The following supplement accompanies the article

Using salinity to identify common bottlenose dolphin habitat in Barataria Bay, LA

Fawn E. Hornsby*, Trent L. McDonald, Brian C. Balmer, Todd R. Speakman, Keith D. Mullin, Patricia E. Rosel, Randall S. Wells, Andrew C. Telander, Peter W. Marcy, Lori H. Schwacke

*Corresponding author: fhornsby@west-inc.com

Endangered Species Research 33: 181–192 (2017)

SUPPLEMENT 1

PREDICTED SALINITY IN VARIOUS BAYS, SOUNDS, AND ESTUARIES WITHIN THE NORTHERN GULF OF MEXICO

INTRODUCTION

We developed a model to predict salinity in five bays, sounds, and estuaries within the northern Gulf of Mexico (GOM). Salinity measurements compiled from numerous state and federal agencies (Table S1.1, Figure S1.1) provided historical data from 2005 through 2012. Spatially, the database covered most estuaries from Vermilion Bay, LA (in the west) to Mobile Bay, MS (in the east). A spatio-temporal geostatistical interpolation (kriging) model (Szpiro et al. 2010, Sampson et al. 2011, Lindström et al. 2014) estimated daily salinity for the entire period at points in a 200-meter by 200-meter grid covering areas of interest (Figure S1.1). We validated the model using three separate hold-out cross-validation procedures.

METHODS

Data

Eleven state and federal agencies maintain public salinity data in the northern GOM (Table S1.1). Data we compiled consisted of two types: 1) discrete—temporally and spatially sporadic one-time, hand-collected values; and 2) continuous—fine temporal scale (e.g., 15 minutes, hourly, or daily) measurements collected by autonomous water-quality monitoring stations. The number and placement of collection stations varied annually. Sporadic temporal gaps existed in both the discrete and continuous datasets when studies stopped and re-started or stations temporarily malfunctioned.

Following download, we housed the data in a single Microsoft SQL database and subjected all records to quality assurance procedures designed to identify and flag erroneous values. Our routines flagged raw values outside an expected range (0 to 40 ppt), potential outliers (>1.5 inner quartile range), and potential anomalies (>2 standard deviations different from a 4 day centered moving average at the same site). We removed values outside the expected range. For all outliers and anomalies, we examined values recorded during similar times and at nearby sites to determine whether the value was plausible or implausible. We left plausible values intact, but removed implausible values. With one exception, less than 0.5% of raw salinity values from any single data source were ultimately excluded as out of range or implausible. Finally, for modeling we computed daily averages if a continuous series reported multiple values per day.

The total number of verified daily salinity records exceeded 675,000 (Table S1.2). Based on availability of data and hydrologic similarity, we identified five estimation areas where salinity estimation was practical (Figure S1.1). We interpolated values in the following areas, moving west to east:

- The Morganza/Vermilion estimation area includes Vermilion Bay, West Cote Blanche Bay, the Atchafalaya Delta, Atchafalaya Bay, Caillou Bay, Terrebonne Bay, and Timbalier Bay off the Louisiana coast.
- The Davis Pond estimation area includes Barataria Bay, Louisiana.
- The Caernarvon estimation area includes areas east of the Mississippi River and west of the Mississippi River Gulf Outlet "(MRGO) canal skirting the western shore of Lake Borgne off the coast of Louisiana.
- The Bonnet Carré estimation area includes Lake Borgne, Bay St. Louis, Biloxi Bay, and Mississippi Sound to Mobile Bay.
- The Mobile Bay estimation area encompassed Mobile Bay, Alabama.

The number of stations providing data appears in Table S1.3.[1] Supplement 2 contains a detailed description of data sources, quality assurance procedures, and additional data we compiled beyond salinity (e.g., temperature).

Spatio-Temporal Model

In Supplement 2, we describe the spatio-temporal kriging model (Lindström et al., 2013) and bi-cubic interpolation (Akima, 1996) which estimated salinity at a 200-meter by 200-meter lattice overlaid upon the two areas of interest. This section contains a brief summary of the model and interpolation.

