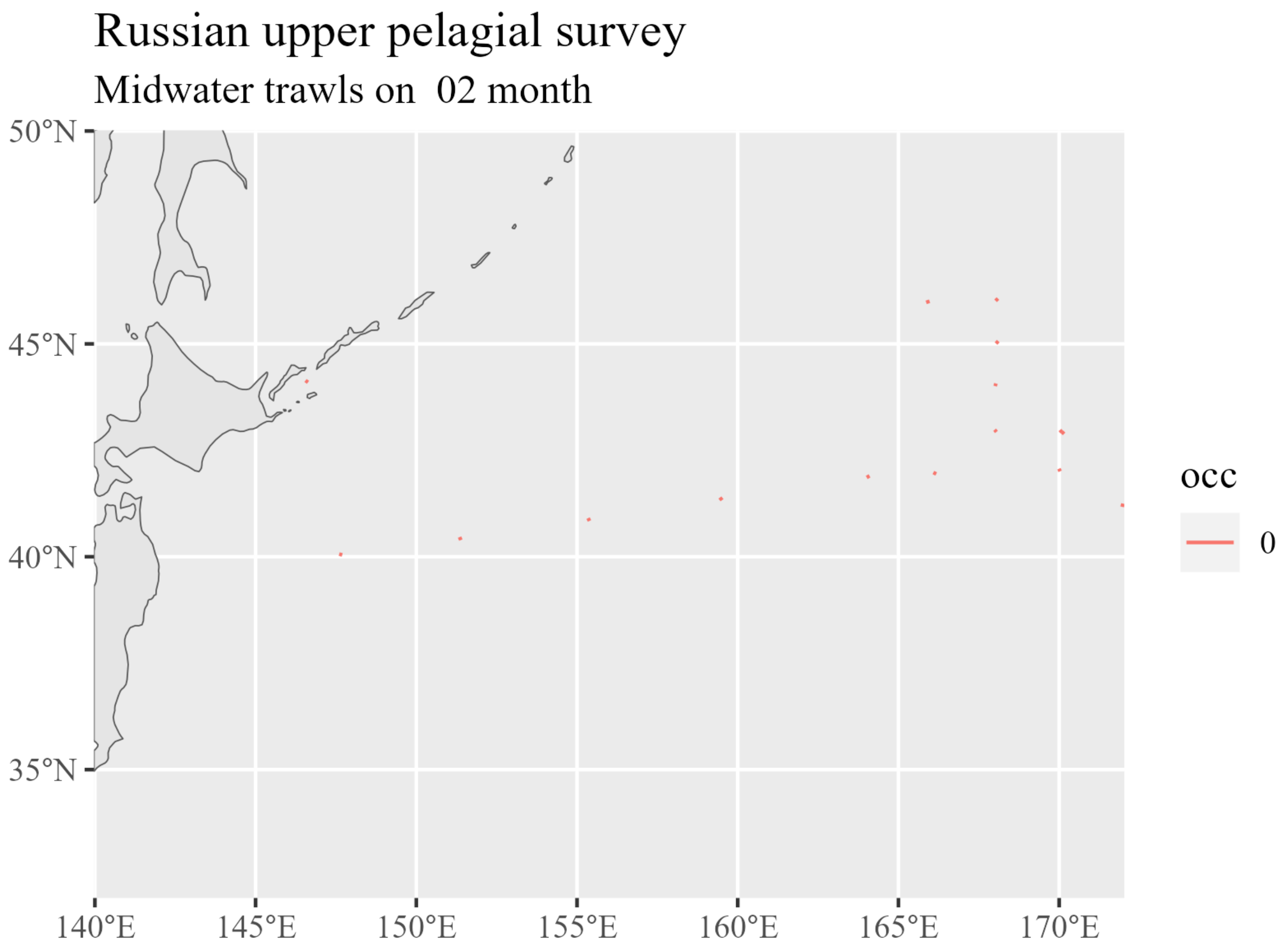
The final goal of our study is in improvement of the SDM for saury.

Consequently, we must test

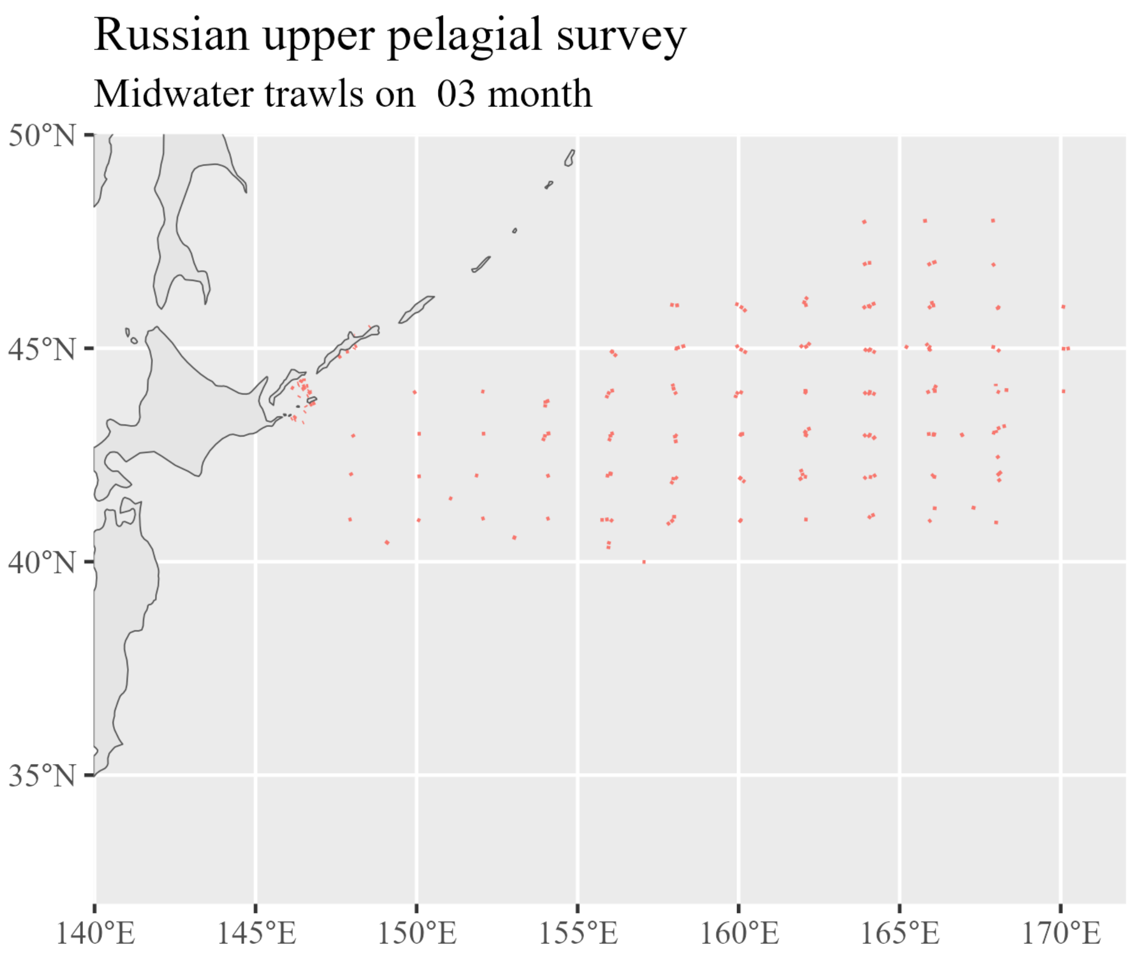
- New data on occurrences
    - Expanded timeframe
    - Added 0 catch of saury from Russian scientific surveys
    - *Hope to cooperate with PICES scientists to extend the range in geographical space more into the High Seas*
- New features for predictors:
    - Added from GLORYS12V1

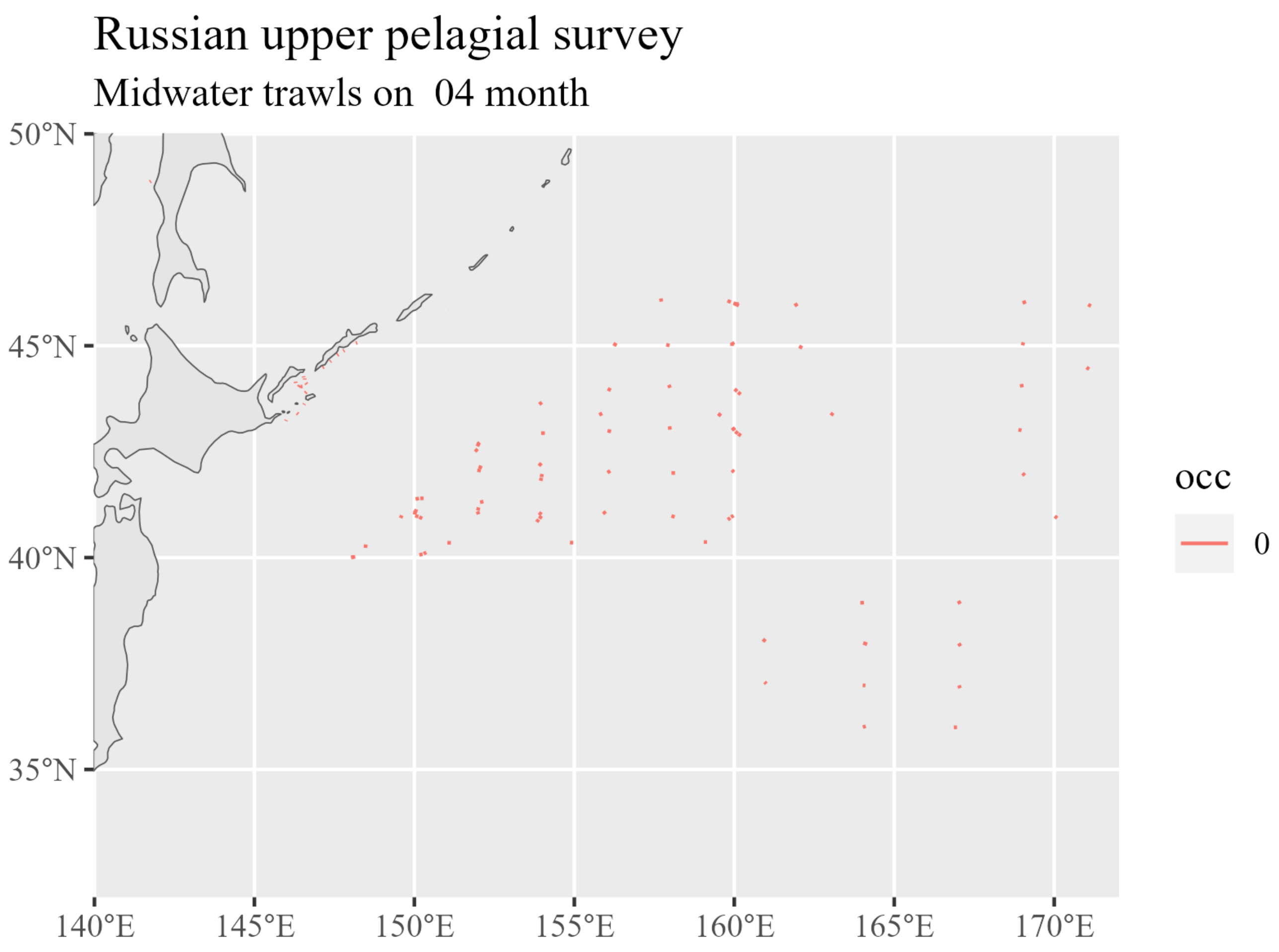
New data on saury absence in the midwater trawl survey of pelagial in **January** and **February**



