Using meteorological reanalysis data for multi-decadal hindcasts of larval connectivity in the coastal ocean

Hilde Oliver[1,2,*], Rhiannon L. Rognstad[1,3], David S. Wethey[1]

[1]Department of Biological Sciences, University of South Carolina, Columbia, SC 29208, USA
[2]Department of Marine Sciences, University of Georgia, Athens, GA 30602, USA

[3]*Present address:* Keck Science Department, Claremont McKenna College, Claremont, CA 91711, USA

ABSTRACT: Ocean circulation models are useful for determining population connectivity, but are only available for a limited number of years. In contrast, meteorological reanalyses are available over decades. Since planktonic larvae are typically found in surface waters which are highly influenced by winds, the relationship between connectivity as estimated using an ocean circulation model and wind was used to develop a long-term hindcast of larval dispersal. The University of California Santa Cruz (UCSC) 31 yr Regional Ocean Modeling System (ROMS) hindcast of the California Current System was used to model inter-estuarine transport of larvae with a 6 d larval duration from 1981 to 2010, and between 3 and 8 connectivity patterns were identified using the self-organizing map (SOM) clustering algorithm. Regression models were developed for those connectivity patterns using meteorological reanalyses of winds. Training periods of 5, 10, and 30 yr were used for model development; in all cases there were strong associations between SOM connectivity estimates and winds. Regression models were validated using connectivity estimates from the ocean model. Validated regression models were used with winds from 1950 to 1980 to hindcast connectivity beyond the time range of the original ocean model. Connectivity as estimated from winds was correlated with the Pacific Decadal Oscillation and with upwelling from 1950 to 2010. Multi-decadal hindcasts of population connectivity can be carried out using meteorological reanalysis winds and statistical clustering of connectivity patterns derived from ocean hindcasts of 5 to 10 yr duration.

KEY WORDS: Connectivity · Dispersal · Self-organizing maps · Metapopulations

INTRODUCTION

Dynamics of marine organisms with planktonic larvae are strongly influenced by the degree of connectivity within metapopulations. Changes in connectivity affect the geographic range limits of those organisms (Gaylord & Gaines 2000, Ayata et al. 2010), which can cause broader-scale consequences for marine ecosystems (Siegel et al. 2008). To estimate the maximum distance to which larvae can be dispersed, ocean circulation models are often used to estimate connectivity (Pineda 2000, Connolly & Baird 2010, Lett et al. 2010, Sponaugle et al. 2012). This approach is possible if the ocean models have adequate spatial resolution to resolve flow along the coastline of the region of interest, and if they have adequate temporal resolution. Short time scales (<6 h) are needed for velocity estimates due to the importance of tides and variations in winds, and a high-resolution grid is needed to resolve flow around peninsulas and within bays. However, estimates of connectivity are constrained to the particular dates

---

described by the ocean circulation model, and most long-term ocean hindcasts are of too coarse a spatial resolution and do not have the short temporal resolution needed for informative connectivity hindcasts. For example, the 37 yr Bay of Biscay hindcast (Huret et al. 2013) has a 4 km horizontal resolution, but has a temporal resolution of 5 d, which is too long for adequate estimates of larval trajectories. The 44 yr NW European Shelf Reanalysis (Hines et al. 2013) has a 12 km horizontal resolution, but only monthly temporal resolution, which is far too long for larval transport studies. Other recent multidecadal ocean reanalyses have high temporal resolution but spatial resolutions that are too coarse to describe coastal features important to larval transport (NOAA Climate Forecast System Reanalysis [CFSR]: Saha et al. 2006, European Centre for Medium-Range Forecasts [ECMWF] Ocean Reanalysis System 4 [ORAS4]: Balmaseda et al. 2013). Thus, there is a need for ways to estimate connectivity on long time scales at fine spatial resolutions, despite the scarcity of appropriate oceanographic data, models, or reanalyses. Here we tested the feasibility of using statistical clustering and correlation methods to develop long-term hindcasts of connectivity in marine populations with planktonic larvae. A multi-decadal time series of surface winds was used to test the potential of a significant relationship between meteorological data and larval transport as a means for predicting connectivity when long-term ocean velocity field hindcasts of the necessary resolutions are not available.

Traditionally, velocity fields from ocean models are used to predict dispersal patterns of larvae that are modeled either as Lagrangian passive particles or behaving particles with limited swimming ability (e.g. Ellien et al. 2000, Siegel et al. 2003, Ayata et al. 2009, Mitarai et al. 2009, Huret et al. 2010, Drake et al. 2011, 2013). This approach is constrained by the availability of velocity fields, which are derived from computationally intensive ocean models, so there are often only a few years (frequently less than a decade) of model data available for predicting dispersal patterns. By contrast, gridded meteorological reanalysis data, including surface wind fields, are available in uninterrupted sub-daily time series spanning multiple decades (NOAA CFSR 1979 to the present, Saha et al. 2006; Twentieth Century Reanalysis 1871 to 2011, Compo et al. 2011). Since surface ocean currents are strongly influenced by surface wind stresses and dispersal rates in surface waters are high (e.g. Levin 2006), we expected maximum larval connectivity to be correlated with surface winds. If connectivity estimated from ocean circulation models and that estimated from surface winds were to be strongly correlated, this would provide a possible means of extending connectivity estimates throughout the period of available meteorological reanalysis data sets. Our approach has philosophical similarities to the work of Logemann et al. (2004), who used statistical correlations between air pressure gradients and ocean currents to estimate ocean circulation patterns.

We discuss the following questions:

(1) Is there a relationship between surface winds and maximum larval connectivity in species with short planktonic larval duration, whose larvae remain in surface waters?

(2) If the answer to Question 1 is affirmation, can winds be used to estimate maximum larval connectivity during time periods when ocean model data are not available, e.g. hindcasting of the past?

(3) What are the implications of hindcasts of long-term and large-scale connectivity patterns to our understanding of the dynamics of coastal recruitment in our model system, the California Current System?

METHODS

Assumptions

We carried out this analysis as a proof of concept of a method for carrying out multidecadal analyses of connectivity in marine systems. We recognize that there are many complexities of larval transport in the coastal zone, with possible barriers to transport in the near shore (e.g. Rilov et al. 2008, Morgan et al. 2009, Pineda et al. 2009), all of which can affect connectivity estimates. We also recognize that the spatial scale of the model grid (~10 km) cannot resolve some coastal features, so it is at the upper end of utility for the analysis we propose. However, our analysis is designed to ask (1) if one has a time-series of connectivity estimates that one believes are correct, can one extend that time series further into the past, by using a combination of statistical clustering methods and wind data, and (2) how many years of connectivity data are necessary for the development of a multidecadal hindcast? Here we make the assumption that our connectivity estimates derived from an ocean circulation model are reasonable since they are similar to estimates made by other authors (e.g. Drake et al. 2011, 2013) which were validated with drifting buoys, and use them to test the method of hindcasting. The test of the hindcasting method itself depends only upon the statistical relationships between winds and the connectivity estimates. As higher resolution oce-

an models become available, and as larval behavior of individual species is better understood, the biological accuracy of the hindcasts derived from our method will increase.

Model organism

We chose to estimate population connectivity for benthic species with a short planktonic larval duration, and larvae that remain in surface waters. The intertidal sediment-dwelling polychaete annelid *Abarenicola pacifica* is an example of such a species. *A. pacifica* spawns in shallow coastal waters in March and April (Strathmann 1981), and has a planktonic larval duration of less than 6 d, so winds are likely to be influential in the larval transport process. We use a hypothetical species with larval characteristics similar to *A. pacifica* as a model for other organisms with short planktonic larval duration whose larvae remain in surface waters.

Estimation of larval connectivity

Connectivity matrices describing transport of larvae were generated using output from the University of California Santa Cruz (UCSC) 31 yr Regional Ocean Modeling System (ROMS) hindcast of the California Current System. This model had the longest time series of high temporal resolution ocean velocities that was available anywhere. The hydrodynamics in the ROMS model were forced using meteorological data (short-wave and long-wave radiation, precipitation, atmospheric pressure, wind velocities at 10 m, air temperature) (Veneziani et al. 2009). Drake et al. (2011, 2013) used a similar ROMS model in the same region and obtained good correspondence of model predictions to trajectories of surface drifting buoys.

The following procedures are outlined in Table 1; each step is represented by 1 row in the table, and the input data used, the data processing methods used, and the output data generated are all defined.