The general form of the spatio-temporal model was,

$$\mathrm{y(s,t)} = \mu(\mathrm{s,t}) + \varepsilon(\mathrm{s,t}),$$

where:

$y(s,t)$ = the observed salinities at location *s* and day *t*;

$\mu(s,t)$ = the mean salinity at location *s* on day *t*; and

$\varepsilon(s,t)$ = the residual for location *s* on day *t*.

A subset of some discrete data sets contained contemporaneous measurements of surface and bottom salinity at the same location. We averaged surface and bottom salinity in these cases to produce a single observation for the site because the shallow depths involved enhanced mixing of the water column and because correlation between top and bottom measurements was extremely high ($R^2 > 0.9$).

The model partitioned $\mu(s,t)$ into a sum of two linear combinations, one that potentially contained environmental covariates and one that contained a set of (smooth) temporal basis functions with spatially varying coefficients (Lindström et al. 2013). The temporal basis functions and their spatial coefficients allowed us to estimate different (and complex) temporal trends at each site. Because agencies rarely reported potential environmental covariates, we simplified the first linear combination in $\mu(s,t)$ to contain only annual means, while we modeled the residual field as time-independent and spatially-correlated.

We used the SpatioTemporal package in R (Lindström et al. 2013) to estimate separate spatio-temporal models for each area (Davis Pond and Caernarvon). The spatio-temporal models estimated salinity at points in a courser-than-desired 1000-meter X 1000-meter prediction grid because estimation of the model at our desired resolution (200-meters) was not computationally feasible. To achieve our desired resolution, the bi-cubic interpolation procedure of Akima (1996) refined spatio-temporal model estimates from the initial 1000-meter prediction grid to the desired 200-meter grid (akima R package, Akima et al. 2013). The final grids contained daily salinity estimates at all locations in the 200-meter lattice from 1-Jan-2006 through 31-Dec-2012.

Model Validation

Following estimation, we applied three procedures to validate the accuracy and reliability of values in the final grids. The first validation procedure held-out discrete values collected at stations associated with oyster sites sampled in 2010 and 2011 during the natural resource damage assessment (NRDA data from oyster studies).

[1] Note that a small portion of eastern Caernarvon estimation area overlaps the western portion of the Bonnet Carré area. Due to this overlap, a few stations were located within both areas, their data contributed to models in both estimation areas, and counts in Tables S1.1.S1.3 for Bonnet Carré and Caernarvon are not mutually exclusive.

These stations were located nearby sampled oyster sites and this validation specifically tested performance in areas of known interest. To explore agreement between modeled and measured salinities, we computed correlations (R^2) and match rates, defined as the percentage of cases where the hold-out measurement and its interpolated value were either both above or both below 3ppt, 5ppt and 8ppt, respectively.

The second validation procedure temporarily removed 20% of the continuous salinity sites reporting data in 2010 near sampled oyster sites. A spatially balanced random sampling procedure (available in ESRI's *Geostatistical Analyst Sampling Network Design Toolset*) of continuous salinity recorder sites near oyster sites selected the specific continuous stations for removal. For this validation procedure, we again compared predictions from the re-estimated spatio-temporal model to held-out data at the locations and dates of the held-out data. Again, we computed correlations and match rates for 3ppt, 5ppt, and 8ppt thresholds.

The third validation procedure compared salinity predictions to measured values in an external dataset collected in 2011 by the DWH Fish Technical Workgroup (NOAA 2011a, 2011b). This procedure computed monthly correlation, average difference, average mean square error, and percent differences to explore seasonal agreement of predictions and measured salinities.

RESULTS

Model Output

The estimation procedure produced 14,600 200-meter resolution raster maps (365 days per year (discounting leap years) × 5 estimation areas × 8 years = 14,600 maps). The full set of these maps is available via the Data Integration, Visualization, Exploration and Reporting (DIVER) system (NOAA 2015). For illustration here, we show one map during spring (1-May), summer (1-Aug), and fall (1-Nov) of 2009, 2010, and 2011 (Figures S1.2 through S1.6).

Model Validation Results

Results from the first cross-validation exercise using discrete data near oyster sites indicated strong correlation between modeled salinity and hold-out values (Table S1.4). Match rates exceeded 90% with two exceptions (5ppt and 8ppt threshold in Caernarvon during 2010), and reached 100% in several cases. Similarly, the second hold-out validation against a spatially balanced sample of continuous sites near oyster sites showed high match rates and strong correlation between held-out and modeled values (Table S1.5).