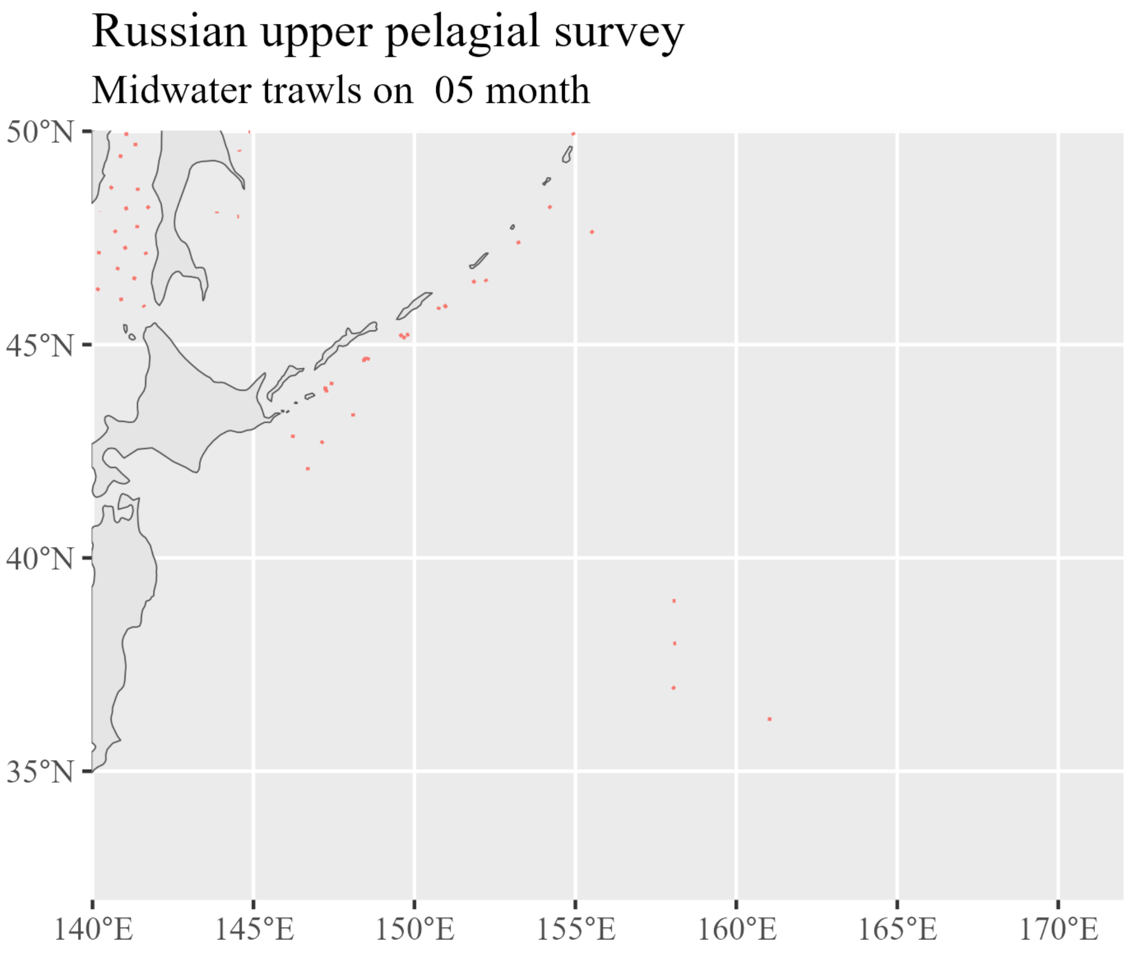


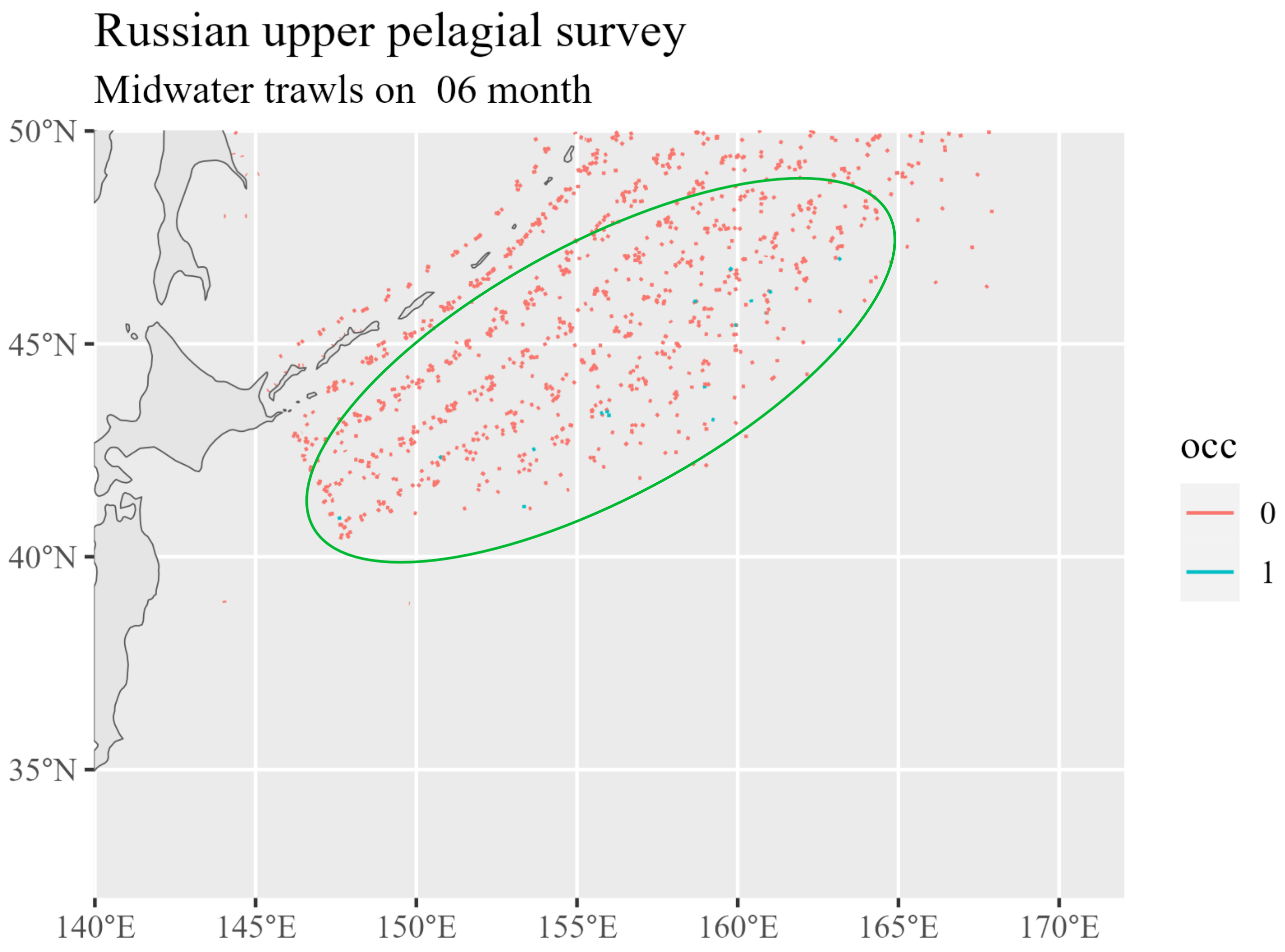
New data on saury absence in the midwater trawl survey of pelagial in **March** and **April**



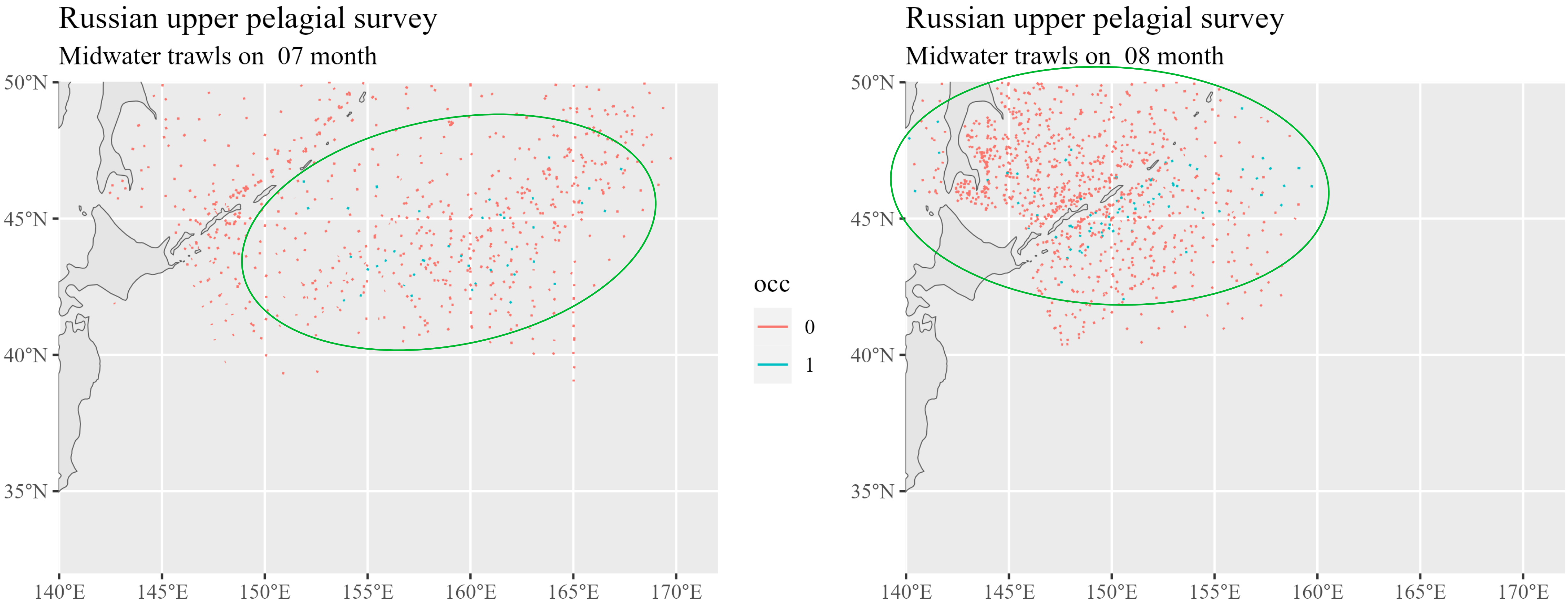


New data on saury absence (occ=0) and presence (occ=1) in the midwater trawl survey of pelagial in **May** and **June**

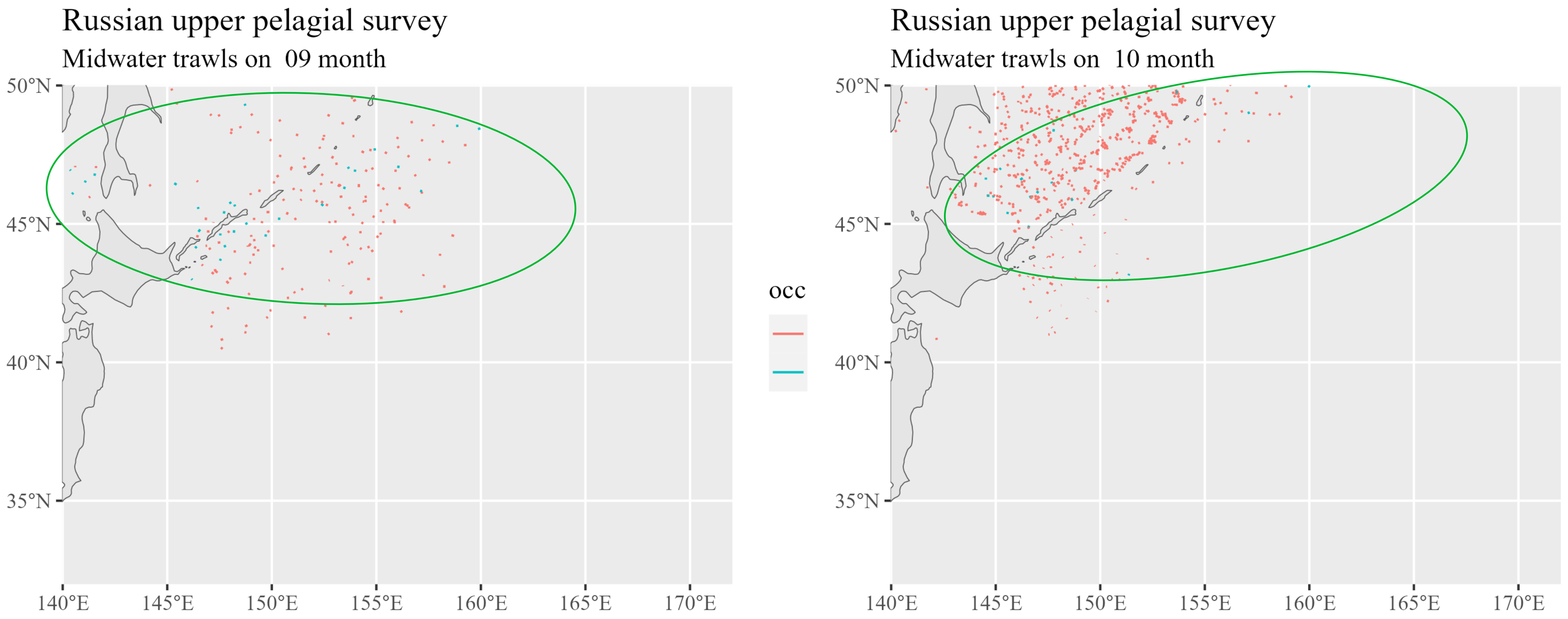




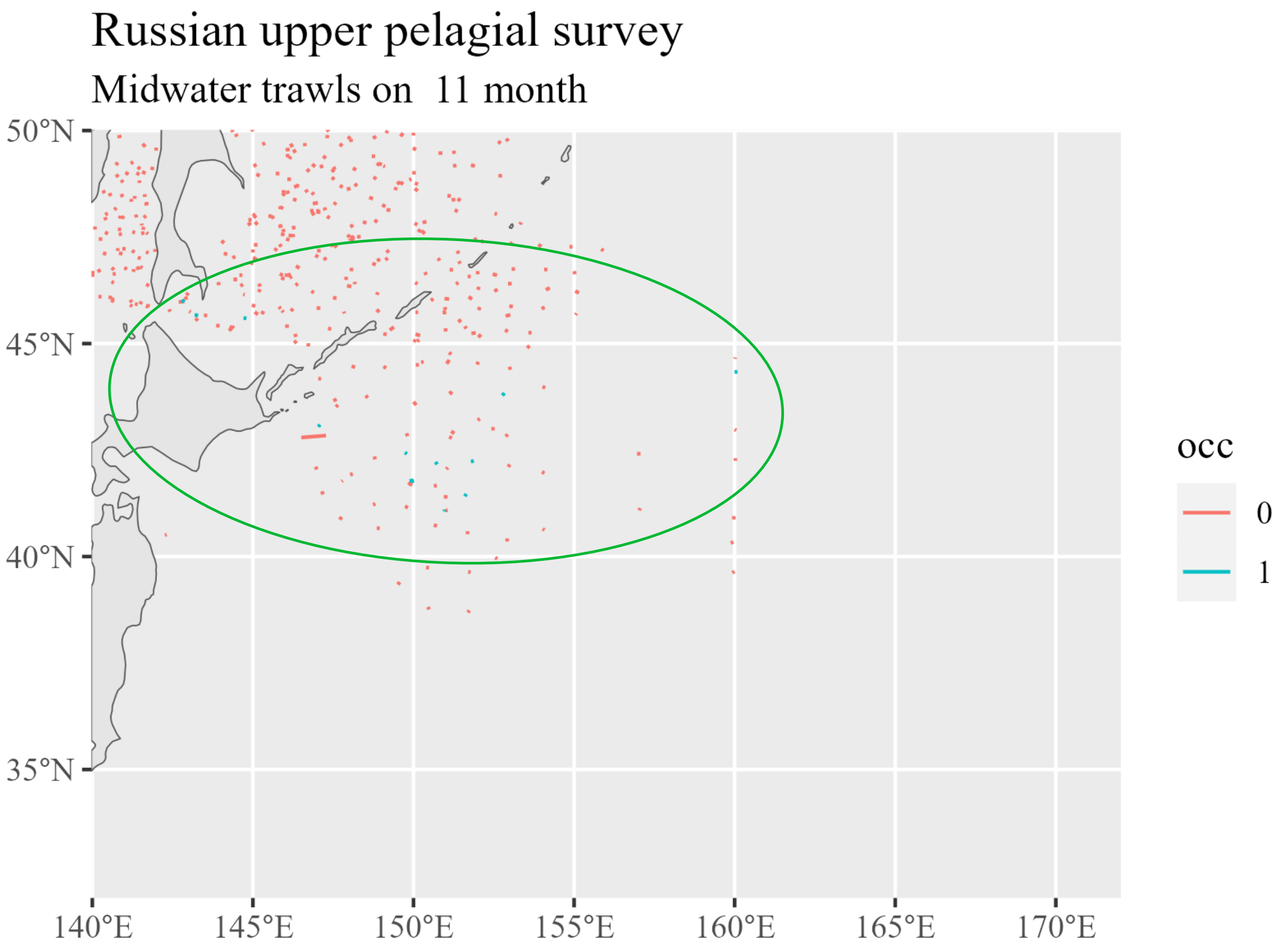
New data on saury absence (occ=0) and presence (occ=1) in the midwater trawl survey of pelagial in **July** and **August**

