An Open Salmon Knowledge Graph

Scott A. Akenhead[1]

[1]Fisheries and Oceans Canada (retired),

Corresponding author: Scott A. Akenhead, `scott@s4s.com`

Abstract

Goals: A knowledge graph for all aspects of salmon science and management, with a labels schema (faceted classification) by and for humans. Rich nodes richly linked. All content from public sources.

Plain Language Summary

A mind map, but large, to share and discover knowledge in the salmon community data. Who is doing what, where, how, and why? …

1 Introduction

1.1 Goal: data to decisions

Overall, conduits from data through analyses to improved decisions about salmon.

1. **Predictions.** Global warming effects – climate chaos– means the historical patterns and correlations in salmon ecology no longer apply; the required assumption of stationarity, of an ergodic system, has been invalidated.Predictions now require understanding how a salmon will react – via physiology, behaviour, survival - to the conditions it will encounter in every habitat throughout its life.

   A. This implies models that track individual salmon's lives across a time-and-space map of habitats, wherein *observed* environmental variables are translated into *perceived* variables that a salmon will respond to.

   B. The result is a **life path**, a trajectory through multiple dimensions (time, location, fish length, energetics, ontology,) including probability of mortality.

   C. This is an integro-differential equation, where every *rate* (of swimming, of growth,) is a function of the fish *state* and the habitat *state*, when the fish is there.

2. **SDM, SKM.** Salmon Data Mobilization Diack et al. (2023) is required to deliver FAIR[1] datasets Wilkinson et al. (2016) to for the analyses, insights, and datasets underlying models such as above. This initiative is furthered as Salmon Knowledge Mobilization (**SKM**) to discover and effect collaboration opportunities: **Who is doing what, where, how, and why?**

   A. The IT to effect this is a Salmon Knowledge Graph (**SKG**) maintained in neo4j Aura. The size and complexity of *all things salmon-related* is challenging but approachable via:

      i. a **labels schema** that is meaningful to, and created/maintained by, end-users. The SKG must reflect how the salmon community thinks about topics (nodes, entities).

      ii. **personalization** to reduce presented information to that relevant and important to individuals, to avoid overwhelming, irrelevant information. This cannot preclude discovery of previously remote but potentially valuable knowledge.

      iii. myriad **user interfaces**, reflecting myriad activities: planning and management, data processing and analysis, field and lab work, research and modelling, communications and decisions, documents and data products,. Interface components are shared for building new interfaces, and continuously improved.

---

[1] FAIR: findable, accessible, interoperable, reusable.

3. **Decisions.** From the preceding, decision-support products are proffered to effect better decisions via predictions about **competing scenarios**, presented as interactive graphics and dashboards. Overall, conduits from data through analyses to improved decisions about salmon.

4. **Engagement.** Recognizing that drive-by science is obsolete, the engagement required to communicate effectively with decision-makers is critical Chapman (2019), Archibald et al. (2021). How to effect such **inclusivity** is a question in hand.

1.2 How?

Beyond mobilizing data and mobilizing knowledge, it is necessary that the salmon community mobilize: embrace and implement a paradigm (a change in culture, perspective, attitudes,) that
(a) reflects urgency to respond to recent and impending extirpations of salmon populations, and
(b) embraces collaboration to deliver that response. This is largely a communications problem with two aspects:

1. reducing barriers to collabortion

2. clarifying and enabling benefits from sharing.

2 Proposed

"Knowledge graph" means a neo4j labelled properties graph DB.