The third validation exercise against the independent set of measurements collected by NRDA fish studies showed primarily positive monthly differences, indicating over prediction of measured salinity (Table S1.6, Figure S1.7). Mean monthly differences ranged from 1.31ppt to 2.16ppt for the months of June through August 2011 (average = 1.77 ppt). During the drier season (September and October 2011) mean differences ranged from 0.40 ppt to -0.67 ppt, and averaged -0.13 ppt. The root mean squared error (RMSE) of salinity prediction differences (Table S1.6) was generally greater than the mean, resulting in an average coefficient of variation (CV) of 163%. Because most values were small, relative differences averaged 26% during June through August and 7% during September and October.

DISCUSSION

This study leveraged data collected by multiple agencies to predict salinities over large geographic regions with high resolution. The larger database from which the salinity records came contained over 35 million verified records collected from 1-Jan-2005 through 31-Dec-2012 at stations from eastern Texas to western Florida. We used approximately 675,000 of the records to fit a spatio-temporal kriging models in five estimation areas and produced daily salinity maps on a 200-meter grid covering both areas. Our database is among the largest single salinity databases ever compiled for the northern Gulf of Mexico, and our maps are among the highest resolution and largest. Our model validation, which compared of model predictions to field measurements, showed high correlations and high match rates, but also approximately 1 to 2 ppt over-prediction at lower (<15ppt) salinity levels. Overall, cross validation correlations between modeled and measured salinity was generally very high, with match rates often exceeding 90% in various investigated basins and years.

Due to high correlation with measured values, the modeled predictions of salinity are adequate for use in other studies as explanatory covariates. Based on maps of prediction standard errors, regions within 400 meters to 600 meters of a measurement station that has provided concurrent observations should be <1 ppt from the actual salinity value.

LITERATURE CITED

Akima H (1996) Rectangular-grid-data surface fitting that has the accuracy of a bicubic polynomial. J ACM 22:357-361.

Akima H, Gebhardt A, Petzold T, Maechler M (2013) akima: Interpolation of irregularly spaced data. R package, Version 0.5-11. http://CRAN.R-project.org/package=akima

Lindström J, Szpiro A, Sampson PD, Bergen S, Oron AP (2013) SpatioTemporal: Spatio-Temporal Model Estimation. https://cran.r-project.org/web/packages/SpatioTemporal/index.html

Lindström J, Szpiro A, Sampson PD, Oron AP, Richards M, Larson TV, Sheppard L (2014). A Flexible Spatio-Temporal Model for Air Pollution with Spatial and Spatio-Temporal Covariates. Environ Ecol Stat 21:411-433.

Loosanoff, V.L. (1948). Effects of turbidity on feeding of oysters. Proc. Natl. Shellfish. Assoc. 1947:40-44.

National Oceanic and Atmospheric Administration (NOAA) (2011a) Submerged Oil Characterization Across Multiple Habitats for Assessment of Persistent Exposures in Nearshore Sediments Deepwater Horizon Oil Spill (DWHOS). http://www.gulfspillrestoration.noaa.gov/wp-content/uploads/2012/05/2011_07_06_NEARSHORE_Submerged-Oil-NOAA-BP_LA-Signature.redacted2.pdf.

National Oceanic and Atmospheric Administration (NOAA) (2011b) Assessment Plan for Marsh Edges and Sandy Shorelines, Version 5. http://www.gulfspillrestoration.noaa.gov/wp-content/uploads/MarshedgeSandyShoreline_AllSigned.Redacted.pdf.

National Oceanic and Atmospheric Administration (NOAA) (2015) Deepwater Horizon Natural Resource Assessment Data: Data Integration, Visualization, Exploration and Reporting (DIVER). https://dwhdiver.orr.noaa.gov.

Sampson PD, Szpiro AA, Sheppard L, Lindström J, Kaufman JD (2011) Pragmatic estimation of a spatio-temporal air quality model with irregular monitoring data. Atmos Environ 45:6593-6606.

