# A quantitative & reproducible approach for biophysical connectivity models

Steven Hawes1,2 and Will Figueira1,2  
1 Sydney Institute of Marine Science, (Mosman, NSW, (Australia); 2 University of Sydney, School of Life & Environmental Sciences (Sydney, NSW, (Australia)

## Abstract

*LEAVE THIS TO LAST*

## Introduction

In recent times, biophysical models have become a popular tool in advancing our understanding on the connectivity patterns of various marine species. (Andrello *et al.*, 2014; Koeck *et al.*, 2015; Sanvicente-Añorve *et al.*, 2014). Biophysical models, as the name suggests, produce potential patterns of connectivity using a physical model coupled with a biological model. These models are particularly relevant for connectivity in marine systems, as dispersal during a pelagic larval stage is a key driver of connectivity. We couple the two models together, because while oceanic processes are the presiding driver of dispersal, biological processes also have highly influential roles (Cowen and Sponaugle, 2009). The usefulness of these models is their ability to quickly and cheaply give us connectivity patterns over temporal and spatial scales; and also due to the high costs associated with empirical research techniques.

Ecological modellers inherently have underlying concerns about the validity of their models - how realistically have they modelled the system processes? This is a central, if not the primary, concern in all ecological modelling studies. Validation through empirical studies is the gold standard, yet constraints of cost, time and resources generally do not make this possible. Another practical and cheaper approach is to use other published studies, both empirical and modelling, for validation.

With the goal of validating our own biophysical model, we reviewed connectivity studies, published in the last 5 years. Unfortunately, there was inherent difficulty in comparisons, due to the number of competing metrics and lack of transparent data. For a number of these studies, not enough information was provided to be able to reproduce the connectivity patterns presented. Thus prompting two following important questions which need to be address for future studies:

1. How we can start useful comparisons between papers - which metrics should we use?
2. How do we make connectivity studies reproducible?

## Methods

We conducted a database search for connectivity studies, in the last 6 year period (2010-2015) that used a biophysical model in their methods. The start date of 2010 was chosen arbitrarily, to limit the search to a manageable amount of papers; note this was not an exhaustive search. The online search engines, ISI Web of Knowledge and SCOPUS, were queried with the keywords connectivity, model\* and marine. The search returned 59 papers that fit within the scope of our study, consisting of 248 authors and published in 31 unique journals. The full list of references is available at (*provide reference to .csv at github perhaps or in supplementary material?*).

## Connectivity metrics & measures

Two prominent problems we came across were that researchers created their own unique metric that was not embraced in other studies or that only a single parameter of a metric was reported, e.g. mean, median, or upper quartile. These two issues made comparisons between studies almost impossible. In fact, we found over 30 different measures used to describe the patterns of connectivity & dispersal in these papers.

The de-facto standard for publishing connectivity matrices and dispersal kernels is to use a heat map. While heat maps are a powerful visualisation tool for both these metrics, it restricts what information which can be extracted from the graph, which bins the values to a colour legend, resulting in information loss. Comparative quantitative metrics are essential for analysis between studies, needed for model validation.

In a field where there are so many measures, which ones should be consistently reported? The most common parameters found in the dataset related to distance the particles travelled and measures of self-recruitment, local-retention and settlement success (Table 1). Although this isn't as clear cut as it may seem, as there is confusion about the definitions of self-recruitment and local retention. These terms were often used interchangeably. Burgess *et al.*, 2014, stated the definitions clearly and succinctly, i.e. local retention is the fraction of offspring produced by a population that also recruits into that population; and self-recruitment is the fraction of recruitment to a site comprised of individuals born in that site.

Table 1: Those most common metrics published in the connectivity studies and the parameters reported per metric used.

|  |  |  |
| --- | --- | --- |
| Metric | Measure parameters | Reported(%) |
| Connectivity matrix |  | **84.6** |
|  | Self-recruitment | 70.0 |
|  | Local retention | 62.0 |
|  | Settlement success | 42.0 |
|  | Source-sink indices | 32.0 |
| Dispersal kernel |  | **71.2** |
|  | Mean distance | 57.1 |
|  | Maximum distance | 33.3 |
|  | Median distance | 11.9 |
|  | Upper quartile distance | 11.9 |
|  | Minimum distance | 4.8 |
| Graph theory |  | **10.2** |

Our suggested approach is to utilise the two most common metrics used, these provide useful measures and also they should be readily accepted amongst the research community (Table 2). The first metric, a connectivity matrix, is the most important for connectivity studies. The measures of self-recruitment, local retention and settlement success can all be calculated easily from the underlying matrix data. The other important metric is a dispersal kernel. The dispersal kernel gives up valuable data on the spread of the particles and the distance they travel, including the density.

Table 2: The five measures that are required to make comparative analysis from studies using biophysical connectivity models.

|  |  |
| --- | --- |
| Metric | Measure parameters |
| Connectivity matrix | Self-recruitment |
|  | Local retention |
|  | Settlement success |
| Dispersal kernel | Spread |
|  | Distance |

## 

## Reproducibility

### Why reproducible research

Reproducibility, a term often synonymous with open science, is an important facet in assessing the scientific merit of a published paper. It can be difficult to achieve in large empirical studies due to experimental variance (Open Science Collaboration, 2015), the same challenges do not exist in computational modelling (Peng, 2011). Ideally for reproducibility, the model configurations (both physical and biological), all source code and the output data is made available at the time of publishing. Reproducibility allows us to run other published models to compare and contrast again our own.