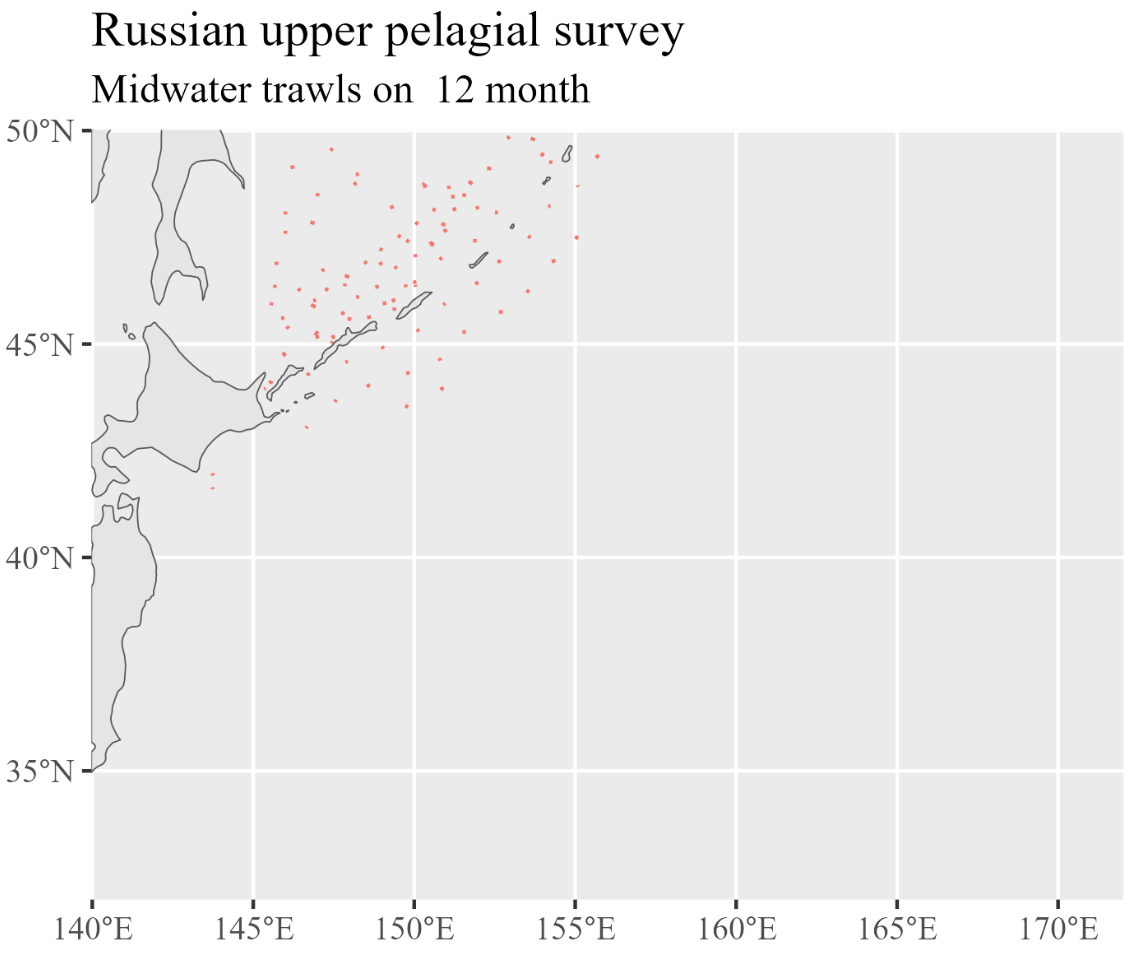


New data on saury absence (occ=0) and presence (occ=1) in the midwater trawl survey of pelagial in **September** and **October**



New data on saury absence (occ=0) and presence (occ=1) in the midwater trawl survey of pelagial in **November** and **December**





In addition to used earlier by us for shorter timeframe (2004–2018)
Lagrangian indicators back calculated in time for the period of 1 month :

L — Lyapunov exponent (day^{-1}) and

S — particle path length, km

and coordinates (**Lon,Lat**) with an ordinal day in a Year (**jDay**) without Year

This time, for a wider period of 1994–2021, we included
new features to predict saury occurrence in RF:

Dshore — Distance to the shore (from GFW),

wT — water temperature and

Sal — salinity

at horizons #0, 1, 12, 17, 22, 25, 28

(or at depth around 0.5, 1.5, 21.6, 47.3, 110, 186 and 318 m)