Modeling steps 1 & 2: Ocean model file preparation. Zonal (west to east) and meridional (north to south) velocities were used to predict larval trajectories (Table 1, Step 1). The UCSC ROMS model output is on a longitude-latitude-depth grid with a horizontal spacing of 0.1° × 0.1° and velocity grids are available at time intervals of 6 h. As we wanted to obtain an estimate of **maximum** potential transport of particles in surface waters, we extracted near-surface 2-

dimensional (2-D) velocities from the 3-D ROMS model data files. ROMS model velocities at each longitude-latitude and time were interpolated in the vertical direction to a depth of 5 m below the surface to allow simulation of 2-D particle movement near the surface (Table 1, Step 2).

Modeling steps 3 & 4: Particle transport simulations. Lagrangian passive particles representing larvae were transported using numerical integration of these gridded velocities at 5 m depth. Particle trajectories were estimated using Euler's method with a time step of 5 min. Euler's method assumes that water velocities at the location of the particle do not change during the 5 min integration step, so that the distance moved is initial water velocity × elapsed time:

$$X_{t+\Delta t} = X_t + U_{t,x,y} \times \Delta t \text{ and } Y_{t+\Delta t} = Y_t + V_{t,x,y} \times \Delta t \qquad (1)$$

where X_t and Y_t define the particle position at time t, $U_{t,x,y}$ and $V_{t,x,y}$ are water velocities in the E–W and N–S directions at the position of the particle at time t, and Δt is the integration time step (in this case 5 min) (Table 1, Step 4).

To estimate water velocities $U_{t,x,y}$ and $V_{t,x,y}$ for use in the numerical integration described above, ocean velocities in the ROMS model output grids were linearly interpolated in time and bilinearly interpolated in space to the location of each particle at each 5 min interval during the transport simulation (Table 1, Step 3). The time interpolation was necessary because water velocities from the ROMS model were at 6 h intervals yet they were needed at 5 min intervals for the calculations in Eq. (1). The linear interpolation of velocity with respect to time was carried out as:

$$U_{x,y,t} = U_{x,y,T} + (U_{x,y,T+6} - U_{x,y,T}) \times t/6 \text{ and}$$
$$V_{x,y,t} = V_{x,y,T} + (V_{x,y,T+6} - V_{x,y,T}) \times t/6 \qquad (2)$$

where $U_{x,y,T}$ and $U_{x,y,T+6}$ are the E–W velocities at position (x,y) in the ROMS model files at times T and $T+6$ (h), respectively, and $U_{x,y,t}$ is the interpolated E–W velocity at time t (h) after ROMS model time T.

Interpolation of water velocities in space was necessary because the ROMS model velocities were at discrete points on a 0.1° grid (~11.1 km apart in the N–S direction and ~8.5 km apart in the E–W direction at 40° latitude), and we needed estimates of velocities at the locations of individual particles which could be anywhere in the grid. Bilinear interpolation assumes that there are no fronts or other discontinuities, so velocities change linearly among points on the grid. The bilinear interpolation was done with the akima package in R (Gebhardt 2013).

Table 1. Outline of steps used in modeling process, data sources used and their spatial and temporal scales. UCSC: University of California Santa Cruz; ROMS: Regional Ocean Modeling System; SOM: self-organizing map; NPGO: the North Pacific Gyre Oscillation; ENSO: El Niño-Southern Oscillation; CCWI: California Current Winter Index. For descriptions of modeling steps 1 to 14 see 'Methods: Estimation of larval connectivity'

Modeling step	Data type	Data source grid	Model scale	Horizontal scale	Vertical scale	Time	Interpolation/Averaging/ Other calculation
1	ROMS water velocities	UCSC	3-D	0.1° × 0.1° (8.5 × 11.1 km)	42 depths	6 h	None
2	Interpolate ROMS 2-D water velocities	Step 1	2-D	0.1° × 0.1° (8.5 × 11.1 km)	5 m depth	6 h	Interpolation Linear in depth dimension
3	Interpolate particle positions at 5 min intervals	Step 2	2-D	1 × 1 m	5 m depth	5 min	Interpolation Bilinear in space Linear in time
4	Particle transport calculation	Step 3	2-D	1 × 1 m	5 m depth	5 min	Euler's method for numerical integration of water velocities
5	Connectivity matrix calculation	Step 4	2-D	5 km source and destination area	5 m depth	6 d larval transport period	None
6	1800 (6 d) connectivity matrices	Step 5	1-D	Estuary scale (19 estuaries)	None	1 per 6 d larval transport period	None
7	Random sample of connectivity matrices	Step 6	1-D	Estuary scale (19 estuaries)	None	6 d	None
8	3–8 cluster connectivity matrices	SOM or *K*-means clustering	1-D	Estuary scale (19 estuaries)	None	6 d	None
9	Surface wind velocities	20th Century Reanalysis	2-D	1.875° × 1.904°	10 m above water surface	6 d averages of 6 h data	Region average 37°–47° N, 123°–125° W
10	Wind-SOM regression model	Steps 8, 9	1-D	Estuary scale (19 estuaries)	None	6 d	None
11	Wind-SOM regression model validation	Steps 9, 10 connectivity compared to Step 5 connectivity	1-D	Estuary scale (19 estuaries)	None	1 yr	Yearly averages of 6 d connectivity
12	Hindcast	Step 9 winds, Step 10 wind-SOM regression	1-D	Estuary scale (19 estuaries)	None	6 d	Yearly averages of 6 d connectivity
13	Estimation of data requirements for hindcasts	Step 9 winds, Step 10 regressions From 5, 10, 30 yr periods					Yearly averages of 6 d connectivity
14	Relation of connectivity to large scale ocean processes	Step 12 and indices: upwelling NPGO ENSO CCWI	1-D	Region scale (19 estuaries)		6 d	Yearly average of cluster frequency

Lagrangian passive particles (proxies for larvae) were released into the ROMS ocean at high tide from 5 km offshore from 19 different estuaries along the US west coast (see Fig. 1 and Table S1 in the Supplement at www.int-res.com/articles/suppl/m530p047_supp.pdf) and followed for the 6 d planktonic larval duration. A total of 500 uniformly distributed larvae were released in a 5 km diameter patch centered on each release location. We released particles offshore because the ROMS model did not resolve all shallow water features, and to avoid flow artifacts associated with regions immediately adjacent to the land–water boundary. Transport calculations were conducted for each date of probable fecundity (March 11 to May 31) for the years 1981 to 2010, and more than 10 million larval particle trajectories were simulated.

Modeling steps 5 & 6: Calculation of connectivity from larval particle trajectories. Connectivity of larvae released from each site on each date was calculated by determining what fraction passed within 5 km of each of the 19 release sites during the 6 d larval period (Table 1, Step 5). Some larvae never passed within a 2.5 km radius of any of the 19 sites and were therefore lost from the system. Connectivity was estimated as the proportion of larvae originating from each release (source) site that passed by each destination site. The connectivity values were put in a matrix whose rows represented the sites of origin of the larvae and the columns represented the destinations. As there were 19 sites, each connectivity matrix had 19 rows and columns. This process yielded 1800 individual ROMS connectivity matrices, 1 for each larval release date (Table 1, Step 6). Since the model species has 1 reproductive season per year, a yearly connectivity matrix was calculated from the mean of each element in the individual matrices for each year.

Modeling steps 7 & 8: Clustering methods. The connectivity analysis (Table 1, Step 6) generated 1800 matrices, 1 for each of the release dates. Cluster analysis was used to find groups of similar patterns among individual ROMS connectivity matrices (Table 1, Steps 7 and 8). The clustering methods condense the primary differences among 1800 individual ROMS connectivity matrices into a small number of representative matrices or clusters. Clustering methods like these have been used very successfully in oceanography (e.g. Cassou & Terray 2001, Liu & Weisberg 2005, Barrier et al. 2013) and climatology (e.g. Johnson et al. 2008, Lee & Feldstein 2013) to detect large scale patterns and their drivers.

We compared *K*-means (Hartigan & Wong 1979) and self-organizing map (SOM) (Kohonen 2001) clustering methods, constraining the maximum number of clusters between 3 and 8 (Table 1, Step 8). In the SOM clustering method, a competitive learning process is used, in which a random (training) subsample of the connectivity matrices is used to obtain a specified number of clusters. As each connectivity matrix is entered into the SOM algorithm during the training process, its Euclidean distance to all existing clusters is calculated, and then the cluster that minimizes the distance is identified as the best matching unit of that matrix. The best matching unit cluster is then updated to be more similar to the newly added matrix. The *K*-means clustering algorithm matches each matrix to the cluster that minimizes the within-cluster sum of squares before updating the cluster to which the matrix is matched.