Szpiro AA, Sampson PD, Sheppard L, Lumley T, Adar SD, Kaufman JD (2010) Predicting Intra-Urban Variation in Air Pollution Concentrations with Complex Spatio-Temporal Dependencies. Environmetrics 21:606–631.

TABLES

Table S1.1. Data sources contributing salinity information delineated by type of data supplied and contact method.

Data Source	Name	Data Type	Website/Contact Method
National Estuarine Research Reserve System	NERRS	Continuous	http://www.nerrs.noaa.gov/
Louisiana Office of Coastal and Protection and Restoration	OCPR-Continuous	Continuous	http://coastal.louisiana.gov/
U.S. Geological Survey	USGS	Continuous	http://waterdata.usgs.gov/nwis
Alabama Department of Public Health	ADPH	Discrete	Contact agency personnel
Louisiana Department of Environmental Quality	LDEQ	Discrete	Contact agency personnel
Louisiana Department of Health and Hospitals	LDHH	Discrete	Contact agency personnel
Louisiana Department of Wildlife and Fisheries	LDWF	Discrete	Contact agency personnel
Mississippi Department of Environmental Quality	MDEQ	Discrete	Contact agency personnel
Mobile Bay National Estuary Program	MBNEP	Discrete	http://www.mobilebaynep.com/
NRDA Oyster TWG sampling plans	NRDA	Discrete	https://dwhdiver.orr.noaa.gov/
Louisiana Office of Coastal and Protection and Restoration	OCPR-Discrete	Discrete	Contact agency personnel
STOrage and RETrieval and Water Quality eXchange	STORET	Discrete	http://www.epa.gov/storet/

Table S1.2. Number of salinity and temperature records contributed by each of eleven agencies, by year. Numbers reflect observations which fell in one of the five estimation areas considered in this report. Additional records exist pre-2005, post-2012, and outside the five estimation areas that were not used in this study (See Supplement 2).

	Continuous Data			Discrete Data								
Year	NERRS	OCPR-Continuous	USGS	ADPH	LDEQ	LDHH	LDWF	MDEQ	MBNEP	NRDA Oyster	OCPR-Discrete	STORET
2005	0	18995	5937	0	266	5814	1268	279	0	0	580	0
2006	0	34050	6400	0	182	6778	1529	272	0	0	1938	0
2007	0	56710	6682	0	0	7288	1540	275	0	0	4535	0
2008	0	78579	6220	0	108	6962	1560	265	0	0	5650	0
2009	2084	81614	7332	0	171	7412	1743	0	1405	0	6893	172
2010	1911	83848	7718	69	109	7257	1622	15	1398	243	6683	77
2011	1980	83939	8096	43	119	6405	1486	277	1425	948	6784	523
2012	1980	69154	7891	31	0	0	1431	0	1464	496	823	254
Total	7955	506889	56276	143	955	47916	12179	1383	5692	1687	33886	1026

Note: ADPH = Alabama Department of Public Health; LDEQ = Louisiana Department of Environmental Quality; LDHH = Louisiana Department of Health and Hospitals; LDWF = Louisiana Department of Wildlife and Fisheries; MDEQ = Mississippi Department of Environmental Quality; MBNEP = Mobile Bay National Estuarine Program; NERRS = National Estuarine Research Reserve System; NRDA Oyster = Natural Resource Damage Assessment Oyster Technical Working Group; OCPR = Office of Coastal and Protection and Restoration; USGS = U.S. Geological Survey.

Table S1.3: Number of stations providing continuous and discrete data within each of the five estimation areas.

Year	Data Type	Number of Stations within estimation area					Total
		Morganza/ Vermilion	Davis Pond	Caernarvon	Bonnet Carré	Mobile	
2005	Continuous	24	42	23	6	NA	95
	Discrete	230	218	160	152	NA	760
2006	Continuous	87	59	22	6	NA	174
	Discrete	376	290	139	152	NA	957
2007	Continuous	116	75	33	16	NA	240
	Discrete	1038	572	276	242	NA	2128
2008	Continuous	140	84	35	19	NA	278
	Discrete	1087	601	263	253	NA	2204
2009	Continuous	140	84	34	21	NA	279
	Discrete	1674	868	393	287	NA	3222
2010	Continuous	140	83	32	22	4	281
	Discrete	1692	842	409	345	27	3315
2011	Continuous	140	83	32	22	4	281
	Discrete	1707	931	450	507	65	3660
2012	Continuous	139	74	26	20	4	263
	Discrete	276	163	85	140	39	703

Table S1.4. Hold-out cross validation results of the spatio-temporal model for salinity with discrete samples near oyster sites temporarily removed, by basin and year. Match rate is the percentage of cases where the hold-out measurement and its interpolated value were either both above or both below the given salinity threshold (either 3ppt, 5ppt, or 8ppt).