from the Global Ocean Physics Reanalysis (GLORYS12V1),

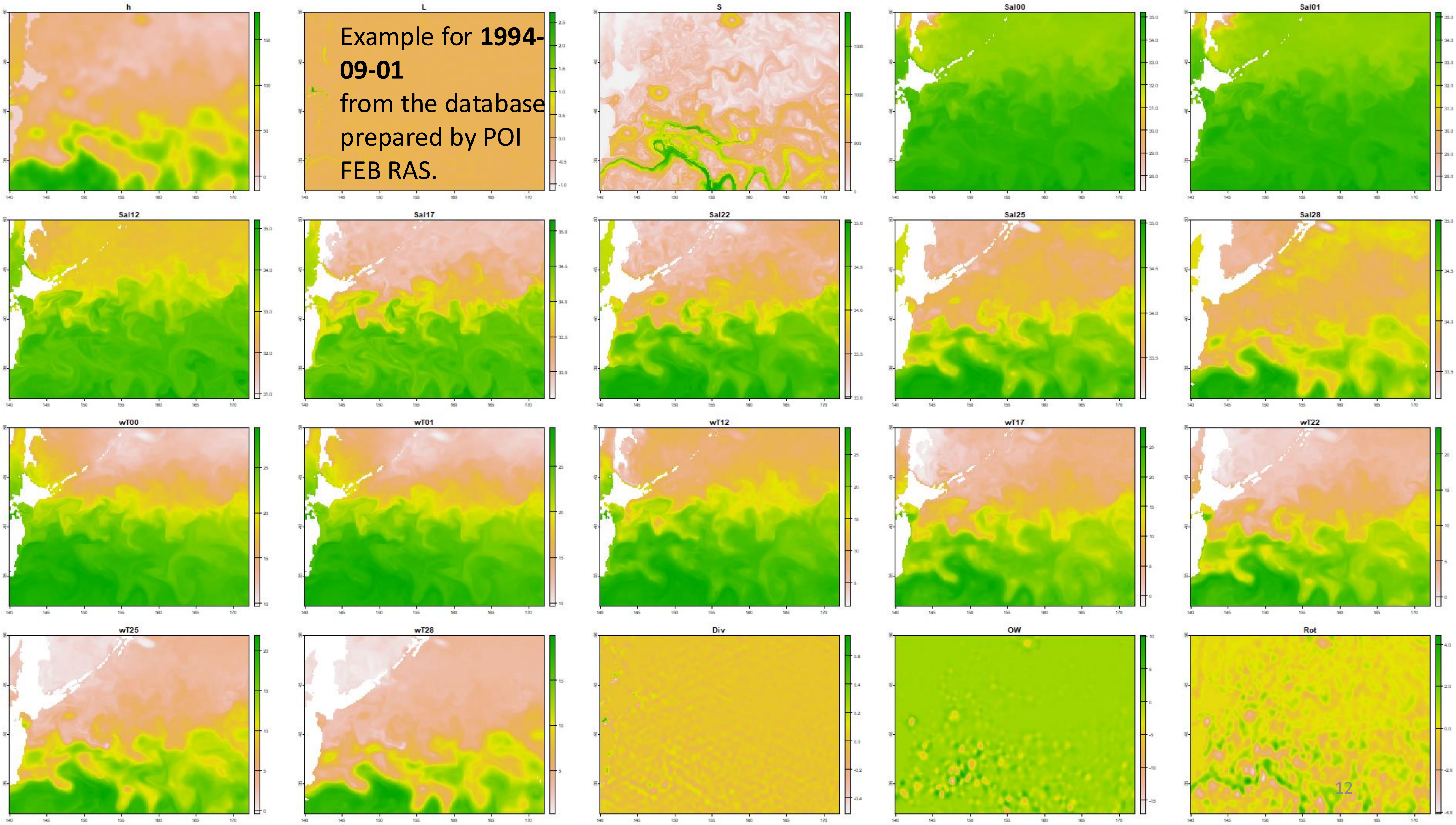
Div — Divergence,

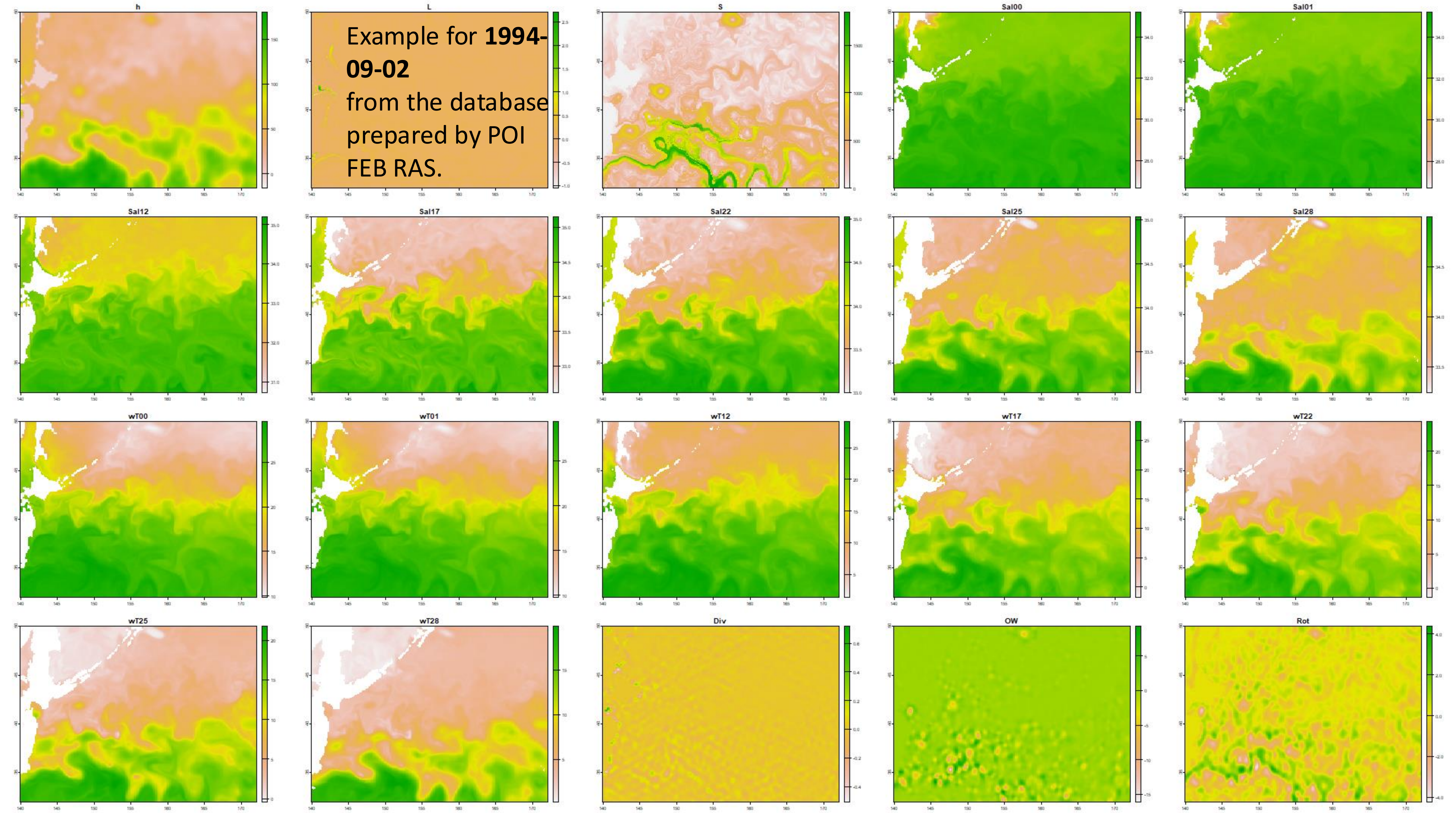
OW— Okubo–Weiss parameter,

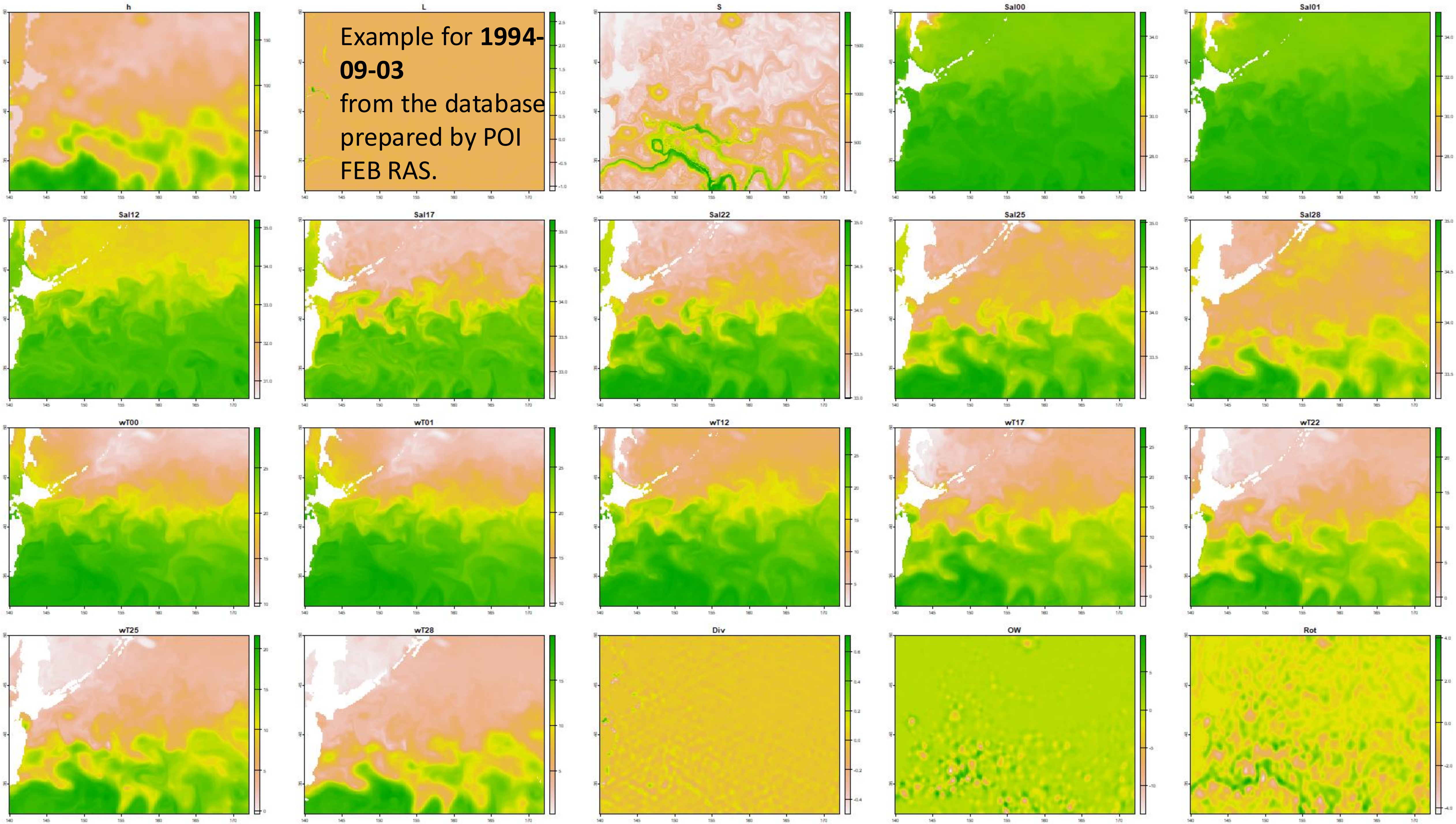
Rot — Rotor,

SSH — Sea Surface Height

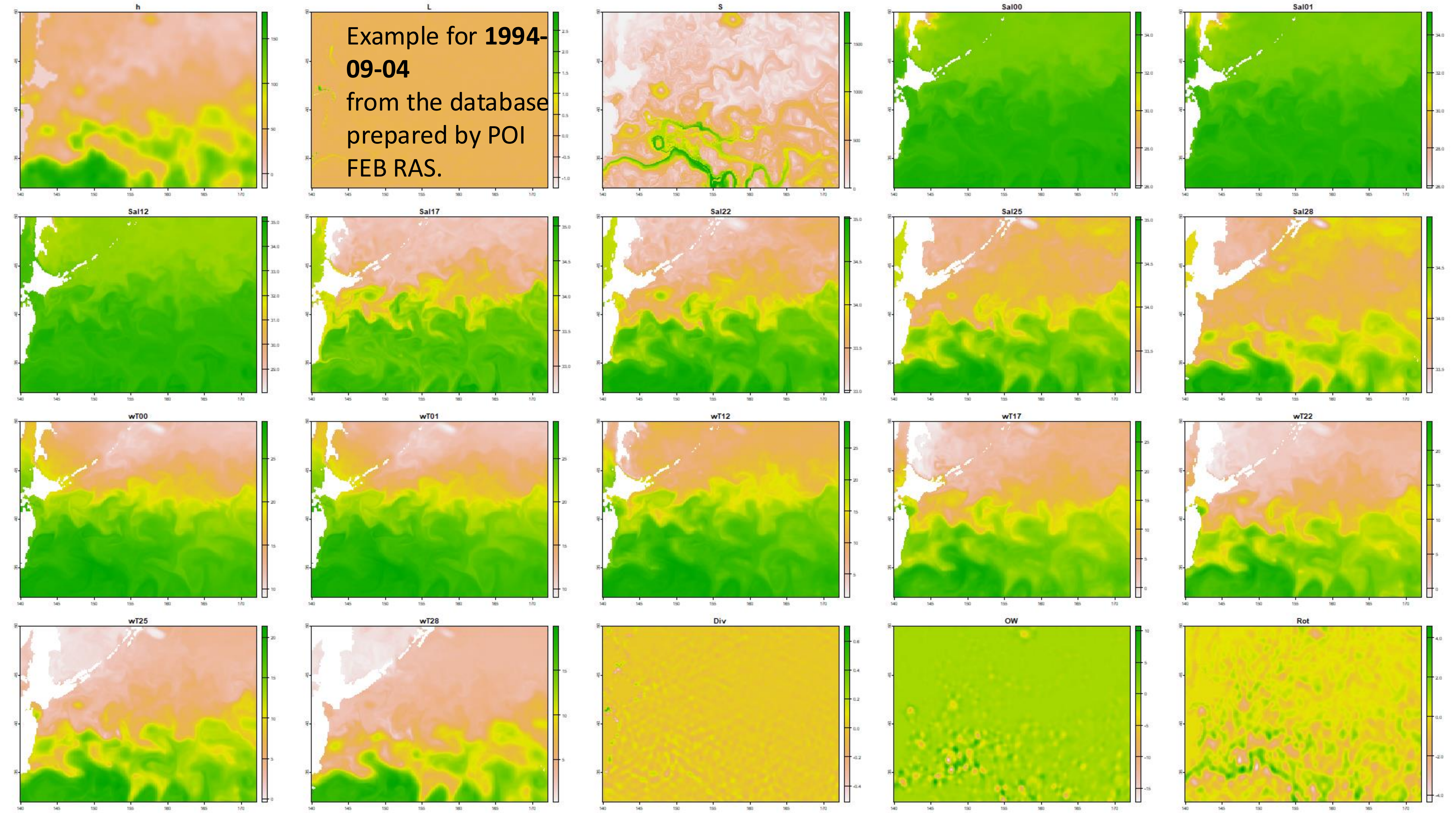
Example for **1994-09-01** from the database prepared by POI FEB RAS.

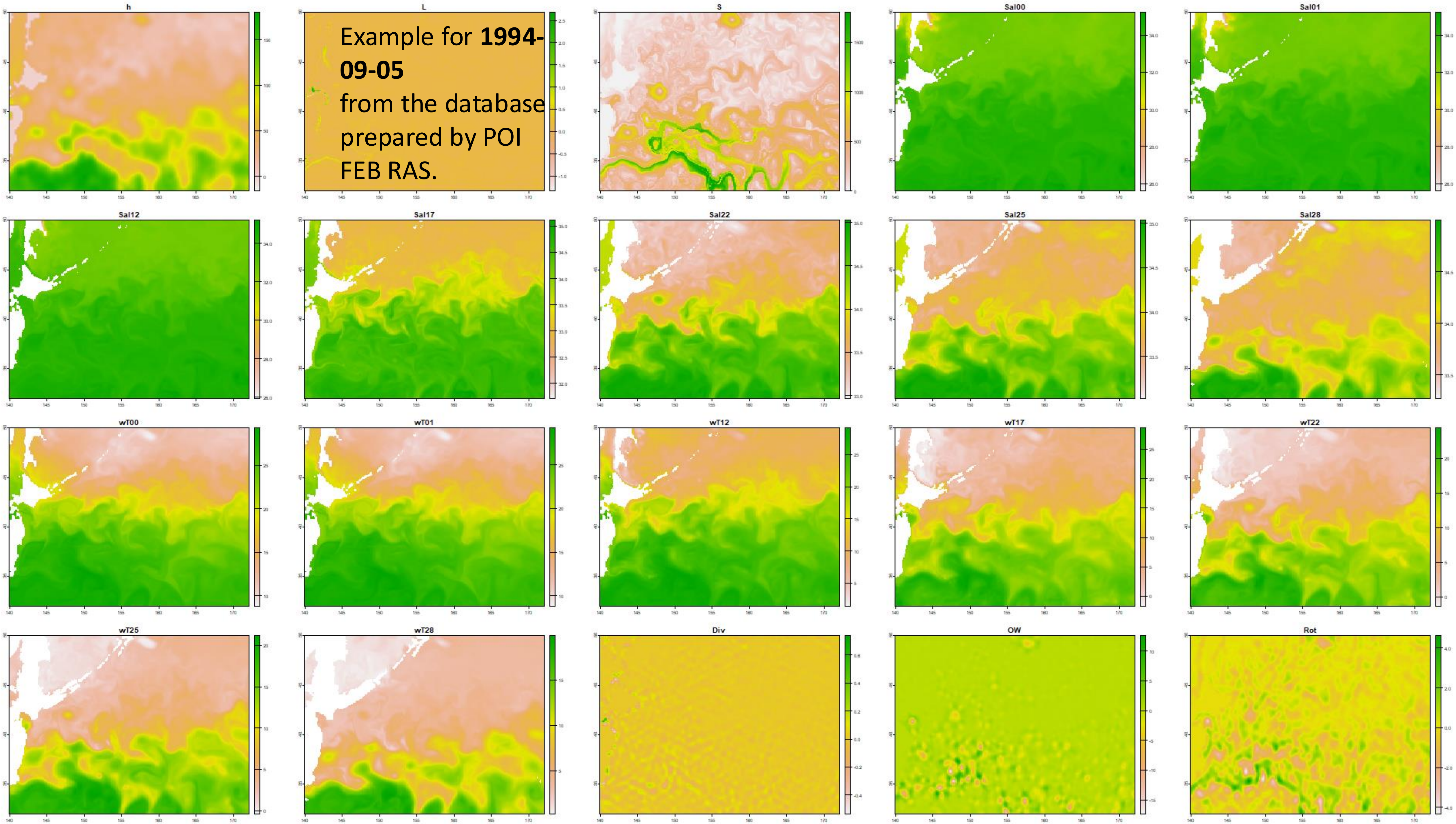



Example for **1994-09-02**
from the database prepared by POI FEB RAS.
h
L
S
Sal00
Sal01
Sal12
Sal17
Sal22
Sal25
Sal28
wT00
wT01
wT12
wT17
wT22
wT25
wT28
Div
OW
Rot



Example for **1994-09-03**
from the database
prepared by POI
FEB RAS.