The cost of reproducible research is minimal, many free online repositories exist for both source code and data, two popular examples being GitHub (http://github.com) and Dryad (http://datadryad.org) . Several journals are also making reproducibility a requirement and providing mechanisms for online storage of data (e.g. PLOS Biology).

### Reproducibility in connectivity studies

We're not going to go into what is required for reproducible research, others have summarised it better than we ever could (Sandve *et al.*, 2013). Instead, the goal is to get researchers using biophysical models for connectivity studies to embrace the concept of reproducible research and to think about how they can meet this goal when publishing their next study.

Complexity in biophysical models derives from coupling two distinct models, biological and physical, creating an additive effect on the number of input parameters. However, it is impossible to reproduce a biophysical connectivity study without all of these parameters, hence the importance of publishing these parameters along with the results. We have supplied a non-exhaustive list of the sort of parameters required to reproduce these studies ({Table 3}). The strong trend towards supplementary information in journals allows for publication of these parameters without obscuring the primary scientific message.

Table 3: List of parameters that should be published with every connectivity study that uses a biophysical model.

|  |  |  |
| --- | --- | --- |
| Model | Name | Description |
| Particle | Particle model | Link to the source code of the particle model used and specifications with information on type, movement and integration methods e.g. Individual based model, with Lagrangian model and it uses Runge-Kutta fourth order method for integration |
|  | Random number seed | If random numbers are used, it is important to publish the seed used to generate the random number so the study can be replicated {Sandve:2013gh} |
| Physical | Oceanographic model | Link to the specifications of the model, including all sub-models, bathymetry & tidal models used |
|  | Spatial scale | The area the particles can move |
|  | Time step | The time step used to move the particles using the oceanographic model (preferably in seconds) |
|  | Diffusion / turbulence | Diffusion strategy (random walk, horizontal and/or vertical) and the turbulence values used |
| Biological | Release times | Release times of the particles |
|  | Release strategy | How the particles were released (e.g. daily, lunar, temperature cues, depths) |
|  | Release sites | The GPS locations and sizes of the release sites |
|  | Settlement sites | The GPS locations and sizes of the settlement sites (if different from the release sites) |
|  | Particles released | The number of particles released |
|  | Pelagic larval duration | The pelagic larval duration of the taxa and whether you used a variable or fixed duration |
|  | Pre-competency period | The pre-competency period given before settlement could occur |
|  | Movement strategies | What movement strategies were the particles given and the values associated with these strategies (e.g. passive, diel, horizontal) |
|  | Mortality strategies | The mortality strategy used (if any) |
|  | Settlement strategies | How did the particles choose a settlement site (e.g. passively, sites were buffered 10 kms, sensory cues) |

# Conclusion

The purpose of this paper is to advance a dialogue about how we can compare biophysical models, used for measuring connectivity in marine systems, with the goal of maturing the field. We identified two main problems in the literature with connectivity studies, that are preventing us from making meaningful comparisons. The first is the inconsistency in the metrics used in connectivity models and how we present the data. The second is the lack of reproducibility and clarity of the model parameters used in studies.

The main take home message is that connectivity modelling studies need to start publishing quantitative metrics of both dispersal and connectivity; too often studies have just focused on one of these facets. We hope that, in addition to providing these measures, researchers will also feel comfortable about publishing the data behind their connectivity matrices and dispersal kernels, allowing for future meta-analyses. Research grants are increasingly asking for publication of datasets and as such, open data publications should start becoming standard practice in the near future.

Our chosen metrics are suggestions only and, realistically, we need a community driven approach—perhaps a discussion facilitated by a workshop—to ensure we come up with agreed common metrics. We hope this paper goes some way to starting this conversation.

## Acknowledgements

## References

\*\*NEED TO FIX UP ERRORS IN END NOTE\*\*

Andrello, M., Mouillot, D., Somot, S., Thuiller, W., and Manel, S. 2014. Additive effects of climate change on connectivity between marine protected areas and larval supply to fished areas. Diversity and Distributions, 21: 139–150. 12 pp.

Burgess, S. C., Nickols, K. J., Griesemer, C. D., Barnett, L. A. K., Dedrick, A. G., Satterthwaite, E. V., Yamane, L., *et al.* 2014. Beyond connectivity: how empirical methods can quantify population persistence to improve marine protected-area design. dx.doi.org, 24: 257–270. Ecological Society of America. 14 pp.

Cowen, R. K., and Sponaugle, S. 2009. Larval Dispersal and Marine Population Connectivity. Annual Review of Marine Science, 1: 443–466. Annual Reviews. 24 pp.

Koeck, B., Gérigny, O., Durieux, E. D. H., Coudray, S., Garsi, L.-H., Bisgambiglia, P.-A., Galgani, F., *et al.* 2015. Connectivity patterns of coastal fishes following different dispersal scenarios across a transboundary marine protected area (Bonifacio strait, NW Mediterranean). Estuarine Coastal and Shelf Science, 154: 234–247. Elsevier Ltd. 14 pp.

Open Science Collaboration. 2015. Estimating the reproducibility of psychological science. Science, 349: aac4716–aac4716. 1 pp.

Peng, R. D. 2011. Reproducible Research in Computational Science. Science, 334: 1226–1227. 2 pp.

Sandve, G. K., Nekrutenko, A., Taylor, J., and Hovig, E. 2013. Ten Simple Rules for Reproducible Computational Research. PLoS Computational Biology, 9: e1003285. 1 p.

Sanvicente-Añorve, L., Zavala-Hidalgo, J., Allende-Arandía, M. E., and Hermoso-Salazar, M. 2014. Connectivity patterns among coral reef systems in the southern Gulf of Mexico. Marine Ecology Progress Series, 498: 27–41. 15 pp.