The proportion of data used for training the algorithms was varied, incorporating a random sampling of 1/3, 1/2, or 2/3 of the original data matrices. Once the algorithms assigned the connectivity matrix for a particular larval release date to a cluster, the cluster ID number (between 1 and 8) was used to record the assignment. This process was repeated with 1000 different random training data sets for each combination of number of clusters, fraction of data used for training, and clustering algorithm. Due to the unusually long duration of the ROMS model hindcast, the same process was repeated with 3 sets of 5 yr of connectivity matrices (1991–1995, 1996–2000, 2006–2010), 10 yr of connectivity matrices (2001–2010) and 30 yr of connectivity matrices (1981–2010).

Modeling steps 9 & 10: Relationship between connectivity and wind. Since the passive larval particles were transported in surface currents, which are expected to be strongly influenced by surface winds, the relationship between connectivity and the average surface winds during the period of larval transport was examined (Table 1, Steps 9 and 10). The 20th Century Reanalysis Version 2 (Compo et al. 2011) was used as the source of the time series of gridded wind data because it has the longest time series available among all reanalysis data sets (1871 to 2011). This product has a spatial scale of 1.875° × 1.904° and a temporal scale of 6 h. For the period of transport from each larval release, we made spatio-temporal averages of the zonal (*u*) and meridional (*v*) wind components, and wind direction at 10 m elevation, over the coastal ocean in the geographic region of our study (37° to 47° N, 123° to 125° W) (Table 1, Step 9). Wind direction was calculated as the arctangent of the average *u* and *v* components.

Linear regressions were calculated to predict clusters from wind magnitude and direction (Table 1, Step 10). Separate regressions were calculated using zonal wind, meridional wind, wind direction, and combinations of the three. These analyses were carried out for each of the combinations of clustering method (SOM and *K*-means), number of clusters (3 to 8), proportion of original data used in training (1/3, 1/2, 2/3), and duration of training data (5, 10, 30 yr) using matrices from the 31 yr ROMS dataset. In each case, the regression models with the largest R^2 values were chosen for further analysis. Aikake's Information Criteria (AIC) were calculated for the regressions as an alternative method of model selection. Since all but one of the best regression models (see Tables 1 & 2) used N–S wind or N–S wind and wind direction as the independent variable(s), we did the remainder of our analyses using N–S wind and wind direction.

Modeling step 11: Validation of the wind model for predicting larval dispersal. Yearly connectivity matrices were created by calculating the mean values of each element of the matrices for each year, using both the matrices calculated directly from the ROMS model and using the cluster matrices predicted solely from winds (Table 1, Step 11). In order to estimate the population consequences of the ROMS connectivity matrices versus the yearly connectivity patterns predicted from winds, 1000 larvae were released from each of the 19 sites and the geographic distribution of recruiting larvae was calculated using the connectivity matrices. The site-specific recruitment densities predicted from the ROMS and wind-based connectivity matrices were regressed against one another. This analysis was repeated using wind-based regressions derived from 3 to 8 SOM clusters, to determine how model performance changed as a function of the number of clusters used. The regression R^2 values were used to evaluate model performance.

Modeling step 12: Connectivity hindcast using winds. The regression model predicting connectivity from wind is likely to be valid if the conditions in the hindcast period are similar to those used in estimation of model parameters. Therefore, it is important to determine whether there were extreme events in the hindcast period (1950 to 1980) that lay outside the range of conditions that occurred during the 5,10, and 30 yr model estimation periods. Ranges of wind speeds and directions were compared between the model estimation period and the hindcast period.

To carry out the hindcast, we calculated 6 d averages of the regional zonal and meridional winds from the Twentieth Century Reanalysis following all potential larval release dates during the period 1950 to 1980. These average winds were used with the validated regression model from the 5, 10 and 30 yr training datasets with the largest R^2 values to predict cluster IDs for the larval transport period during the years 1950 to 1980. The cluster matrices predicted for each release date were averaged to generate yearly connectivity matrices for the duration of the hindcast period (1950–1980) (Table 1, Step 12).

Modeling step 13: Estimation of number of years needed for connectivity model development. Many ocean model hindcasts span periods shorter than 3 decades and therefore have less data available for generating connectivity estimates. It was therefore necessary to test the sensitivity of our methods to differences in the total amount of data available (Table 1, Step 13). The majority of models available have between 5 and 10 yr of data, which usually occur in the past 5 to 10 yr, so we used data from the

past decade to test our methods. To test how differences in the duration of the ocean model time series would affect the quality of the hindcast, we performed our clustering and wind analyses on sets of connectivity matrices spanning 5 yr (2006–2010), 10 yr (2001–2010) and 30 yr (1981–2010) (see Fig. 1). We also examined whether there was an effect on model performance depending on the time period used for building the model. We did this by using 5 yr of data from 1991–1995, 1996–2000, and 2006–2010.

Modeling step 14: Relation to large-scale atmospheric and oceanographic processes. To determine whether connectivity patterns were associated with large scale atmospheric or oceanographic processes, we calculated regressions of connectivity patterns versus the Multivariate El Niño-Southern Oscillation (ENSO) Index (MEI, Wolter & Timlin 2011), the Pacific Decadal Oscillation (PDO, Mantua & Hare 2002), the North Pacific Gyre Oscillation (NPGO, Di Lorenzo et al. 2008), and the regional upwelling index (Bakun 1973) (Table 1, Step 14). MEI data were obtained from the National Oceanic and Atmospheric Administration (NOAA). PDO data were obtained from the Joint Institute for the Study of the Atmosphere and Ocean at the University of Washington. NPGO data were obtained from E. Di Lorenzo (Georgia Institute of Technology). Coastal upwelling index data were obtained from the NOAA Pacific Fisheries Environmental Laboratory. Regional average upwelling indices for the planktonic larval period were calculated from the monthly time series for the US West coast at 39°N, 42°N, and 45°N, the locations encompassing the study region (see Fig. 1).

Each year from 1950 to 2010, we calculated the frequency of occurrence of each of the connectivity cluster matrices that were predicted by the winds, using regression equations developed from the 5, 10 and 30 yr training periods. We regressed the cluster frequencies against the seasonal values of the MEI, PDO, NPGO and upwelling indices corresponding to the larval transport period in each year.

Statistics

All calculations were carried out in the statistical programming language R (Ihaka & Gentleman 1996, R Core Team 2012) Version 2.15, using the akima library (Gebhardt 2013) for interpolation, the clim. pact library (Benestad 2004) for conversion between latitude/longitude coordinates and meters, the kohonen library (Wehrens & Buydens 2007) for SOM calculations, and the stats library for *K*-means clustering calculations.

RESULTS

Clustering method selection

Surface winds explained up to 57, 48, and 38% of the variance in the respective 5, 10, and 30 yr connectivity patterns, based on linear regressions of cluster identity versus wind magnitude and direction. Comparisons of the best regression R^2 values from different clustering methods, number of clusters, proportion of data used for training, and wind data (Table 2) showed that the clusters generated by K-means and SOM algorithms correlate similarly to the Twentieth Century Reanalysis wind data. The SOM algorithm did a slightly better job of clustering the connectivity data than the K-means algorithm, based on the relationships between connectivity patterns and surface winds. In general, more clusters and larger proportions of the data used for training resulted in smaller correlation coefficients. The best R^2 values that were used for further analysis were found by correlating the clusters generated by the SOM clustering method (7 clusters for 5 yr [2006–2010] and 10 yr training data [2001–2010], and 4 and 7 clusters for 30 yr training data [1981–2010] with meridional wind magnitude and wind direction, using 1/3 of the data for training). Smaller numbers of clusters had higher correlation with meridional wind magnitude and wind direction, while larger numbers of clusters generally had higher correlation with only meridional wind magnitude (Table 2). AIC analyses

(Table 3) indicated that the smallest number of clusters should be used, but we felt that fewer clusters captured less of the variation in connectivity, and possibly underestimated maximum dispersal distance, so we chose 5 clusters as a compromise between best AIC and best R^2. As the vast majority of the results in Table 2 were based on meridional wind magnitude and wind direction, we decided to use these 2 variables as predictors in the remainder of our analyses.