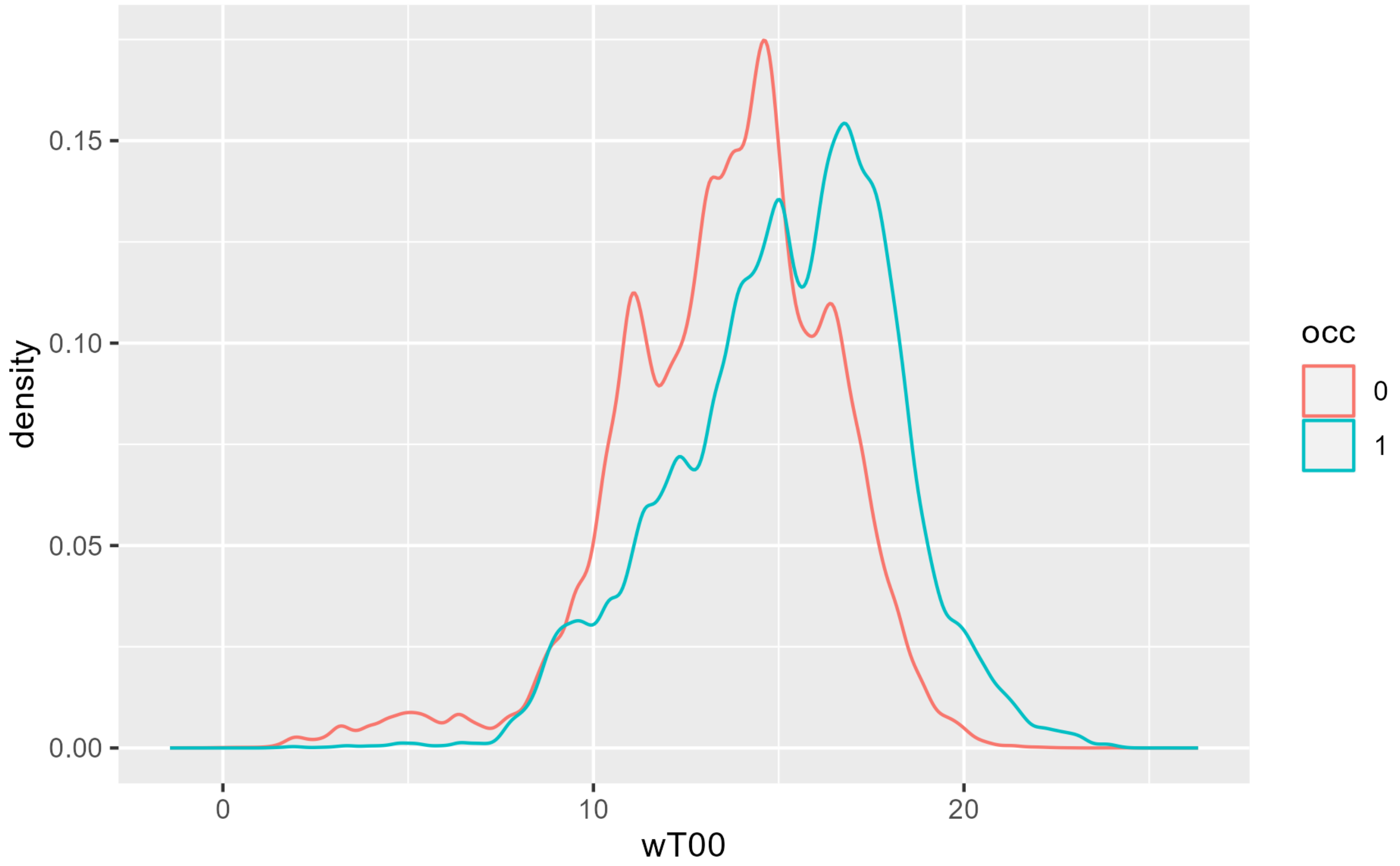

h
L
S
Sal00
Sal01
Sal12
Sal17
Sal22
Sal25
Sal28
wT00
wT01
wT12
wT17
wT22
wT25
wT28
Div
OW
Rot
Example for **1994-09-04** from the database prepared by POI FEB RAS.



Example for **1994-09-05**
from the database prepared by POI FEB RAS.

New data full set. Only SST (wT00) is not enough to split the data

Occurrence of saury	0 (No)	1 (Yes)
Number of points	409615	586229



```r
# Reproducibility
set.seed(1234)

# Divide data into training and test for hyperparameter tuning:
dat_split <- dat %>%
  initial_split(prop = 0.5, strata=occ)

dat_train_df <- training(dat_split)
dat_test_df  <-  testing(dat_split)
```

Table 1: Descriptive Statistics Training Data

Characteristic	Overall, N = 446,040[1]	occ: 0, N = 168,718[1]	occ: 1, N = 277,322[1]	p-value[2]
lon	152.93 (5.63)	150.22 (4.51)	154.59 (5.61)	<0.001
lat	42.99 (1.73)	42.94 (1.62)	43.02 (1.80)	<0.001
Div	0.00 (0.02)	0.00 (0.02)	0.00 (0.02)	<0.001
Dshore	393.41 (285.99)	257.12 (223.05)	476.32 (288.28)	<0.001

[1] Mean (SD)

[2] Welch Two Sample t-test

Table 2: Descriptive Statistics Training Data Lagrangian

		occ		
Characteristic	**Overall**, N = 446,040[1]	**0**, N = 168,718[1]	**1**, N = 277,322[1]	**p-value**[2]
h	32.76 (11.52)	31.60 (10.12)	33.47 (12.25)	<0.001
L	0.07 (0.04)	0.07 (0.04)	0.07 (0.04)	<0.001
S	255.53 (164.33)	246.84 (147.85)	260.83 (173.38)	<0.001

[1] Mean (SD)

[2] Welch Two Sample t-test

Table 3: Descriptive Statistics Training Data Salinity

		occ		
Characteristic	**Overall**, N = 446,040[1]	**0**, N = 168,718[1]	**1**, N = 277,322[1]	**p-value**[2]
Sal00	33.03 (0.56)	32.84 (0.47)	33.14 (0.58)	<0.001
Sal01	33.03 (0.56)	32.84 (0.47)	33.15 (0.58)	<0.001
Sal12	33.11 (0.51)	32.92 (0.42)	33.22 (0.53)	<0.001
Sal17	33.35 (0.44)	33.18 (0.35)	33.46 (0.46)	<0.001
Sal22	33.50 (0.36)	33.39 (0.28)	33.58 (0.39)	<0.001
Sal25	33.62 (0.20)	33.56 (0.17)	33.66 (0.21)	<0.001
Sal28	33.79 (0.12)	33.76 (0.13)	33.81 (0.11)	<0.001

[1] Mean (SD)

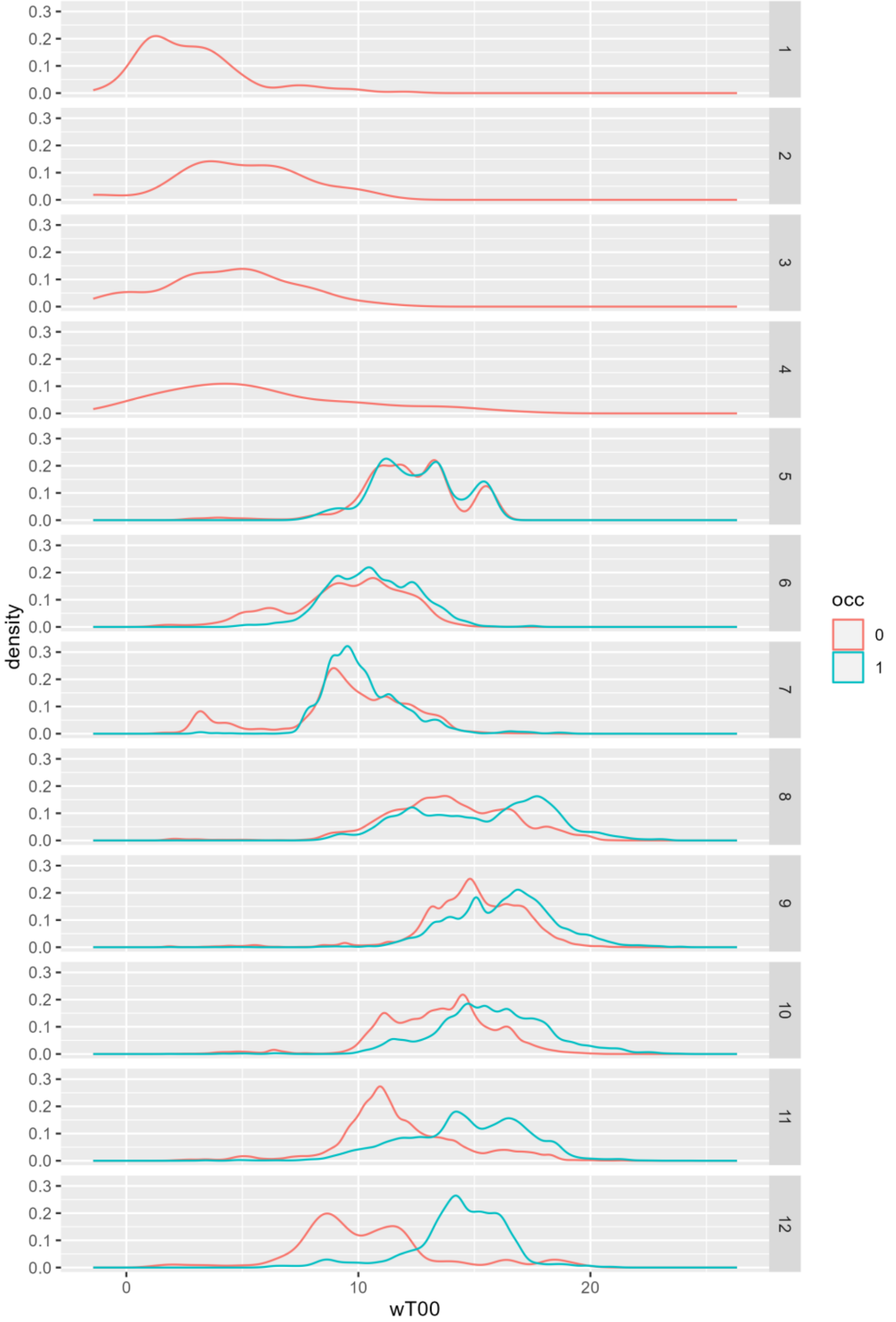
[2] Welch Two Sample t-test

Table 4: Descriptive Statistics Training Data Temperature

		occ		
Characteristic	**Overall**, N = 446,040[1]	**0**, N = 168,718[1]	**1**, N = 277,322[1]	**p-value**[2]
wT00	14.52 (3.10)	13.32 (2.97)	15.25 (2.94)	<0.001
wT01	14.49 (3.09)	13.30 (2.96)	15.22 (2.94)	<0.001
wT12	13.08 (3.34)	11.92 (2.91)	13.78 (3.39)	<0.001
wT17	8.10 (4.32)	6.86 (3.62)	8.86 (4.53)	<0.001
wT22	4.78 (3.33)	3.64 (2.61)	5.48 (3.52)	<0.001
wT25	4.15 (2.26)	3.28 (1.71)	4.69 (2.38)	<0.001
wT28	3.75 (1.16)	3.25 (0.91)	4.05 (1.19)	<0.001

[1] Mean (SD)

[2] Welch Two Sample t-test


density
wT00
occ
0
1

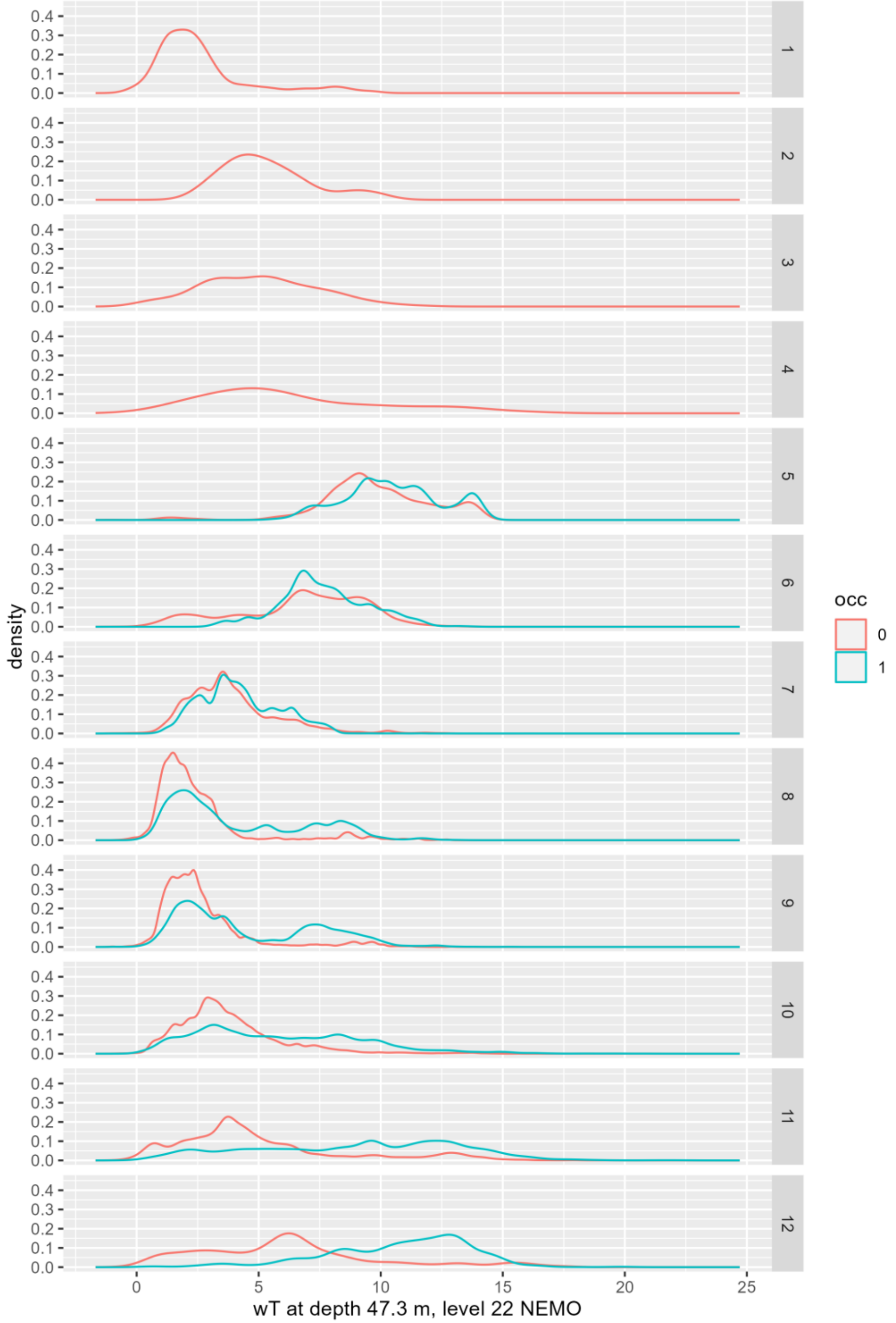
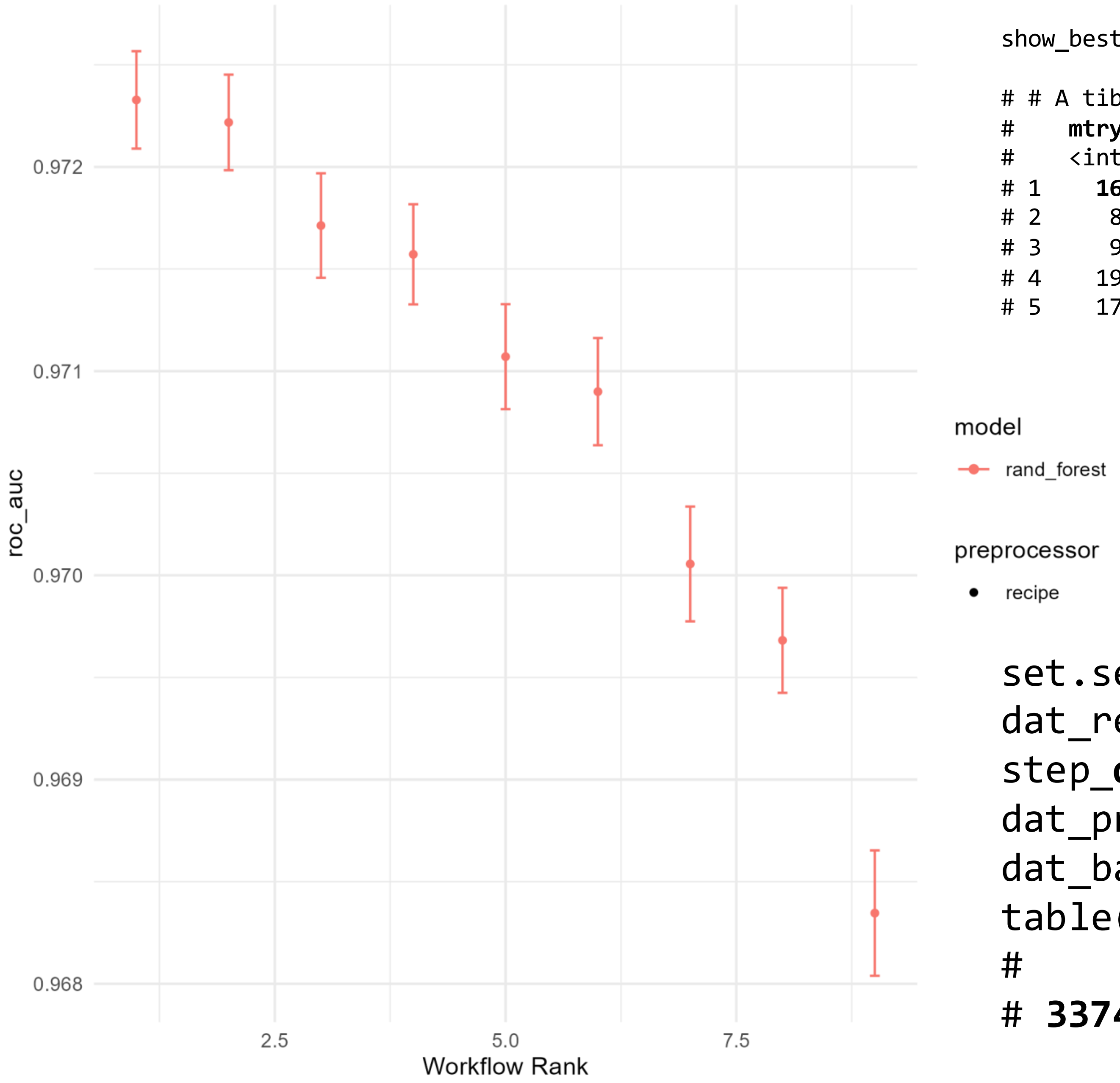

