Use of data science and machine learning techniques for study Sea lice (*Caligus rogercresseyi*) infestation on Atlantic salmon (*Salmo salar*)

De los Ríos-Escalante P.[1,2*]; Ibáñez-Arancibia E.[1,3]

Received: April 2023 Accepted: July 2023

Abstract

Salmon farming, mainly Atlantic salmon (*Salmo salar*) is the main productive activity in Chilean Patagonia (38-53°S), one of the main problems for *Salmo salar* farming is the infestation of sea lice *Caligus rogercresseyi*. The aim of the present study was to analyze the infestation rate of sea lice *Caligus rogercresseyi* on Salmo salar farmed in the Aysen region in central Chilean Patagonia (43-50° S). The results revealed the existence of a weak but not significant relation between latitude and infestation rate, whereas it was found inverse direct associations between temperature and salinity with infestation rate. The possible cause would be due in Southern latitudes, the temperature and salinity decrease, that are conditions that limit the infestation rate of *Caligus rogercresseyi* on Salmo salar in Southern Chile. The exposed results would be similar with literature descriptions, and would indicate that use of data science and machine learning can be a powerful tool for study of *Caligus rogercresseyi* infestation on Chilean farmed salmonids.

Keywords: *Salmo salar*, *Caligus rogercresseyi*, Machine learning, Parasites, Chile

1-Departamento de Ciencias Biológicas y Químicas, Facultad de Recursos Naturales, Universidad Católica de Temuco, Temuco, Chile

2-Núcleo de Estudios Ambientales, Universidad Católica de Temuco, Temuco, Chile

3-Programa de Doctorado en Biología Celular y Molecular Aplicada, Universidad de la Frontera Temuco, Chile

*Corresponding author's Email: prios@uct.cl

Introduction

The salmon farming is an important economic activity in Chilean Patagonia, mainly in Lakes, Aysen and Magallanes region where are located the main aquaculture sites, where *Salmo salar* is the main farmed species (Soto *et al*., 2019).

The sea lice infestation by *Caligus rogercreseyi*, Boxshall and Bravo, 2000 is the main problem for salmon farming, this species is an ectoparasite that affects the skin of fishes that in consequence generates decreasing in economical inputs (Boxshall and Bravo, 2000; Bravo *et al*., 2009; 2013; 2014; De los Rios, 2019). In this scenario is necessary control management for decrease the infestation rate of this plague on farmed fishes (González *et al*., 2016), that is the cause of control from National Fisheries and Aquaculture Service (SERNAPESCA-Chile), that management large data volumes of sea lice infestation in farmed fishes.

On the basis of large data volume of sea lice infestations, it would be possible apply "machine learning" techniques for study these data with an interdisciplinary viewpoint, with the aim of obtain patterns in large volume data using computation programming and statistic techniques (VanderPlas, 2017). The aim of the present study is to analyze using "machine learning" techniques a large data volumen of sea lice infestations in *S. salar* farmed in Aysen region (43-46º S), in Chilean Patagonia based in information provided by National Fisheries and Aquaculture Service (SERNAPESCA-Chile)

Materials and methods

Sea lice infestation (juvenile, ovigerous females, males and total adults) on *S. salar* from Aysen region (43-46°S) between 2018-2019, also these data included temperature, salinity and geographical location of different farmed sites, these data were obtained from National Fisheries and Aquaculture Service (SERNAPESCA-Chile)

These data were imported and processed using Machine learing techniques using the software Python (Van Rossum and Drake, 1995), with libraries numpy (Harris *et al*., 2020), pandas (McKinney, 2010), matplotlib (Hunter, 2007), seaborn (Waskom, 2021) and statsmodel (Seabold and Perktold, 2010), to these data was applied a multiple regression analysis (Halswanter, 2016). To this regression analysis was included as independent variables geographical location, temperature, salinity, and as independent variables the total sea lice infestation for each stage (juvenile, ovigerous females, males and total adults).

Results and discussion

The results of multiple regression model denoted significant inverse associations with latitude and temperature, and significant direct associations with longitude and salinity for each sea lice life stages (Table 1).

The results revealed of low infestation rate at high latitudes and longitude would be due oceanographic conditions that involves low salinity and temperature. The infestation rate is important because high infestation rate

can be vector for viral and bacterial pathogens (Oelkers *et al*., 2014; Labra *et al*., 2020).