Hindcast model validation

A measure of the quality of the wind-derived cluster connectivity estimates is the consequence of connectivity for population dynamics. We used yearly averages of the ROMS matrices and the wind-derived cluster connectivity matrices to predict the geographic distribution of arriving larvae, if 1000 larvae were released from each of the 19 estuaries (Fig. 1), and then regressed the distribution using the winds against the predicted distribution of larvae from the original connectivity matrices found using the ROMS model. The slope of the relationship ranged from 0.86 to 1.45 with a mean of 1.15, and the R^2 value ranged from 0.64 to 0.98 with a mean of 0.85. The quality of the wind-derived connectivity estimates did not change with the number of clusters used, within the range of 3 to 8 clusters (Fig. 1). For the remainder of our analysis, we used 5 clusters as a

Table 2. Highest R^2 values resulting from regressions correlating cluster ID to reanalysis wind data. **Bold**, non-italicized, and non-shaded data resulted from using meridional wind magnitude and wind direction as predictors; all values in normal typeface resulted from using just north–south wind magnitude as a predictor; the value in italics resulted from using total wind magnitude and wind directions as predictors; the value in the shaded cell resulted from using wind direction and zonal wind magnitude as predictors; and the underlined values are the highest maximum R^2 values resulting from each regression analysis for each model duration. SOM: self-organizing map

Cluste-ring method	No. of clusters	2006–2010 model			2001–2010 model			1981–2010 model		
		1/3 training	1/2 training	2/3 training	1/3 training	1/2 training	2/3 training	1/3 training	1/2 training	2/3 training
SOM	3	**0.51**	**0.47**	**0.40**	0.41	**0.41**	0.37	**0.33**	**0.34**	**0.33**
	4	*0.54*	**0.47**	**0.42**	**0.45**	**0.43**	**0.41**	<u>0.38</u>	**0.32**	**0.32**
	5	**0.52**	0.44	**0.41**	**0.45**	**0.40**	**0.38**	**0.35**	**0.33**	0.30
	6	**0.53**	0.44	**0.39**	0.45	**0.42**	**0.43**	0.35	0.30	0.30
	7	0.57	0.44	0.43	<u>0.48</u>	0.39	**0.37**	**0.33**	0.30	**0.28**
	8	<u>0.57</u>	0.49	0.41	**0.45**	**0.40**	0.37	0.33	0.30	0.31
K-means	3	0.47	**0.46**	0.41	**0.43**	**0.40**	0.37	**0.37**	**0.33**	**0.33**
	4	**0.51**	**0.47**	**0.45**	**0.44**	**0.41**	**0.40**	**0.37**	**0.32**	**0.31**
	5	**0.52**	0.47	0.43	**0.44**	**0.40**	**0.38**	**0.35**	0.31	0.31
	6	**0.51**	**0.45**	0.43	**0.45**	**0.40**	**0.39**	0.35	0.31	0.30
	7	0.54	**0.46**	**0.41**	**0.46**	**0.42**	**0.38**	0.34	0.32	0.30
	8	0.51	0.46	0.44	0.43	**0.40**	0.39	0.34	0.30	0.30

Table 3. Aikake's information criterion (AIC) values resulting from regressions with highest R^2 values correlating cluster ID to reanalysis wind data as used in Table 2. **Bold**, non-italicized, and non-shaded data resulted from using meridional wind magnitude and wind direction as predictors; all values in normal typeface resulted from using just north–south wind magnitude as a predictor; the value in italics resulted from using total wind magnitude and wind directions as predictors; the value in the shaded cell resulted from using wind direction and zonal wind magnitude as predictors; and the underlined values are the AIC values from regressions with highest maximum R^2 values from analyses for each model duration. SOM: self-organizing map

| Cluste-ring method	No. of clusters	2006–2010 model			2001–2010 model			1981–2010 model		
		1/3 training	1/2 training	2/3 training	1/3 training	1/2 training	2/3 training	1/3 training	1/2 training	2/3 training
SOM	3	**180**	**260**	**374**	373	**564**	**771**	**1213**	**1825**	**2529**
	4	*244*	**360**	**489**	**519**	**775**	**1065**	<u>**1640**</u>	**2478**	**3377**
	5	**297**	459	**632**	632	**991**	**1305**	**2006**	**2978**	4036
	6	**326**	534	**711**	738	**1087**	**1457**	2265	3466	4491
	7	<u>350</u>	540	775	<u>776</u>	1221	**1623**	2485	3780	**5061**
	8	<u>385</u>	607	839	852	1280	1699	2689	4011	5517
K-means	3	185	251	363	357	575	776	1160	1826	2430
	4	**241**	362	**491**	495	775	1053	1655	2579	3403
	5	**313**	473	638	625	959	1315	**2026**	3085	4055
	6	**337**	523	704	696	1096	1446	2266	3405	4623
	7	341	**584**	**791**	718	1178	1596	2466	3701	5025
	8	407	619	803	843	**1266**	1716	2623	4017	5328
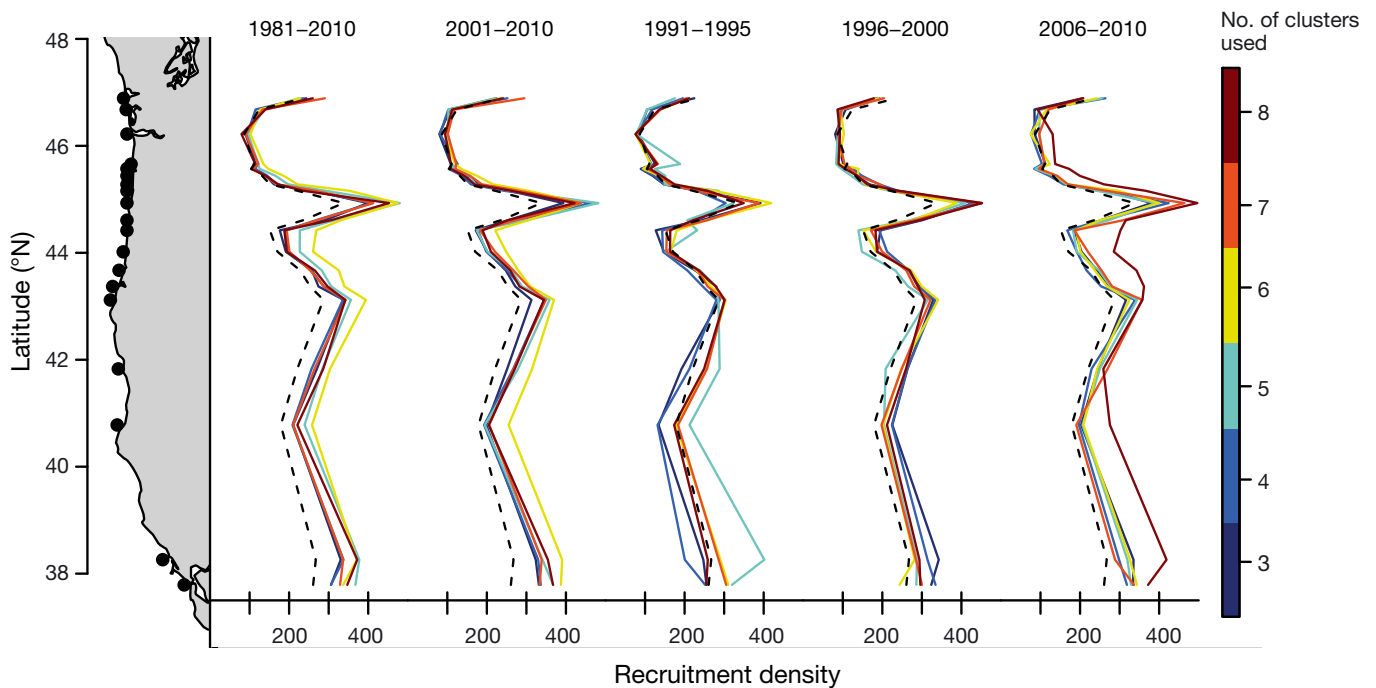
Fig. 1. Site-specific recruitment densities after dispersal modeled by the Regional Ocean Modeling System (ROMS) or winds using yearly connectivity matrices from 1981 to 2010 and comparison of the utility of different numbers of clusters for recruitment density predictions. The yearly averages of ROMS model matrices and cluster predictions were used to predict dispersal, beginning with 1000 passive particles released at each site. The sites of release are marked with black dots on the map of the Pacific coast of North America (California, Oregon, Washington); 1 dot per estuary. The colored curves show the average recruitment using self-organizing map (SOM) cluster models based on 30 yr of training data from 1981–2010 (min $R^2 = 0.89$, max $R^2 = 0.99$), 10 yr of training data from 2001–2010 (min $R^2 = 0.91$, max $R^2 = 0.99$), and 5 yr of training data from 1991–1995 (min $R^2 = 0.69$, max $R^2 = 0.95$), 1996–2000 (min $R^2 = 0.90$, max $R^2 = 0.95$), and 2006–2010 (min $R^2 = 0.77$, max $R^2 = 0.98$), with each color indicating the number of clusters used. Dashed line = prediction of recruitment density from average of original Lagrangian particle tracking data from the ocean model for 1981–2010, colored lines = predictions from SOM connectivity derived solely from winds for each number of clusters

reasonable compromise between model complexity and ability of the model to capture the underlying variability of the system.