density
wT at depth 47.3 m, level 22 NEMO
occ
0
1

Figure 1: Results Hyperparameter Tuning




```
show_best(RF1_results, metric = "roc_auc")

# # A tibble: 5 × 8
#    mtry min_n .metric .estimator  mean     n  std_err .config
#    <int> <int> <chr>   <chr>      <dbl> <int>    <dbl> <chr>
# 1    16    10 roc_auc binary     0.972    10 0.000145 Preprocessor1_Model4
# 2     8     2 roc_auc binary     0.972    10 0.000142 Preprocessor1_Model8
# 3     9    15 roc_auc binary     0.972    10 0.000156 Preprocessor1_Model9
# 4    19    19 roc_auc binary     0.972    10 0.000149 Preprocessor1_Model3
# 5    17    25 roc_auc binary     0.971    10 0.000156 Preprocessor1_Model1
```

```
set.seed(1001) # FINAL
dat_recipe <- recipe(occ ~ ., data=dat) %>%
step_downsample(occ)
dat_prep <- prep(dat_recipe, dat)
dat_bake <- bake(dat_prep, new_data = NULL)
table(dat_bake$occ)
#      0      1
# 337436 337436
```

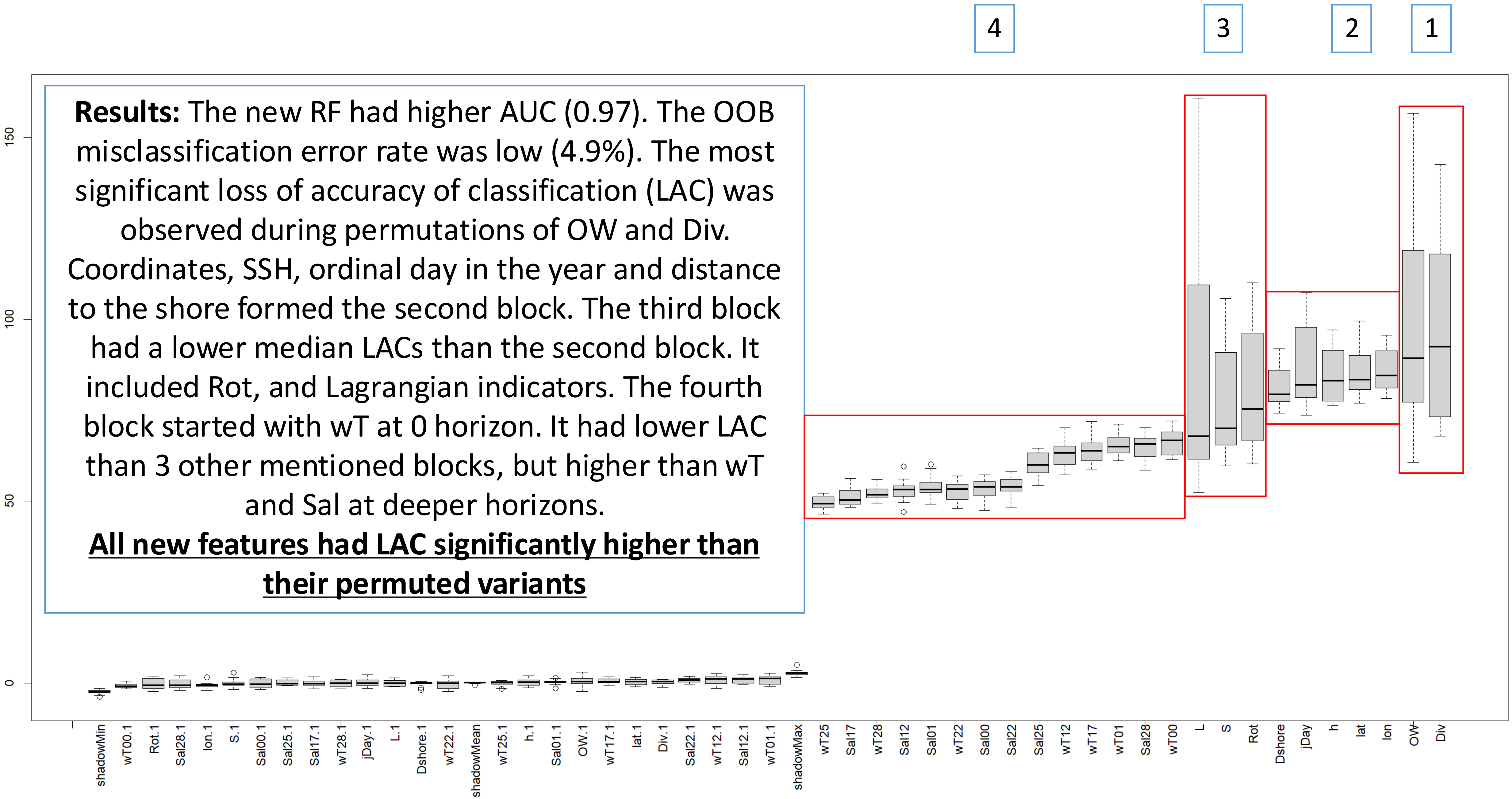
Final method – Random forest

```
set.seed(1001) # FINAL
dat_recipe <- recipe(occ ~ ., data=dat) %>%
step_downsample(occ)
dat_prep <- prep(dat_recipe, dat)
dat_bake <- bake(dat_prep, new_data = NULL)
table(dat_bake$occ)
#      0      1
# 337436 337436
```

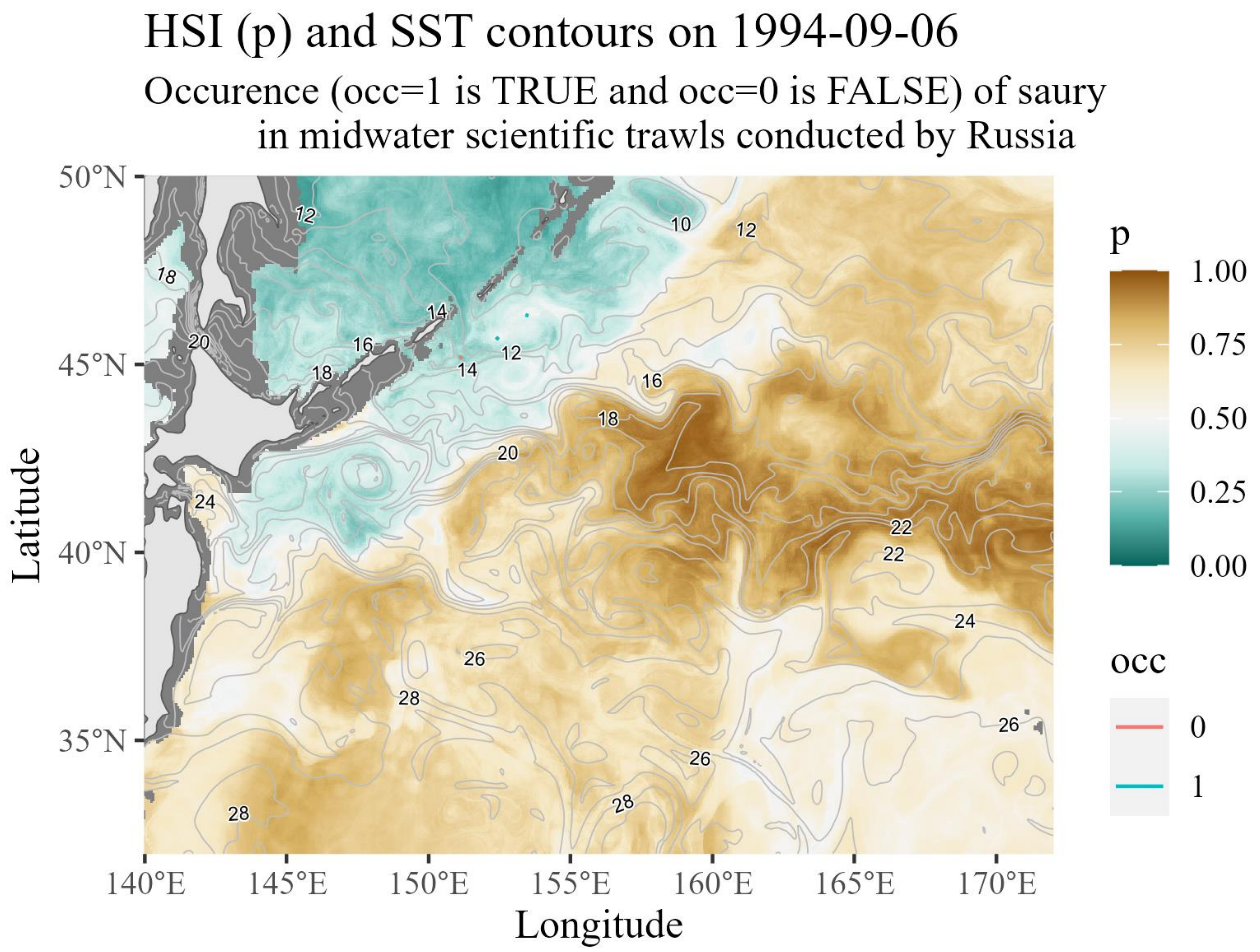
```
# Type:                              Probability estimation
*
# Number of trees:                   500
# Sample size:                       674872
# Number of independent variables:   24
# Mtry:                              16
# Target node size:                  10
# Variable importance mode:          none
# Splitrule:                         extratrees
# Number of random splits:           1
# OOB prediction error (Brier s.):   0.04903609
# *Malley, J. D., Kruppa, J., Dasgupta, A., Malley, K. G.,
& Ziegler, A. (2012). Probability machines: consistent
probability estimation using nonparametric learning
machines. Methods Inf Med 51:74-81. doi:10.3414/ME00-01-
0052.
```