Table 1: Results of multiple regression analysis for sea lice (total, male, female, juvenile) infestation for farmed *S. salar* in Aysen region, Chile

Sea lice total				Female sea lice			
Factor	**Coefficient**	**T observed**	**P**	**Factor**	**Coefficient**	**T observed**	**P**
Latitude	-0.3463	-5.618	< 0.001*	Latitude	-0.748	-9.431	< 0.001*
Longitude	0.7717	9.522	< 0.001*	Longitude	1.572	15.079	< 0.001*
Temperature	-0.1408	-7.146	< 0.001*	Temperature	-0.202	-7.967	< 0.001*
Salinity	0.0757	8.851	< 0.001*	Salinity	0.077	7.059	< 0.001*
R^2	0.085	F observed	153.900	R^2	0.121	F observed	228.600
R^2 adjusted	0.084	P	< 0.001*	R^2 adjusted	0.120	P	< 0.001*

Male sea lice				Juvenile sea lice			
Factor	**Coefficient**	**T observed**	**P**	**Factor**	**Coefficient**	**T observed**	**P**
Latitude	-0.748	-9.431	< 0.001*	Latitude	-1.119	-6.740	< 0.001*
Longitude	1.572	15.079	< 0.001*	Longitude	4.415	20.230	< 0.001*
Temperature	-0.203	-7.967	< 0.001*	Temperature	-0.452	-8.526	< 0.001*
Salinity	0.078	7.059	< 0.001*	Salinity	0.005	0.248	< 0.001*
R^2	0.121	F observed	228.600	R^2	0.109	F observed	203.300
R^2 adjusted	0.120	P	< 0.001*	R^2 adjusted	0.108	P	< 0.001*

Nevertheless, the present data did not include information of size, biomass and reproductive stage of infested fishes, that probably can affect the results (Gonzalez *et al*., 2020; Hemingsen *et al*., 2020; Montory *et al*., 2020). On these antecedents, it is necessary implement of sea lice infestation management on farmed salmonids that involved chemical agents as well as molecular applications (Valenzuela-Muñoz *et al*., 2020).

On the basis of the important role of salmon farming on Chilean economy (Soto *et al*., 2019), the use of large data volume of sea lice infestation would be an important tool for management. The present study was based between 43-46°S, whereas the salmon farming centers are located between 40-53°S (Soto *et al*., 2019; Valenzuela-Muñoz *et al*., 2020). The present study although included a fraction of total available data volume, the use of Data Sciences and Machine Learning techniques would be an important study that can be applicable for other activities related with natural resources management and ecology (Humphries *et al*., 2018). On these antecedents, the use of these techniques for study sea lice infestation would confirm the antecedents described by literature and would be interesting the scale of available data for get more robust models.

Acknowledgements

The present study was financed by project MECESUP UCT 0804, the authors express their gratitude to National Fisheries and Aquaculture Service (SERNAPESCA-Chile), and

M.I., and S.M.A., for their valuable suggestions for improve the manuscript.

References

Boxshall, G.A. and Bravo, S., 2000. On the identity of the common *Caligus* (Copepoda: Siphonostomatoida: Caligidae) from salmonid net pen systems in southern Chile. *Contributions to Zoology (Amsterdam, Netherlands)*, 69, 137-146.

Bravo, S., Erranz, F. and Lagos C., 2009. A comparison of sea lice, *Caligus rogercresseyi*, fecundity in four areas in southern Chile. *Journal of Fish Diseases* 32, 107–113. DOI: 10.1111/j.1365-2761.2008.01012.x

Bravo, S., Pozo, V., Silva, M.T. and Abarca, D., 2013. Comparison of the fecundity rate of *Caligus rogercresseyi* infesting Atlantic salmon (*Salmo salar* L.) on farms in two regions of Chile. Aquaculture 404/405, 55-58. DOI: 10.1016/j.aquaculture.2013.04.002

Bravo, S., Silva, M.T. and Treasurer, J., 2014. Factors affecting the abundance of *Caligus rogercresseyi* (Boxshall and Bravo) on farmed salmonids in Chile in the period 2006–2007. *Aquaculture* 434, 456-461. DOI: 10.1016/j.aquaculture.2014.09.009

De Los Rios, P., 2019. A checklist of marine parasite copepods of the genus *Caligus* (O.F. Müller, 1785) and their respective marine fish hosts in Chile. *Proceedings of the Biological Society of Washington*, 132, 65–72. DOI:10.2988/PBSW-D-19-