Estimation of number of years needed for training

Many ocean models have shorter durations than 3 decades, so it was necessary to determine whether regression models correlating winds to clusters of connectivity from shorter time periods would be appropriate for making multidecadal hindcasts of connectivity. The sensitivity of the regressions to differences in the total amount of data provided for sampling for training (a shorter training period would have less connectivity information available) was first tested by looking at the maximum R^2 values using connectivity matrices from 5 yr periods (1991–1995, 1996–2000, 2006–2010) and a 10 yr period (2001–2010) (Table 2, see Table S2 in the Supplement). Winds explained up to 57% of the variance in connectivity for the 2006–2010 model and up to 48% of the variance in connectivity for the 10 yr model, while they explained 38% of the variance for the 30 yr model. The best regression model for the 10 yr training period used SOM clustering, 1/3 training, 7 clusters, and meridional (N–S) wind magnitude and wind direction as predictors, while one of the best 2 regression models for the 5 yr period used SOM clustering, 1/3 training, and 8 clusters but was anomalous in that it used zonal (W–E) wind magnitude as a predictor for connectivity along a north–south-running coastline ($R^2 = 0.57$). However, while higher cluster numbers tended to account for more of the variance in wind velocity data, the AIC values tended to decrease for decreased numbers of clusters (Table 3, see Table S3 in the Supplement), with AIC values at least doubling by increasing the number of clusters from 3 to 8. Using 1/3 of the data for training and using meridional wind magnitude and wind direction resulted in consistently higher R^2 and lower AIC values whether SOM or *K*-means was used (Tables 2 & 3), and the 2 different clustering methods resulted in no major differences between R^2 and AIC values. Therefore validations of the model were proceeded with using meridional wind magnitude and wind direction, 1/3 sampling used for training, SOM clustering (though *K*-means clustering may also have been used), and all 6 numbers of clusters (3 to 8).

The geographic distribution of larvae arriving at the 19 west-coast estuaries was predicted for the 1981–2010, 2001–2010, 1991–1995, 1996–2000, and 2006–2010 models (Fig. 1) for each number of clusters (3 to 8), 1/3 sampling, and SOM clustering. Regressions were calculated correlating the distribution of larvae predicted using surface winds (meridional magnitude and wind direction) to the distribution of larvae predicted from the original connectivity matrices. Recruitment patterns predicted by the SOM model were highly correlated with those predicted directly from the ocean model. For the SOM model trained using data from the 5 yr period covering 2006–2010, the R^2 values ranged from 0.77 to 0.98 with a mean of 0.92, and all p-values were less than 0.01. For the SOM model trained with 5 yr of data from 1996–2000, the R^2 values ranged from 0.90 to 0.95 with a mean of 0.92, and all p-values were less than 0.01. For the SOM model trained with 5 yr of data from 1991–1995, the R^2 values ranged from 0.69 to 0.95 with a mean of 0.90, and all p-values were less than 0.01. For the 10 yr SOM model (2001–2010), the R^2 values ranged from 0.91 to 0.99 with a mean of 0.95, and all p-values were less than 0.01. The 30 yr SOM model (1981–2010) had R^2 values ranging from 0.89 to 0.99 with a mean of 0.96, and all p-values were less than 0.01.

Validity of winds as predictors of connectivity in hindcast period

To determine whether ranges of wind conditions were similar during both the hindcast period and the dates used to develop the wind-based connectivity model, the ranges of wind speeds and directions in the hindcast period (1950–1980) were compared to those in the 3 training periods. The ranges of wind values in the 5 yr (2006–2010) and 10 yr (2001–2010) training data sets were smaller than the ranges during the hindcast period, so conditions similar to the most extreme events of 1950–1980 did not occur between 2001 and 2010 (Table 4). In this situation, the

Table 4. Ranges of wind reanalysis datasets for the hindcast period (1950–1980) and the period used for training (1981–2010)

	Hindcast period 1950–1980	Training periods		
		2006–2010	2001–2010	1981–2010
Zonal velocity	−1.33 to 3.19	−0.79 to 3.26	−0.80 to 3.26	−1.69 to 3.26
Meridional velocity	−2.66 to 3.17	−2.39 to 3.05	−2.39 to 3.05	−2.64 to 3.28
Direction	−3.04 to 3.12	−2.84 to 2.39	−2.84 to 2.71	−3.08 to −3.12

regression equations are, in a few cases, applied outside the range of the independent variables used to derive the equations. By contrast, the range of wind values in the 30 yr (1981–2010) training data set was similar to the range of values in the 60 yr hindcast period (1951–2010) (Table 4), so the regression equations are being used within the range of the independent variables used to derive the equation in this case.

Relation between connectivity and large-scale oceanographic processes

To test whether long-term connectivity patterns were associated with large scale oceanographic and meteorological oscillations for the hindcast period from 1950 to 2010, the annual predicted frequencies of occurrence of each cluster estimated from the training periods 1981–2010, 2001–2010, and 2006–2010 (Fig. 2) were compared to the ENSO, the PDO, the regional upwelling index, and the NPGO. While using 5 clusters from the 30 yr training period, none of the frequencies correlated with ENSO, but there were correlations between PDO and the first ($p < 0.01$) and third clusters ($p < 0.05$), the regional upwelling index and the first ($p < 0.01$), second ($p < 0.01$), and fourth clusters ($p < 0.01$), and NPGO and the third ($p < 0.05$) cluster. In the 10 yr model with 5 clusters, PDO correlated with the first ($p < 0.05$) and third ($p < 0.01$) clusters, the regional upwelling index correlated with the first ($p < 0.01$), second ($p < 0.01$), and fourth ($p < 0.01$) clusters, and neither ENSO nor NPGO correlated with the cluster proportions. In the 5 yr (2006–2010) model with 5 clusters, PDO correlated with the first ($p < 0.05$) and third ($p < 0.05$) clusters, the regional upwelling index correlated with the first ($p < 0.01$), second ($p < 0.05$), and fourth ($p < 0.01$) clusters, while ENSO and NPGO did not correlate with the clusters.

DISCUSSION

In order to understand long-term dynamics of marine species, it is important to estimate changes in connectivity among populations. Lagrangian particle-tracking methods have been used very successfully to estimate connectivity using velocity fields from ocean models (e.g. Siegel et al. 2003, 2008, Ayata et al. 2010, Drake et al. 2011, 2013, Sponaugle et al. 2012). However, ocean circulation model data are only available for a limited number of years, so

alternative methods become necessary for hindcasts extending beyond the dates described by the ocean models. In contrast, sub-daily meteorological re-analysis data are available for dates extending back to the 19th century (Compo et al. 2011). For species whose larvae are typically found in surface waters, there should be a strong influence of winds on connectivity. Therefore the relationship between winds and larval connectivity patterns predicted by the ocean model was investigated as a potential method for hindcasting connectivity patterns for years presently described by wind data, but not by ocean models. By corresponding connectivity with wind reanalysis data using statistical clustering methods, the proposed method doubled the temporal range of connectivity estimates by accounting for much of the variance of coastal connectivity over 30 yr with only a few connectivity patterns.

Hindcast model validation

Using clustering methods, it was possible to compress large numbers of individual 6 d connectivity patterns in the California Current System from 5, 10 and 30 yr periods (2006–2010, 2001–2010, and 1981–2010) into 7 or fewer major clusters (Fig. 2). These connectivity patterns were associated with winds: 57 % of the variance in cluster identity can be explained by winds in the 5 yr analysis (2006–2010), and 38 % of the variance in the 30 yr analysis (Table 2). Similar correlations have been shown to be characteristic of other SOM analyses that led to robust conclusions about shifting wind patterns (Johnson et al. 2008, Lee & Feldstein 2013).