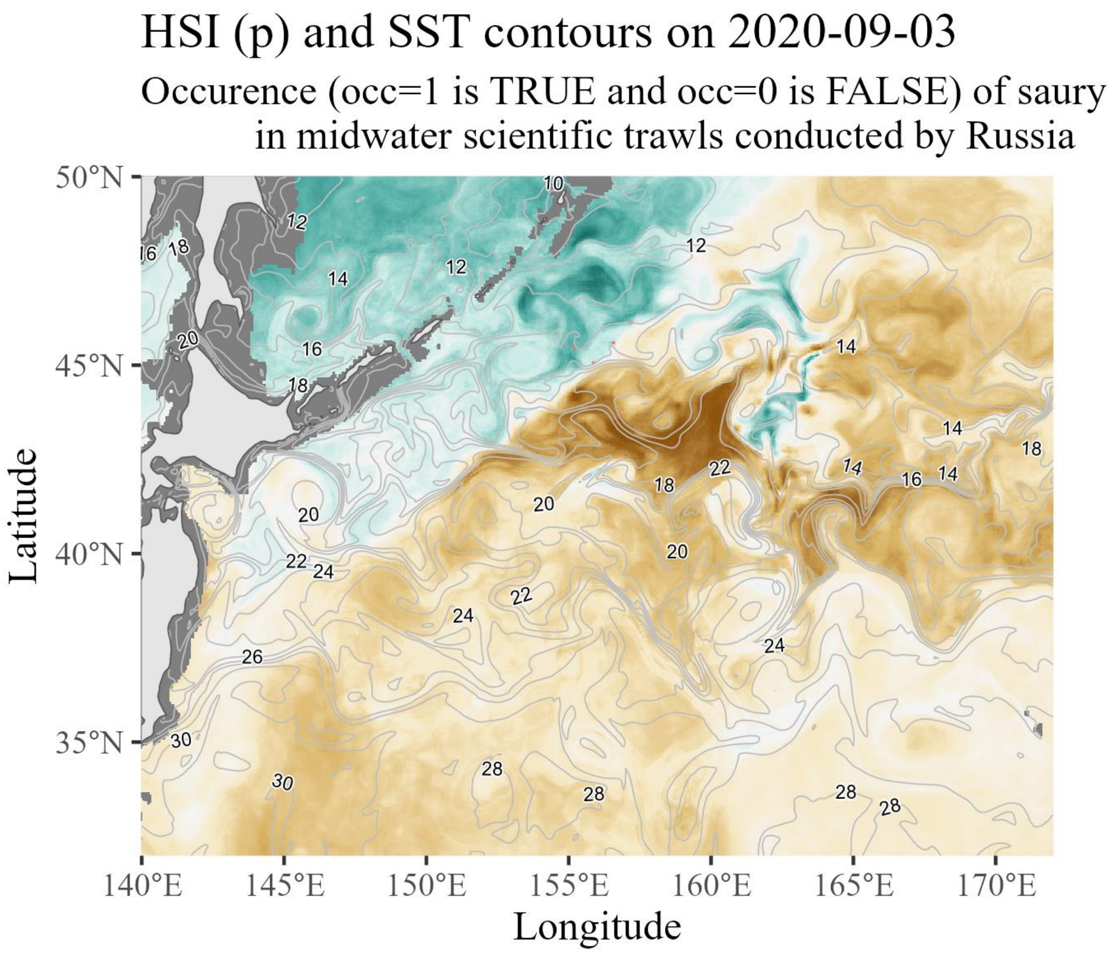
Results: The new RF had higher AUC (0.97). The OOB misclassification error rate was low (4.9%). The most significant loss of accuracy of classification (LAC) was observed during permutations of OW and Div. Coordinates, SSH, ordinal day in the year and distance to the shore formed the second block. The third block had a lower median LACs than the second block. It included Rot, and Lagrangian indicators. The fourth block started with wT at 0 horizon. It had lower LAC than 3 other mentioned blocks, but higher than wT and Sal at deeper horizons.

<u>**All new features had LAC significantly higher than their permuted variants**</u>



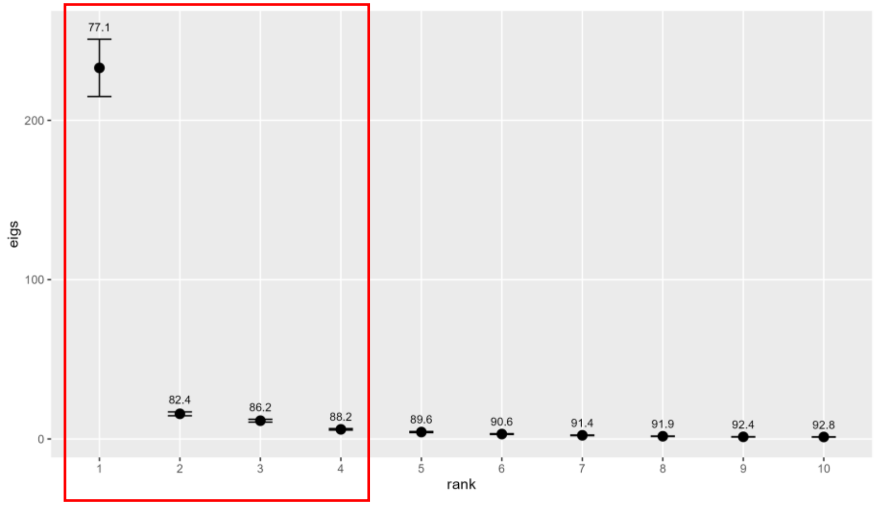
Random forest predictions

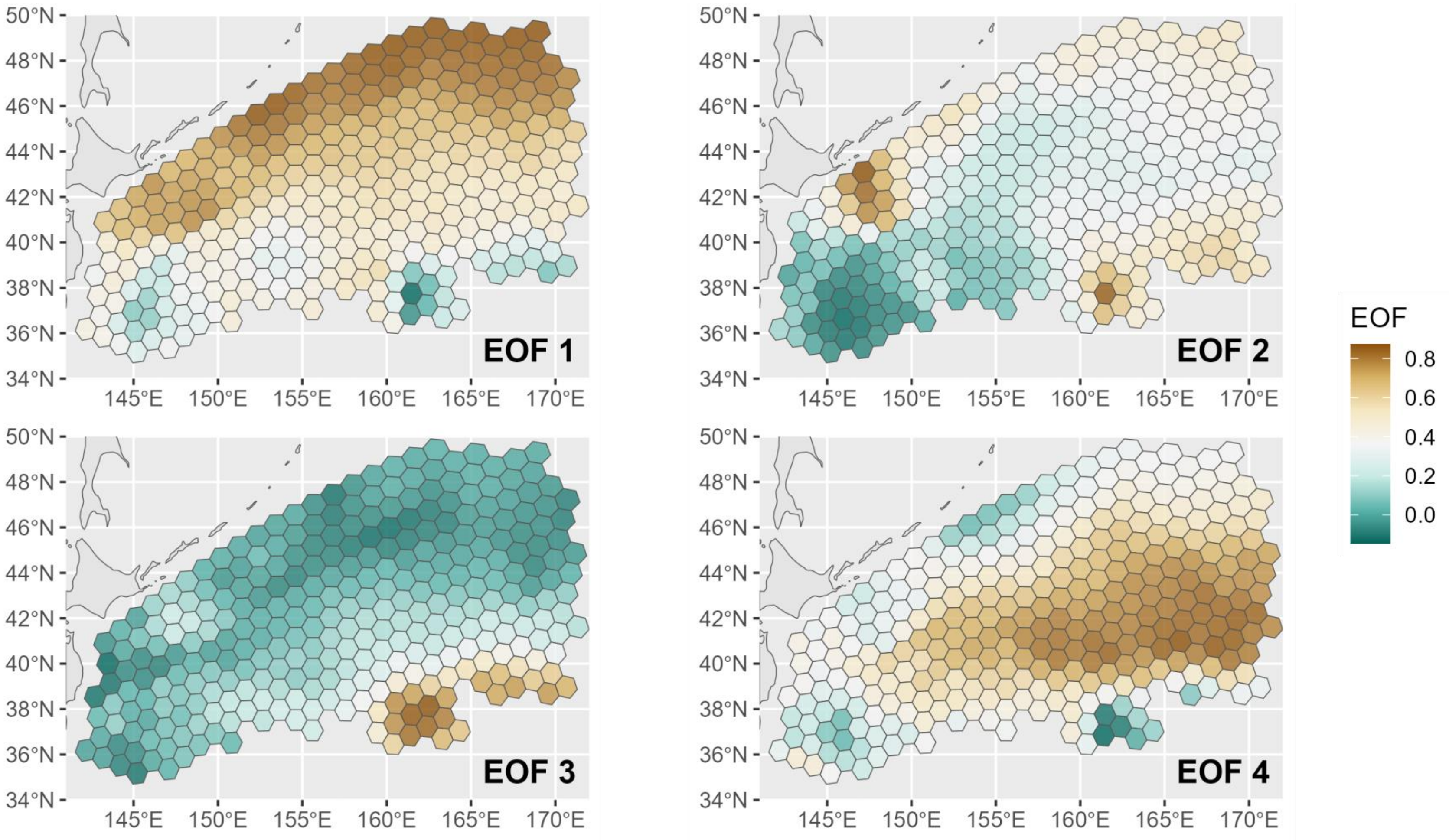




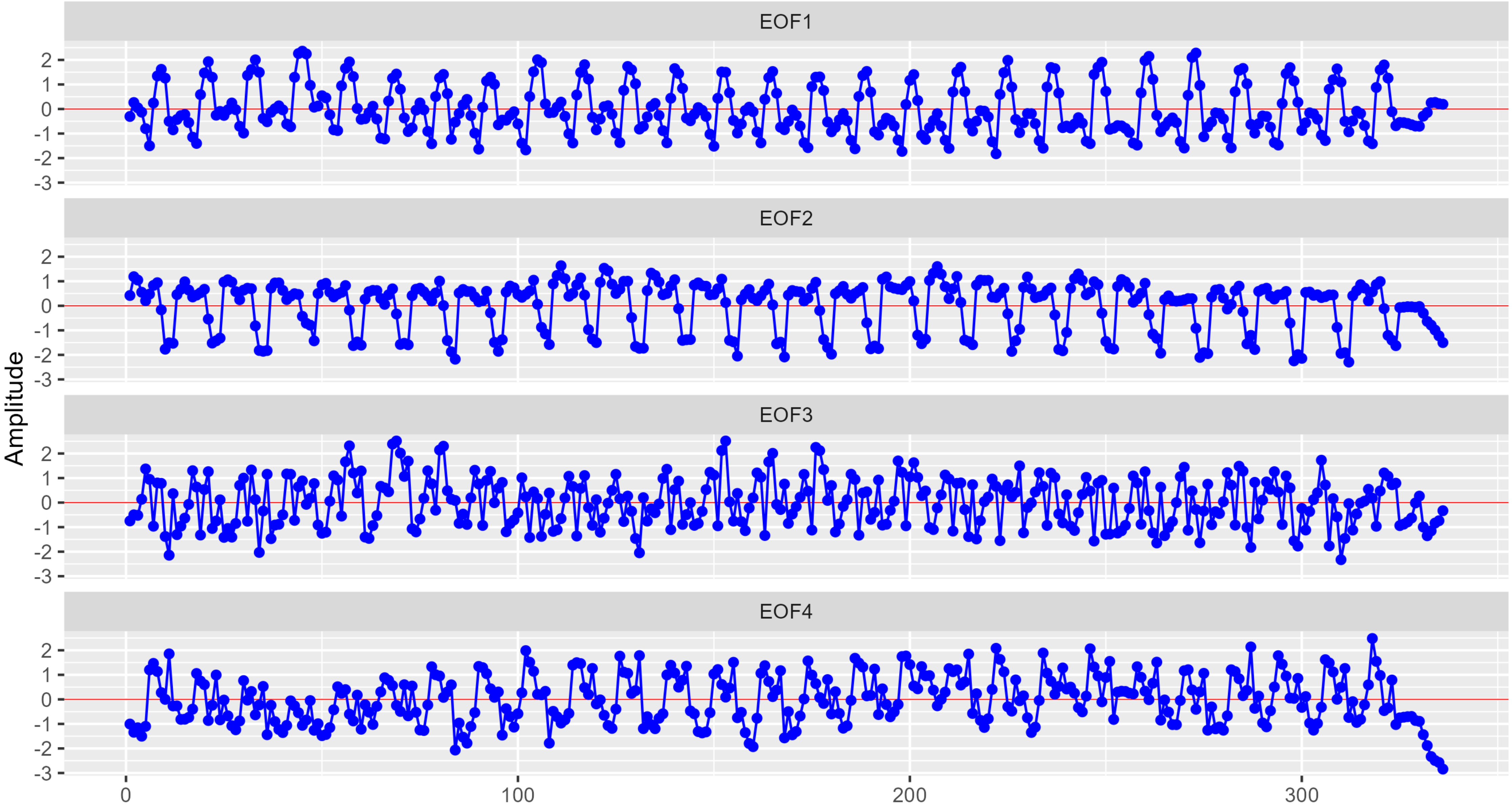
EOF on Random forest predictions averaged by 302 equal area hexagons (8660 km^2 , 100 km height, 115 km width each) for each year and month

