

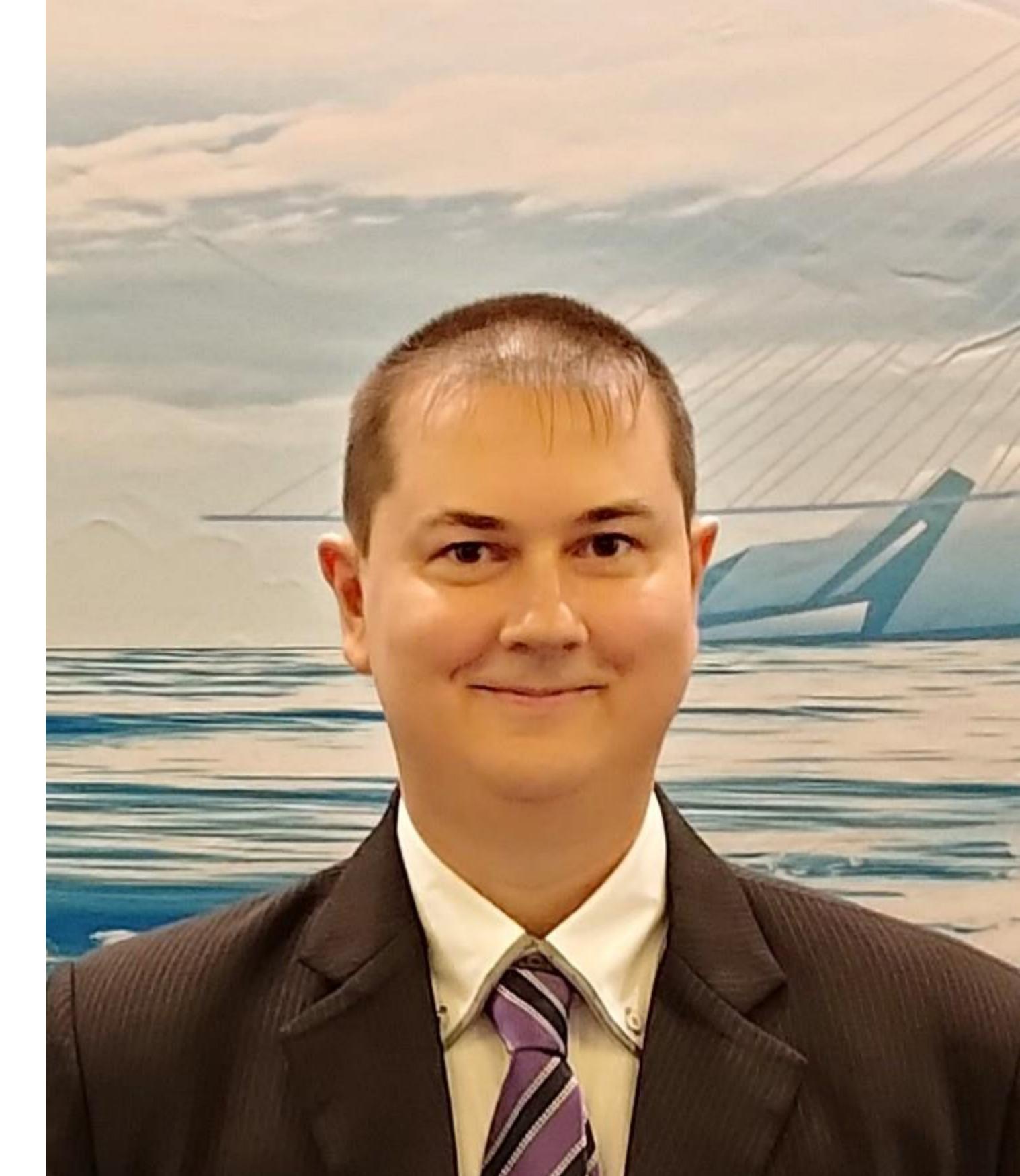
# Species distribution modeling: the importance of zero catches in fishery for the chub mackerel (*Scomber japonicus*)

Vladimir Kulik<sup>1</sup>, Aleksei Baitaliuk<sup>1,2</sup>, Oleg Katugin<sup>1</sup>, Maxim Budyansky<sup>3</sup> and Michael Uleysky<sup>3</sup>

<sup>1</sup> Pacific branch of "VNIRO" ("TINRO"), Vladivostok, Russia E-mail: vladimir.kulik@tinro.vniro.ru