Basin	Year	Hold-out Data		Measured vs Modeled			
		Station Counts	Measurement Counts	R^2	Sig. (p value)	Match Rate (5ppt)	Match Rate (10 ppt)
Morganza-Vermillion	2010	13	59	0.75	< 0.001	100.0%	91.5%
	2011	16	70	0.68	< 0.001	90.0%	87.1%
	2012	18	59	0.16	0.002	100.0%	79.7%
Davis Pond	2010	11	53	0.90	< 0.001	94.4%	100.0%
	2011	33	72	0.77	< 0.001	100.0%	100.0%
	2012	44	114	0.80	< 0.001	97.4%	86.0%
Caernarvon	2010	8	33	0.63	< 0.001	81.8%	81.8%
	2011	22	135	0.68	< 0.001	97.8%	91.9%
	2012	22	92	0.01	0.433	90.2%	82.6%
Bonnet Carré	2010	34	91	0.57	< 0.001	100.0%	100.0%
	2011	79	269	0.58	< 0.001	99.6%	94.4%
	2012	77	254	0.51	< 0.001	98.8%	88.6%

Table S1.5. Hold-out cross validation results of the spatio-temporal model after removing a 20% spatially balanced random sample of the continuous monitoring stations in the vicinity of oyster sites in Davis Pond and Caernarvon. The match rate is the percentage of cases where the hold-out measurement and its interpolated values were both above or both below the given salinity threshold (either 3ppt, 5ppt, or 8ppt).

		Hold-out Data		Measured vs Modelled				
Basin	**Year**	**Station Counts**	**Measurement Counts**	R^2	**Sig. (p value)**	**Match Rate (3ppt)**	**Match Rate (5ppt)**	**Match Rate (8ppt)**
Davis Pond	2010	16	485	0.86	< 0.001	97.7%	93.8%	93.2%
Caernarvon	2010	5	561	0.76	< 0.001	90.7%	87.7%	87.3%

Table S1.6. Results of the salinity model validation in the Davis Pond area using NRDA Fish TWG data as a reference and the predicted values from the spatio-temporal kriging models aggregated by month. In the table, RMSE is root mean squared error for predictions around measured values and is akin to standard error. MRPE represents the mean relative absolute prediction error for the salinity predictions.

Month	Year	Mean Difference (Prediction-Observed)	RMSE	MRPE	n
June	2011	1.845358	2.313245	0.228927	205
July	2011	1.312724	1.705225	0.282058	53
August	2011	2.160987	3.138116	0.269274	53
September	2011	0.405799	0.670206	0.061779	6
October	2011	-0.67232	1.697888	0.07921	17