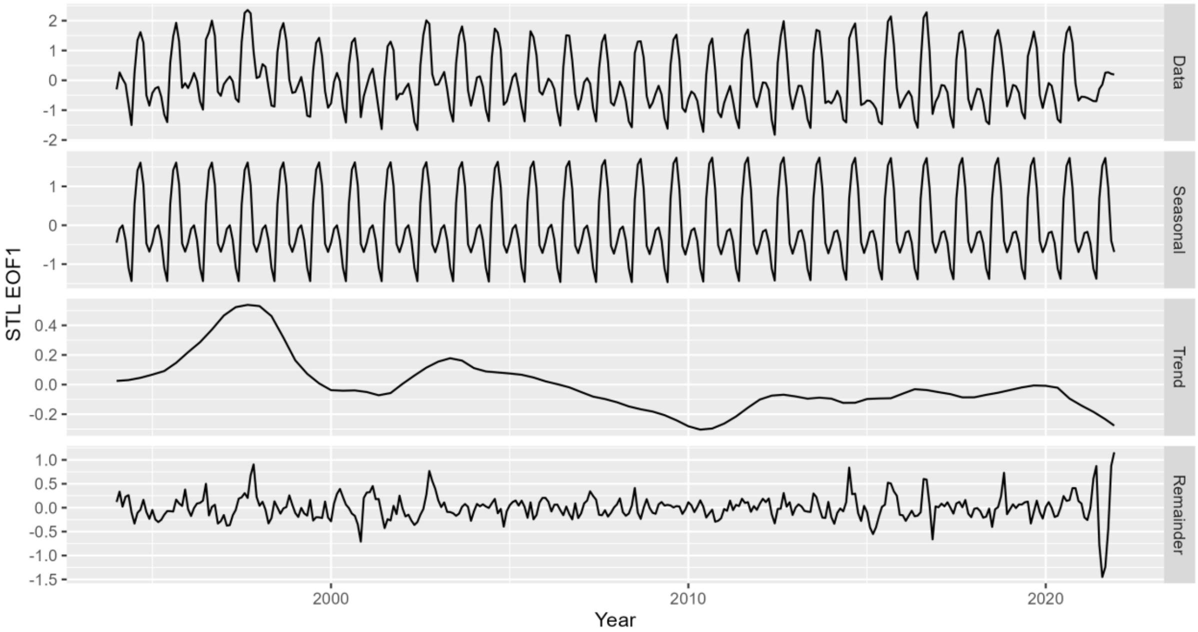
4 EOFs captured > 88% of variance

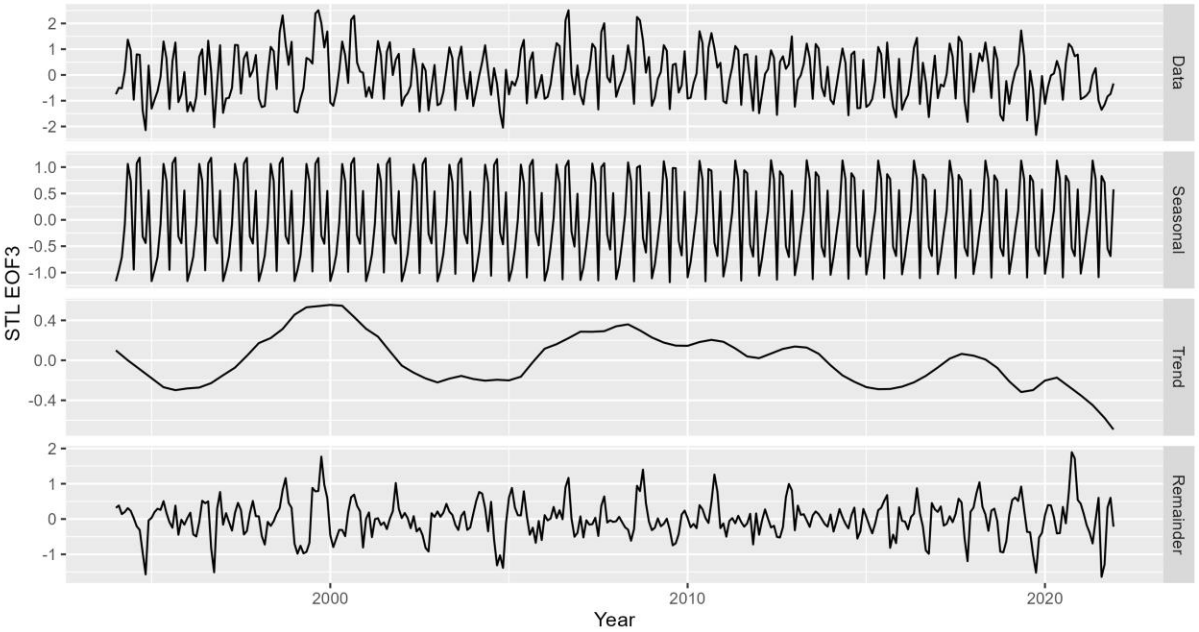


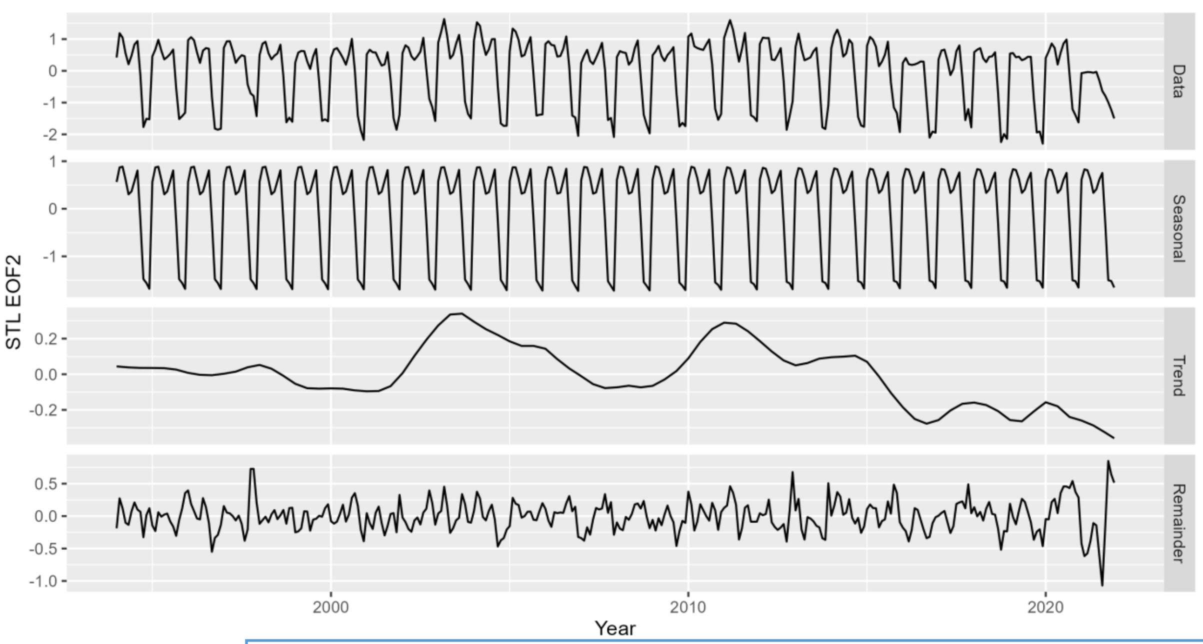


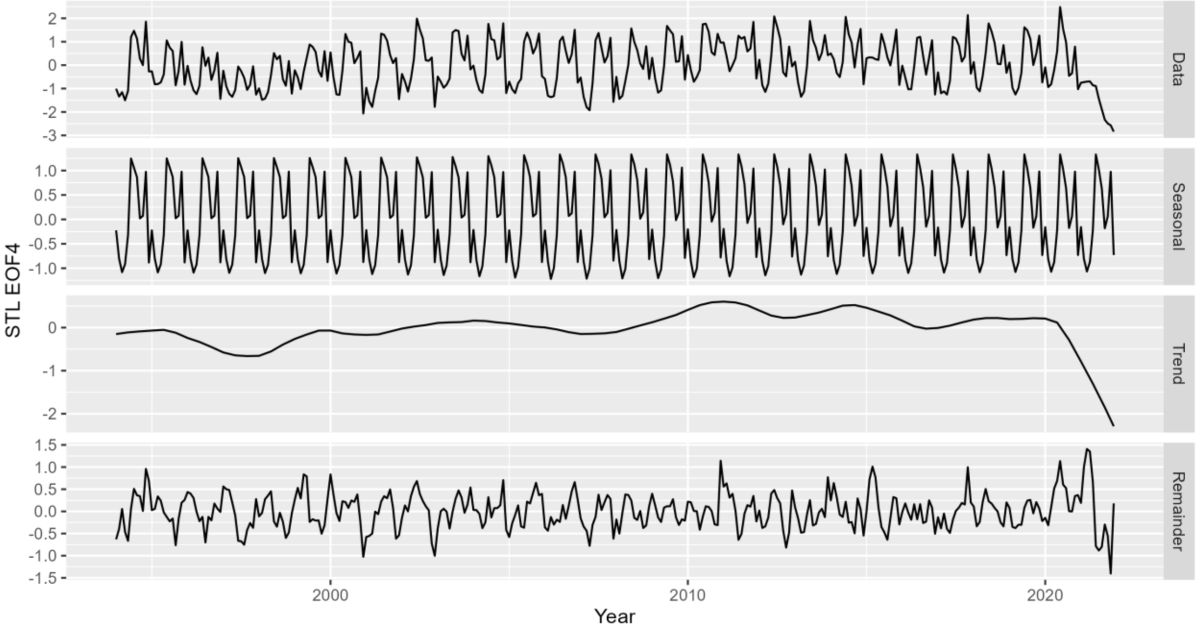
step = 1 month



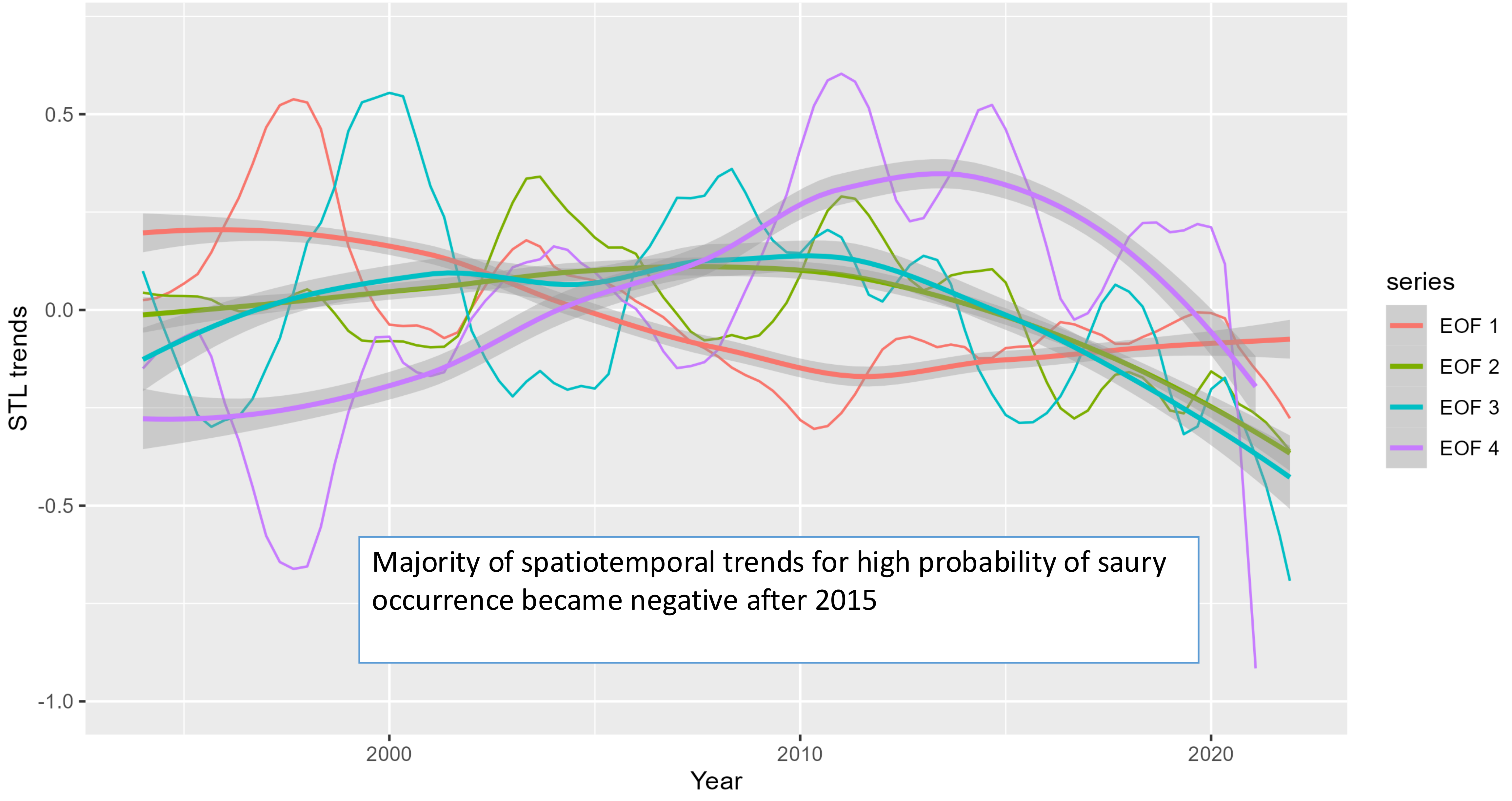








STL (Cleveland et al., 1990) Swindow=25; Twindow=37; step = 1 month


0.5
0.0
-0.5
-1.0
STL trends
2000
2010
2020
Year
series
EOF 1
EOF 2
EOF 3
EOF 4
Majority of spatiotemporal trends for high probability of saury occurrence became negative after 2015

Conclusion

We reached the goal of our study (improvement of the SDM for saury). Published AUC was 0.85 while new SDM reached AUC=0.97. Misclassification error rate of the new SDM is low (4.9%).

Majority of spatiotemporal trends for high probability of saury occurrence became negative after 2015. But we can not update our SDM and test it against new years after 2021, because GLORYS12V1 became unavailable in Russia.

So, we still hope to cooperate with PICES scientists to extend the range of study in geographical space more into the High Seas.
We also would be very glad to test other predictor variables than we did.

Thank you for your attention!