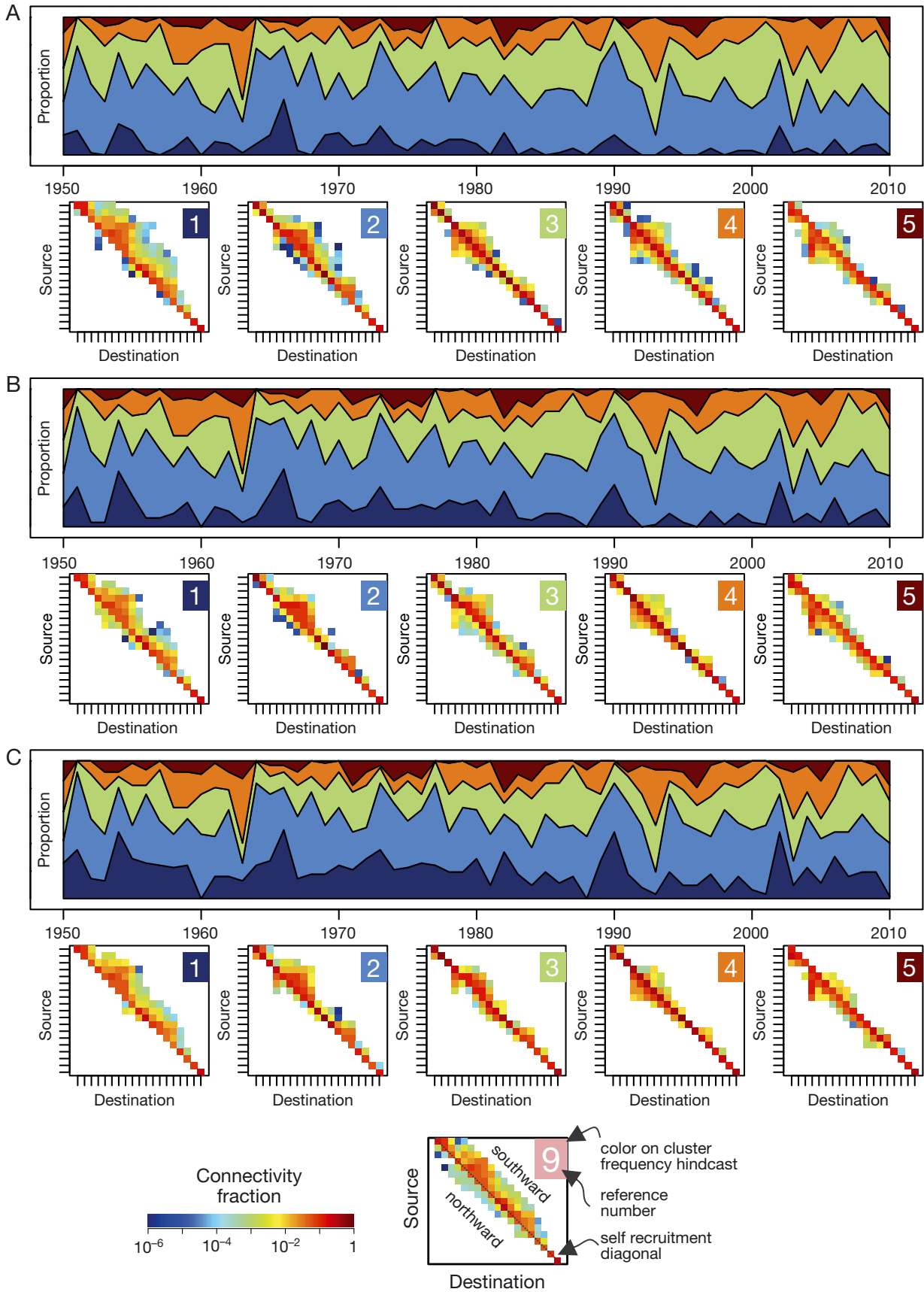The population consequences of dispersal are well described using connectivity estimated with

Fig. 2. Yearly frequencies of each of 5 connectivity clusters in the 1950–2010 hindcast. Each frequency graph is a plot of proportion of each cluster for each year hindcasted using the regression model. The plots are as follows: (A) 5 clusters using a 5 yr training period, (B) 5 clusters using a 10 yr training period, and (C) 5 clusters using a 30 yr training period. Plots of the respective clusters (connectivity matrices plotting the proportion of larvae beginning from each of the 19 start sites that arrive at the 19 destination sites) are below their frequency plots, with the color in the top right-hand corner of each indicating that particular plot's frequency in the graph above it. The scale is given at the bottom, along with an interpretive schematic of the connectivity matrices. Values to the right of the diagonal indicate southward transport, values to the left of the diagonal represent northward transport

Connectivity fraction

10^{-6} 10^{-4} 10^{-2} 1

color on cluster frequency hindcast

reference number

self recruitment diagonal

southward

northward

the regression model using winds and connectivity clusters for all tested total numbers of clusters (3 to 8). Yearly connectivity matrices were used to disperse virtual larvae among 19 estuaries on the US west coast, and strong association was found between virtual recruitment densities estimated using 3 to 8 SOM matrices and recruitment densities estimated using the matrices from the Lagrangian particle tracking over the period 1981 to 2010, with every p-value < 0.01 (Fig. 1). Such statistically significant validations of the regression model suggest that wind data from meteorological reanalysis can be used to temporally extend a time-limited ocean circulation model for predicting larval dispersal.

Our analysis was carried out for an organism with a 6 d planktonic larval duration. We suspect that the accuracy of the method will decline to some extent as planktonic larval duration increases because variation in wind speed and direction during the larval period will not always be well represented by longer averages.

One could argue that since winds are used to force the ocean model, connectivity patterns derived from clustering methods are by definition correlated with winds. In the ROMS model, winds contribute to water movements, just as they contribute to real-world water flows. However, in both the ocean and the ROMS model, there are also influences of bottom topography, bottom friction, viscosity, water density, and tides, all of which contribute to flows, even in the absence of wind. Therefore there should be a correlation between wind and connectivity, but it will be modified by the physics of ocean water and its interactions with the atmosphere, gravitational forces, and topography.

Number of years needed for training

Frequently, ocean models are only available for periods shorter than 30 yr, so the same regression of winds and connectivity analyses were run using the ROMS model output for a 5 yr period and a 10 yr period (Table 2). The maximum R^2 value for the 2006–2010 period was 0.57, which was actually an improvement over using the connectivity matrices from all 30 yr ($R^2 = 0.38$), suggesting that shorter periods reduce the chance of overspecification of the regression models. Virtual recruitment density estimates for that particular 5 yr period (mean $R^2 = 0.88$) were also slightly improved over the 30 yr period (mean $R^2 = 0.85$) (Fig. 1). The similar success of using

shorter periods suggests that multi-decadal connectivity hindcasts are feasible with 5 to 10 yr of ocean model data.

To determine whether a connectivity hindcast using winds would be reliable, the ranges of values for meridional wind magnitude, zonal wind magnitude, and wind direction on the dates described by the original data and the dates intended to be hindcasted were compared. It was expected that a hindcast would be most likely to underestimate connectivity if the range of wind data in the hindcast period fell outside the ranges of the winds for the dates of the original data. While the range of wind data in the 1981–2010 training period was similar to the 1950–1980 hindcast period, the shorter training periods had smaller ranges of data than the 1981–2010 training period and the 1950–1980 hindcast periods because they capture fewer extreme events (Table 4). However, the clusters found using shorter training periods (2001–2010, 1991–1995, 1996–2000, 2006–2010) saw similar success (all p-values < 0.01) during verification analysis (Fig. 1), suggesting that capturing more extreme wind patterns in the training period does not make a significant difference to connectivity, especially when looking at yearly averages of connectivity. The validated regression models were used with winds from 1950–2010 to hindcast yearly connectivity beyond the time range of the original ocean model, effectively producing twice, 6 times, or 12 times the number of years for which connectivity data is available. Based on this analysis, 5 yr of data are likely to be adequate for building a wind-based regression model of connectivity like that which we have proposed here.