example: The abundance of returning salmon, R, that are children of a preceding abundance of spawners, S, has conventionally been treated as a simplistic domed function,

$$R_t = S_{t-4}e^{\alpha - \beta S_{t-4}} \tag{1}$$

where α is initial productivity of a salmon population before the effect, β, of salmon density reduces survival from spawners to returns. The use of $t - 4$ reflect presumption of a four-year life cycle, spawners in 2001 will be returns (to spawn the next generation) in 2005. Using Equation 1 for non-linear fitting is preferable to the linearized version,

$$\log(\frac{R_t}{S_{t-4}}) = \alpha - \beta S_{t-4} + N(0, \sigma) \tag{2}$$

which is not ordinary linear regression for many reasons including error term without a known distribution (log ratio Poisson variables), predictor variable is part of predicted variable, and undue effect from cases with imprecise estimates at low abundances of S.
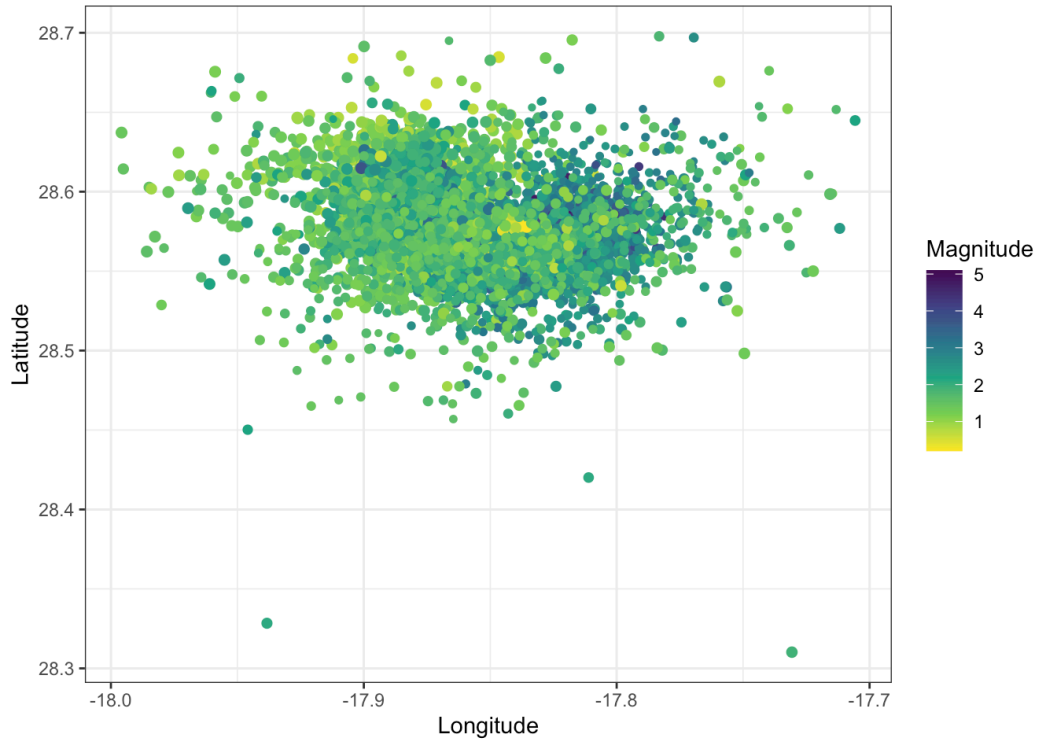
Figure 1: Locations of earthquakes on La Palma since 2017

Figure 1 shows the location of recent Earthquakes on La Palma.

3 Methods

4 Results

5 Conclusion

6 Acknowledgements

7 References

Archibald, D. W., McIver, R., & Rangeley, R. (2021). Untimely publications: De-
layed Canadian fisheries science advice limits transparency of decision-making.
Marine Policy, *132*, 104690. https://doi.org/10.1016/j.marpol.2021.104690

Chapman, K. (2019). First international year of the salmon data laboratory (ISDL)
workshop. In (Vol. 14, pp. 15–20). Vancouver, BC, Canada. Retrieved from
https://npafc.org/wp-content/uploads/technical-reports/Tech-Report
-14/Technical-Report-14_Final.pdf

Diack, G., Bird, T., Akenhead, S., Bayer, J. M., Brophy, D., Bull, C., et al. (2023).
Salmon data mobilization. Retrieved from https://osf.io/hk4gu

Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M.,
Baak, A., et al. (2016). The FAIR guiding principles for scientific data man-
agement and stewardship. *Scientific Data*, *3*(1), 160018. https://doi.org/
10.1038/sdata.2016.18