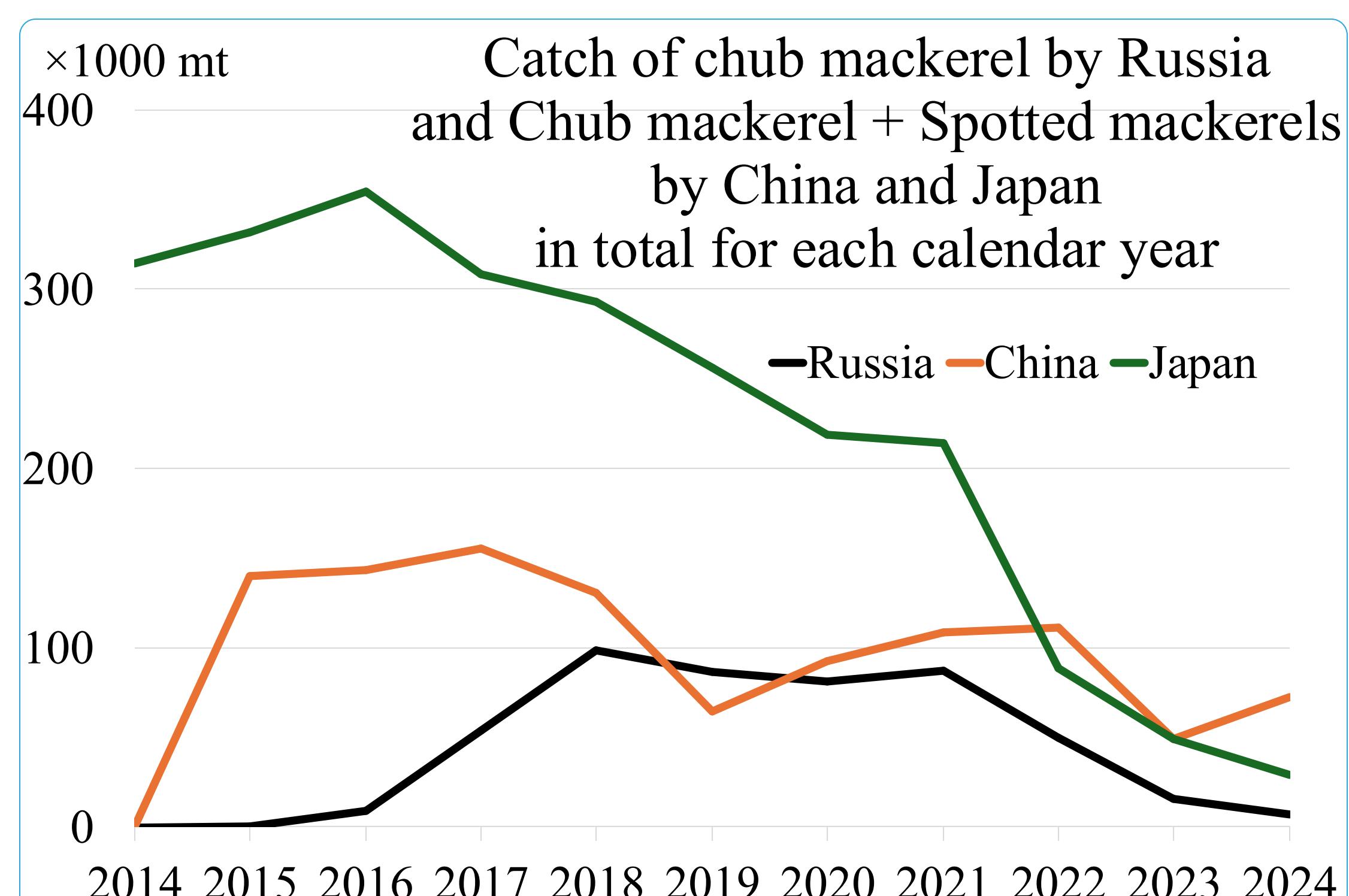
<sup>2</sup> Russian Federal Research Institute of Fisheries and Oceanography "VNIRO", Moscow, Russia

<sup>3</sup> V.I. Il'ichev Pacific Oceanological Institute of the Far-Eastern Branch of the Russian Academy of Sciences "POI FEBRAS"



## Introduction

Catches of the chub mackerel by Russian fleet for the last 10 years sharply increased to a maximum of 98,812 mt in 2018 but then declined to 7,233 mt in 2024.