Relation between connectivity and large-scale oceanographic processes

As connectivity was well correlated with winds, and long-term measures of climatic and environmental variability are linked to patterns of recruitment (Menge & Menge 2013), we expected to find associations between connectivity and large-scale oceanographic processes. PDO and especially the regional upwelling index had very strong correlations with the cluster frequencies. The significant correlation between the frequencies and the regional upwelling index makes sense in that the connectivity clusters with higher self-recruitment occur most often in years when the regional upwelling index is low. Larvae in the surface layer are largely transported offshore and are lost from the system, particularly

during upwelling (Drake et al. 2011), and major differences in larval recruitment have been attributed to shifts in upwelling (Menge & Menge 2013).

Implications of the results

Our results provide an opportunity to examine the importance of long-term, large-scale hindcasts for understanding population dynamics in coastal ecosystems, in this case the California Current System. Spatial and temporal variation in recruitment has been documented in the California Current System since the beginning of the 20th century, and early authors attributed interannual fluctuations to variations in ocean currents (e.g. Coe 1953, 1955). However, at the time of Coe's work, there was no way to test his ideas directly, since there had been few large-scale measurements of ocean currents during the years in which he observed massive recruitment events (1931, 1934, 1938, 1950), since CalCOFI surveys did not start until 1950 (CalCOFI 1950). Though limited drift bottle data and current meter data from the first half of the 20th century does exist, high-resolution patterns of currents were not well characterized at the time (Marmer 1926). While the southern region of the California Current System has been sampled intensively since 1950, Oregon has less data (Wyatt et al. 1972), making the methods described in the present paper a means of providing estimates of patterns for a time when relatively few surface current observations exist for the Northern California and Oregon coast. More recently, large-scale differences in recruitment have been linked to regional differences in upwelling regimes (e.g. Connolly et al. 2001, Broitman et al. 2008, Menge et al. 2011, Menge & Menge 2013). Our connectivity hindcast patterns are consistent with the observations of these authors. The boundary between recruitment regimes that they observed is in the same area where we see a reduction in north–south connectivity. By extending the temporal range of ocean models decades into the past, our methods also offer the opportunity to examine the environmental conditions surrounding historical variations in recruitment on large spatial scales, in conjunction with data on long-term variability in the California Current System (Bakun & Nelson 1991, Bograd & Lynn 2003). In addition, it raises the possibility of extending population connectivity estimates even farther into the past, using long-term reconstructions of general properties of the California Current System (Black et al. 2014).

Degree of self-recruitment versus northward or southward transport across latitudes

The connectivity patterns indicate that there is a great deal of self-recruitment, which should uncouple the population dynamics of sites separated by distances greater than ~20 km. If these patterns are maintained in species with longer planktonic larval duration, they would help explain the spatial patterns of synchrony in space occupiers like barnacles and mussels (e.g. Broitman et al. 2008, Gouhier et al. 2010, Menge et al. 2011). For example, Gouhier et al. (2010) found that the spatial correlations of mussel population dynamics on the US Pacific coast decayed to zero at a distance of 150 km, which is approximately 3× the spatial scale of mussel larval dispersal (Gilg & Hilbish 2003). Broitman et al. (2008) observed high similarity in recruitment timing and intensity among sites from Strawberry Hill (44.25°N) to the north. This region showed high intra region connectivity in our analyses (Fig. 2), which may contribute to the patterns observed by Broitman et al. (2008).

It is possible that the greater number of estuaries north of 43°N compared to south of 43°N influences our ability to estimate connectivity in those 2 regions. However, Drake et al. (2011, 2013) observed similar patterns in their high resolution simulations of connectivity in the same geographic region. They estimated connectivity among locations spaced by 2 to 4 km along the entire coastline and found higher connectivity among northern sites than between the northern and southern regions of our study area. Therefore, we believe that our connectivity estimates are representative of the actual patterns in the region.

Differences in correlations between long-term cycles and recruitment

An association between long-term cycles and recruitment patterns on the US Pacific coast has been documented on large geographic scales by Menge et al. (2011). These authors found associations at some temporal scales and not others, and some relationships were relatively constant over 15 yr whereas others varied in intensity over that period. For example, the seasonal scale association between the NPGO and MEI was strong between 1995 and 2001 and weak otherwise. We found a major shift in connectivity patterns bracketing these dates, with a large increase in the frequency of connectivity cluster 5 before and after those years (Fig. 2).

We suspect that offshore transport of larvae during upwelling events may contribute to changes in the degree of connectivity among sites. This leads to the correlations between some connectivity patterns and the regional upwelling index. Menge et al. (2011) found the strongest associations with upwelling at seasonal time scales, but they also observed associations at multi-annual scales. Since our connectivity analysis was restricted to a 3 mo period of each year, we cannot resolve the seasonal patterns, but we do see the multi-annual association with the upwelling index.

The relationship between connectivity and the PDO (see 'Results: Relation between connectivity and large-scale oceanographic processes') may accentuate shifting biogeographic patterns of coastal species that are associated with the PDO. For instance, during the PDO warm phase in 1994 to 1997, the subtropical mussel *Mytilus galloprovincialis* was common at 41°N, but as the PDO shifted to cold phase by 2005 to 2007, this species was rare to absent north of 39°N (Hilbish et al. 2010). The range contraction was likely due to reproductive failure caused by cold conditions in northern sites, but was probably accelerated by southward transport conditions that are more common in the PDO cold phase (e.g. Fig. 2).

Why peaks in recruitment occur where they do across latitudes

The equilibrium population structure predicted by our connectivity models (Fig. 1) has a peak around 45°N, which is the same region where recruitment rates have been observed to have the highest levels on the US Pacific coast (e.g. Connolly et al. 2001, Broitman et al. 2008). We hypothesize that connectivity contributes to the pattern, perhaps intensifying the effects of upwelling that have been documented by others (e.g. Connolly et al. 2001, Broitman et al. 2008, Menge et al. 2011).

SUMMARY

We proposed means for extending the temporal range of connectivity estimates using a combination of multivariate statistical clustering by SOM and by using relationships between connectivity and wind magnitude and direction. We validated the method with a 30 yr hindcast of the California Current System. Only a few connectivity patterns were needed to account for much of the variance in coastal connec-

tivity over 30 yr. The weekly and yearly frequencies of these connectivity patterns were estimated using surface wind fields from meteorological reanalysis data in place of long time series of ocean model data with fine spatial and temporal scales. Therefore, at least for larvae with short planktonic durations, it is possible to make multi-decadal estimates of maximum connectivity using wind fields, so long as an adequate period of ocean data is available for parameterizing and validating a wind-based model. Analysis of the results of our hindcast indicates that there are multi-annual associations between connectivity and long-term cycles like the PDO, and that there are episodic large-scale shifts in connectivity patterns. These changes in connectivity presumably contribute to the continental scale variations in recruitment intensity and population structure of intertidal invertebrates that have been observed on the US Pacific coast over the past 2 decades.

LITERATURE CITED

Ayata SD, Ellien C, Dumas F, Dubois S, Thiébaut É (2009) Modelling larval dispersal and settlement of the reef-building polychaete *Sabellaria alveolata*: role of hydro-climatic processes on the sustainability of biogenic reefs. Cont Shelf Res 29:1605–1623

Ayata SD, Lazure P, Thiébaut E (2010) How does the connectivity between populations mediate range limits of marine invertebrates? A case study of larval dispersal between the Bay of Biscay and the English Channel (North-East Atlantic). Prog Oceanogr 87:18–36

Bakun A (1973) Coastal upwelling indices, west coast of North America, 1946–71. US Department of Commerce, NOAA Tech Rep NMFS–SSRF–671

Bakun A, Nelson CS (1991) The seasonal cycle of wind-stress curl in subtropical eastern boundary current regions. J Phys Oceanogr 21:1815–1834

Balmaseda MA, Mogensen K, Weaver AT (2013) Evaluation of the ECMWF Ocean Reanalysis ORAS4. Q J R Meteorol Soc 139:1132–1161

Barrier N, Treguier AM, Cassou C, Deshayes J (2013) Impact of the winter North-Atlantic weather regimes on subtropical sea-surface height variability. Clim Dyn 41:1159–1171

Benestad RE (2004) Empirical statistical downscaling in cli-

mate modeling. Eos Trans AGU 85:417–422, doi:10.1029/2004EO420002

Black BA, Sydeman WJ, Frank DC, Griffin D and others (2014) Six centuries of variability and extremes in a coupled marine-terrestrial ecosystem. Science 345:1498–1502

Bograd SJ, Lynn RJ (2003) Long-term variability in the southern California Current System. Deep-Sea Res II 50:2355–2370