FIGURES

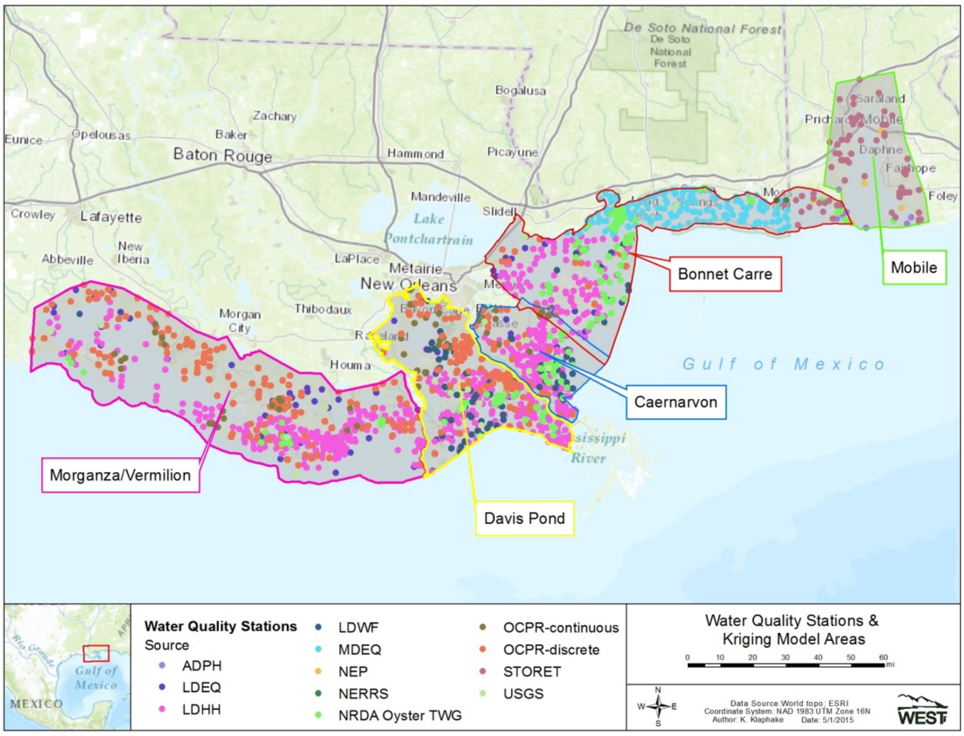

Figure S1.1. Estimation areas, station locations, and data source used to model salinity and temperature during 2005 to 2012. Most stations were discrete and provided only a few raw data values. See Figure 2 in the main article for a pictorial representation of temporal coverage. Additional data outside these estimation areas, pre-2005, and post-2012 are available.