Different SDMs showed very poor performances in the Russian studies, when SDMs were tuned using pseudo absence information. The best published SDM achieved AUC=0.69, on average, growing up to 0.75 during the period of high catches in 2020 (Chernienko, Chernienko, 2021).

## Methods

Low number of minority class (only 465 tracks) does not allow us to use down sampling to it for resolution of imbalance ratio problem. Therefore, we decided to use the geometric mean (Gmean) of true positive rate (TPR) and the true negative rate (TNR):  $Gmean = \sqrt{TPR \times TNR}$  for maximization in the Random Forest implementation of the q-classifier (RFQ) for the two-class imbalance problem (Ishwaran et al., 2021). The Gmean is close to 1 when both the TNR and TPR are close to 1 and the difference between the two is small (Kubat et al., 1997).

Replacement of pseudo absence places with real absence of chub mackerel in catches during midwater trawl surveys in Russian national waters and in High Seas increased **AUC of SDM up to 0.99 (Gmean = 0.96)** for the period when accurate information on each commercial operation was available (between February 2021 and May 2023). Misclassification error rate of occurrence and absence probability were very low, of 4.2% and 4.1%, respectively. The most important predictor was wt1 followed by the Sea surface temperature and salinity. Lagrangian indicators: L and S had lower effect on the accuracy loss than predictors related to productivity (PrP, Chl, Si, NO3, PO4), but appeared higher than other speed-based predictors (W, Div, OW, and Rot). Nevertheless, all predictors were statistically significant. **We need an external validation of our SDM to test its performance in areas where chub mackerel catches are high, and where Russian fleet operates occasionally (in the High Seas).** This step is impossible without merging Russian data with data on chub mackerel catch by China and Japan. **That is where PICES is a good place for collaboration.**

## References

Chernienko, E.P., Chernienko, I.S. (2021). Information support for chub mackerel *Scomber japonicus* fishery in the Pacific waters of the Russian Federation. *Izvestiya TINRO*. 201(2):390-399. (In Russ.) <https://doi.org/10.26428/1606-9919-2021-201-390-399>

Ishwaran, H., O'Brien, R., Lu, M., & Kogalur, U. B. (2021). "randomForestSRC: random forests quantile classifier (RFQ) vignette." <http://randomforestsrc.org/articles/imbalance.html>

Kubat, M., Holte, R., Matwin, S. (1997). Learning when negative examples abound. In: van Someren, M., Widmer, G. (eds) Machine Learning: ECML-97. Lecture Notes in Computer Science, vol 1224. Springer, Berlin, Heidelberg, p. 146–153. [https://doi.org/10.1007/3-540-62858-4\\_79](https://doi.org/10.1007/3-540-62858-4_79)

## Materials

### Response – binary variable.

Each scientific trawl track stored in TINRO with occurrence or absence of Chub mackerel and each track targeted for it in commercial fishing available in Electronic Fishing Journal in the Federal government-financed institution "Centre of Fishery Monitoring and Communications" for the period between February 2021 and May 2023 were used. In total we selected 465 tracks with absence of Chub mackerel and 2049 tracks with occurrence.

**Predictors.** jDay – the number of a day in a year [1-365],

Lagrangian indicators calculated in POI FEBRAS:

L — Lyapunov exponent and S — the length of passive tracers' trajectories in km back calculated in time for 15 days,

Global Ocean Physics Reanalysis (GLORYS12V1):

W — vertical component of velocity,

Div — divergence,

OW — the Okubo–Weiss parameter,

Rot — rotor,

PrP — Net primary production of biomass expressed as carbon per unit volume in sea water (mg/m<sup>3</sup>/day),

Chl — Mass concentration of chlorophyll *a* in sea water (mg/m<sup>3</sup>),

Si — Mole concentration of silicate in sea water (mmol/m<sup>3</sup>),

h and SSH — sea surface height from different sources (altimetry and model),

NO3 — Mole concentration of nitrate in sea water (mmol/m<sup>3</sup>),

PO4 — Mole concentration of phosphate in sea water (mmol/m<sup>3</sup>),

MLD — mixed layer thickness (m),

Sal0 — sea surface salinity, Sal1 — water salinity at the 1 level,

wT0 — sea surface temperature, wT1 — water temperature at the 1 level.

HYCOM: Sal00 — sea surface salinity and wT00 — sea surface temperature.

## Results and discussion