Broitman BR, Blanchette CA, Menge BA, Lubchenco J and others (2008) Spatial and temporal patterns of invertebrate recruitment along the west coast of the United States. Ecol Monogr 78:403–421

CalCOFI (1950) Calif Coop Ocean Fish Invest Rep 1:1–54

Cassou C, Terray L (2001) Oceanic forcing of the wintertime low-frequency atmospheric variability in the North Atlantic European sector: a study with the ARPEGE model. J Clim 14:4266–4291

Coe WR (1953) Resurgent populations of littoral marine invertebrates and their dependence on ocean currents and tidal currents. Ecology 34:225–229

Coe WR (1955) Ecology of the bean clam *Donax gouldi* on the coast of southern California. Ecology 36:512–514

Compo GP, Whitaker JS, Sardeshmukh PD, Matsui N and others (2011) The Twentieth Century Reanalysis project. Q J R Meteorol Soc 137:1–28

Connolly SR, Baird AH (2010) Estimating dispersal potential for marine larvae: dynamic models applied to scleractinian corals. Ecology 91:3572–3583

Connolly SR, Menge BA, Roughgarden J (2001) A latitudinal gradient in recruitment of intertidal invertebrates in the Northwest Pacific Ocean. Ecology 82:1799–1813

Di Lorenzo E, Schneider N, Cobb KM, Franks PJS and others (2008) North Pacific Gyre Oscillation links ocean climate and ecosystem change. Geophys Res Lett 35:L08607, doi:10.1029/2007GL032838

Drake PT, Edwards CA, Barth JA (2011) Dispersion and connectivity estimates along the U.S. west coast from a realistic numerical model. J Mar Res 69:1–37

Drake PT, Edwards CA, Morgan SG, Dever EP (2013) Influence of larval behavior on transport and population connectivity in a realistic simulation of the California Current System. J Mar Res 71:317–350

Ellien C, Thiebaut E, Barnay AS, Dauvin JC, Gentil F, Salomon JC (2000) The influence of variability in larval dispersal on the dynamics of a marine metapopulation in the eastern Channel. Oceanol Acta 23:423–442

Gaylord B, Gaines SD (2000) Temperature or transport? Range limits in marine species mediated solely by flow. Am Nat 155:769–789

Gebhardt A (2013) akima: interpolation of irregularly spaced data. R package version 0.5-10. http://CRAN.R-project.org/package=akima (accessed 17 July 2014)

Gilg MR, Hilbish TJ (2003) The geography of marine larval dispersal: coupling genetics with fine-scale physical oceanography. Ecology 84:2989–2998

Gouhier TC, Guichard F, Menge BA (2010) Ecological processes can synchronize marine population dynamics over continental scales. Proc Natl Acad Sci USA 107:8281–8286

Hartigan JA, Wong MA (1979) Algorithm AS 136: A *K*-means clustering algorithm. J R Stat Soc C Appl Stat 28:100–108

Hilbish TJ, Brannock PM, Jones KR, Smith AB, Bullock BN, Wethey DS (2010) Historical changes in the distribution of invasive and endemic marine invertebrates are contrary to global warming predictions: the effects of decadal climate oscillations. J Biogeogr 37:423–431

Hines A, Siddorn J, Hackett B, McConnell N (2013) Product user manual for north-west shelf physical hindcast products, document MYO2-NWS-PUM-004-005-006. http://catalogue.myocean.eu.org/static/resources/myocean/pum/MYO2-NWS-PUM-004-005-006-V2.1.pdf (accessed 11 April 2014)

Huret M, Petitgas P, Woillez M (2010) Dispersal kernels and their drivers captured with a hydrodynamic model and spatial indices: A case study on anchovy (*Engraulis encrasicolus*) early life stages in the Bay of Biscay. Prog Oceanogr 87:6–17

Huret M, Sourisseau M, Petigas P, Struski C, Léger F, Lazure P (2013) A multi-decadal hindcast of a physical–biogeochemical model and derived oceanographic indices in the Bay of Biscay. J Mar Syst 109–110:S77–S94

Ihaka R, Gentleman R (1996) R: a language for data analysis and graphics. J Comput Graph Stat 5:299–314

Johnson NC, Feldstein SB, Tremblay B (2008) The continuum of northern hemisphere teleconnection patterns and a description of the NAO shift with the use of self-organizing maps. J Clim 21:6354–6371

Kohonen T (2001) Self-organizing maps. Springer Verlag, Berlin

Lee S, Feldstein SB (2013) Detecting ozone- and greenhouse gas-driven wind trends with observational data. Science 339:563–567

Lett C, Ayata SD, Huret M, Irisson JO (2010) Biophysical modelling to investigate the effects of climate change on marine population dispersal and connectivity. Prog Oceanogr 87:106–113

Levin LA (2006) Recent progress in understanding larval dispersal: new directions and digressions. Integr Comp Biol 46:282–297

Liu Y, Weisberg RH (2005) Patterns of ocean current variability on the West Florida Shelf using the self organizing map. J Geophys Res 110:C06003

Logemann K, Backhaus JO, Harms IH (2004) SNAC: a statistical emulator of north-east Atlantic circulation. Ocean Model 7:97–100

Mantua NJ, Hare SR (2002) The Pacific decadal oscillation. J Oceanogr 58:35–44

Marmer HA (1926) Coastal currents along the Pacific coast of the United States. US Coast & Geodetic Survey Spec Pub 121:1–80

Menge BA, Menge DNL (2013) Dynamics of coastal meta-ecosystems: the intermittent upwelling hypothesis and a test in rocky intertidal regions. Ecol Monogr 83:283–310

Menge BA, Gouhier TC, Freidenburg T, Lubchenco J (2011) Linking long-term, large-scale climatic and environmental variability to patterns of marine invertebrate recruitment: towards explaining 'unexplained' variation. J Exp Mar Biol Ecol 400:236–249

Mitarai S, Siegel DA, Watson JR, Dong C, McWilliams JC (2009) Quantifying connectivity in the coastal ocean with application to the Southern California Bight. J Geophys Res 114:C10026, doi:10.1029/2008JC005166

Morgan SG, Fisher JL, Miller SH, McAfee ST, Largier JL (2009) Nearshore larval retention in a region of strong upwelling and recruitment limitation. Ecology 90:3489–3502

Pineda J (2000) Linking larval settlement to larval transport:

assumptions, potentials, and pitfalls. Oceanogr Eastern Pacific 1:84–105

Pineda J, Reyns NB, Starczak VR (2009) Complexity and simplification in understanding recruitment in benthic populations. Popul Ecol 51:17–32

R Core Team (2012) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. www.r-project.org

Rilov G, Dudas SE, Menge BA, Grantham BA, Lubchenco J, Schiel DR (2008) The surf zone: a semi-permeable barrier to onshore recruitment of invertebrate larvae? J Exp Mar Biol Ecol 361:59–74

Saha S, Nadiga S, Thiaw C, Wang J and others (2006) The NCEP Climate Forecast System. J Clim 19:3483–3517

Siegel DA, Kinlan BP, Gaylord B, Gaines SD (2003) Lagrangian descriptions of marine larval dispersion. Mar Ecol Prog Ser 260:83–96

Siegel DA, Mitarai S, Costello CJ, Gaines SD, Kendall BE, Warner RR, Winters KB (2008) The stochastic nature of larval connectivity among nearshore marine populations. Proc Natl Acad Sci USA 105:8974–8979

Sponaugle S, Paris C, Walter KD, Kourafalou V, D'Alessandro E (2012) Observed and modeled larval settlement of a reef fish to the Florida Keys. Mar Ecol Prog Ser 453: 201–212

Strathmann MF (1981) Reproduction and development of marine invertebrates of the Northern Pacific Coast. University of Washington Press, Seattle, WA

Veneziani M, Edwards CA, Doyle JD, Foley D (2009) A central California coastal ocean modeling study: 1. Forward model and the influence of realistic climatological forcing. J Geophys Res 114:C04015, doi:10.1029/2008JC 004774

Wehrens R, Buydens LMC (2007) Self- and super-organizing maps in R: the kohonen package. J Stat Softw 21:1–19

Wolter K, Timlin MS (2011) El Niño/Southern Oscillation behavior since 1871 as diagnosed in an extended multivariate ENSO index (MEI.ext). Int J Climatol 31: 1074–1081

Wyatt B, Burt WV, Pattullo JG (1972) Surface currents off Oregon as determined from drift bottle returns. J Phys Oceanogr 2:286–293