Gonzalez, M.P., Muñoz, J.L.P., Valerio, V. and Vargas-Chacoff, L., 2016. Effects of the ectoparasite *Caligus rogercresseyi* on Salmo salar blood parameters under farm conditions. *Aquaculture*, 457, 29-34. 01.027. DOI:10.1016/j.aquaculture.2016.01.027

González-Gómez, M.P., Ovalle, L., Spinetto, C., Oyarzo, C., Oyarzun, R., Menanteau, M., Alvarez, D., Rivas, M. and Olmos, P., 2020. Experimental transmission of Caligus rogercresseyi between two different fish species. *Diseases of Aquatic Organisms*, 141, 127-138. DOI: 10.3354/dao03513

Halswanter, T., 2016. An introduction to statistics with Python. With applications in the life sciences. Springer International Publishing Switzerland, 278P.

Harris, C.R., Millman, K.J., van der Walt, S., Gommers, R., Virtalen, P., Cournapeau, D., Wieser, E., Taylor J., Berg, S., Smith, N.J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M.H., Brett, M., Haldane, A., Fernandez del Río, J., Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppars, K., Reddy, T., Warren Weckesser, Abbasi, H., Gohlke, C. and Oliphant, T.E., 2020. Array programming with NumPy. *Nature*, 585, 357-362.

Hemmingsen, W., Mackenzie, K., Sagerup, K., Remen, M., Bloch-Hansen, K. and Dagbjartarson Imsland, A.K., 2020. *Caligus elongatus* and other sea lice of the genus *Caligus* as parasites of farmed salmonids: a review. *Aquaculture* 522, 735160. DOI:10.1016/j.aquaculture.2020.735160

Humphries, G., Magness D.R. and Huettmann F., (Eds) 2018. Machine learning for ecology and sustainable natural resource management. Springer Nature Switzerland, A.G., 441P.

Hunter, J.D., 2007. Matplotlib: A 2D graphics environment. *Computer Science and Engineering*, 9, 90–95. DOI:10.1109/MCSE.2007.55

Labra, A., Bravo, S. and Marshall, S.H., 2020. Defining the role of *Caligus rogercresseyi* in transmission and spreading of *Piscirickettsia salmonis*. *Aquaculture* 528, 735489. DOI:10.1016/j.aquaculture.2020.735489

McKinney, W., 2010. Data structures for statistical computing in python. Proceedings 9th Python in Sciences Conferences, 445, 51–56. DOI:10.25080/Majora-92bf1922-00a

Montory, J.A., Cumillaf., J.P., Gebauer, P., Urbina, M., Cubillos V.M., Navarro, J.M., Marin, S.L., and Cruces E., 2020. Early development and metabolic rate of the sea louse Caligus rogercresseyi under different scenarios of temperature and pCO2. *Marine Environmental Research*, 162, 105154. DOI: 10.1016/j.marenvres.2020.105154

Oelckers, K., Vike, S., Duesund, H., Gonzalez, J., Wadsworth, and Nylund, A., 2014. *Caligus rogercresseyi* as a potential vector for transmission of infectious salmon anemia (ISA) virus in Chile. *Aquaculture*, 420/421: 126-132. DOI: 10.1016/j.aquaculture.2013.10.016

Seabold, S. and Perktold, J., 2010. Statsmodels: Econometric and statistical modeling with python. *9th Python in Science Conference*. Pp. 92-96.

Soto, D., Leon-Muñoz, J., Desdner, J., Luengo, L., Tapia, F.J. and Garreaud, R., 2019. Salmon farming vulnerability to climate change in southern Chile: understanding the biophysical, socioeconomic and governance links. *Reviews in Aquaculture*, 11, 354-474. DOI: 10.1111/raq.12336

Valenzuela-Muñoz, V., Gallardo-Escarate, A., Saez-Vera, C., Garces, F., Bonfatti, J. and Gallardo-Escarate, C., 2020. More than bubbles: in vivo assessment and transcriptome modulation of *Caligus rogercresseyi* and Atlantic salmon exposed to hydrogen peroxide (PARAMOVE®). *Aquaculture*, 522, 735170. DOI:10.1016/j.aquaculture.2020.735170

Van Rossum, G. and Drake Jr, F.L. 1995. Python reference manual. Centrum voor Wiskunde en Informatica Amsterdam.

VanderPlas, J., 2017. Python Data Science Handbook. O'Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472. 529P.

Waskom, M.L., 2021. Seaborn: statistical data visualization. *Journal of Open Software*, 6, 3021. DOI: 10.21105/joss.03021