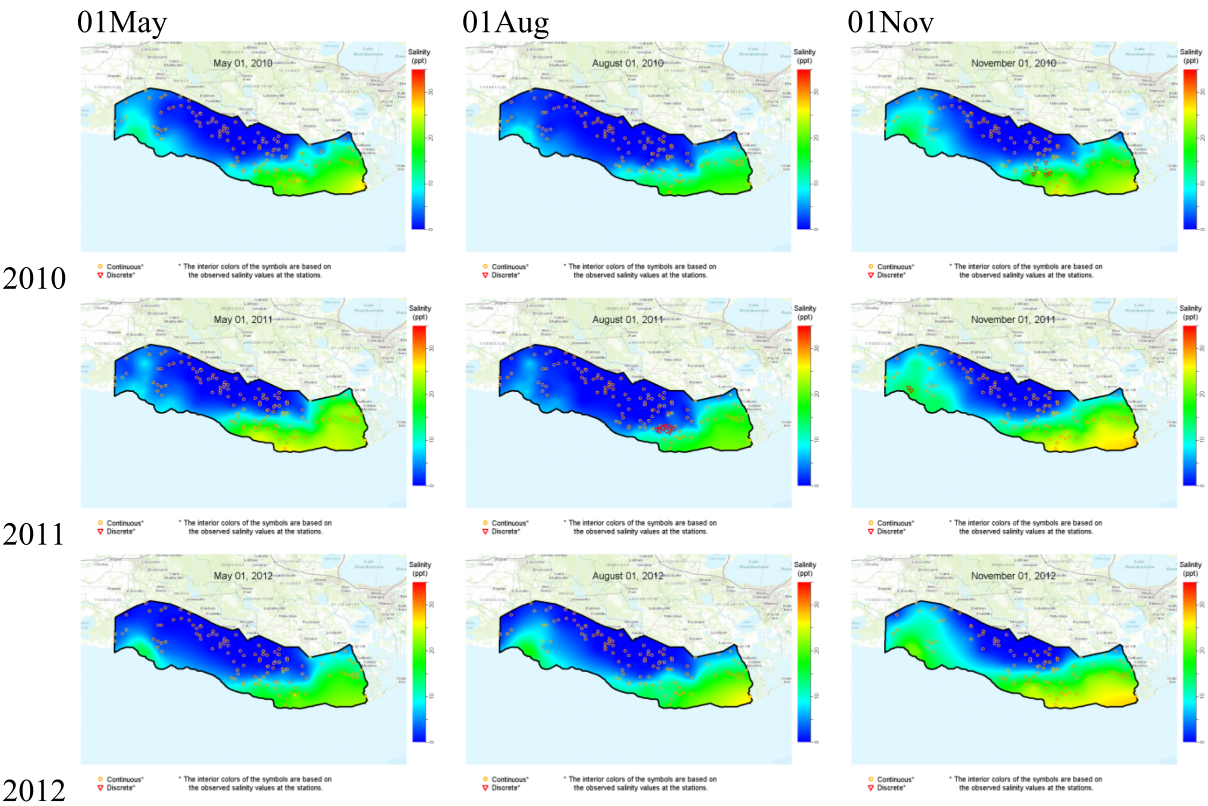

Figure S1.2. Estimated salinity in Morganza/Vermilion (MV) estimation area on three representative days in 2010, 2011, and 2012. Dots in the map indicate the location of salinity stations used to produce the estimated salinity values for the given day/year.

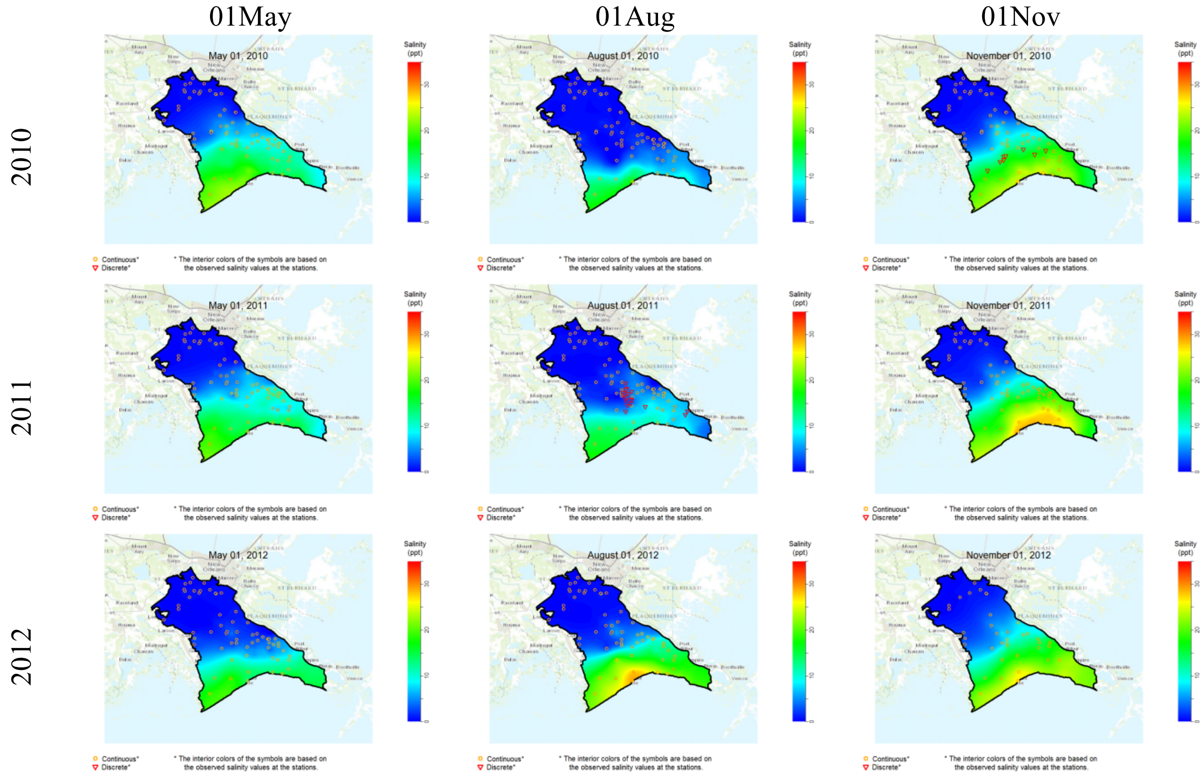

Figure S1.3. Estimated salinity in Davis Pond (DP) estimation area on three representative days in 2010, 2011, and 2012. Dots in the map indicate the location of salinity stations used to produce the estimated salinity values for the given day/year.

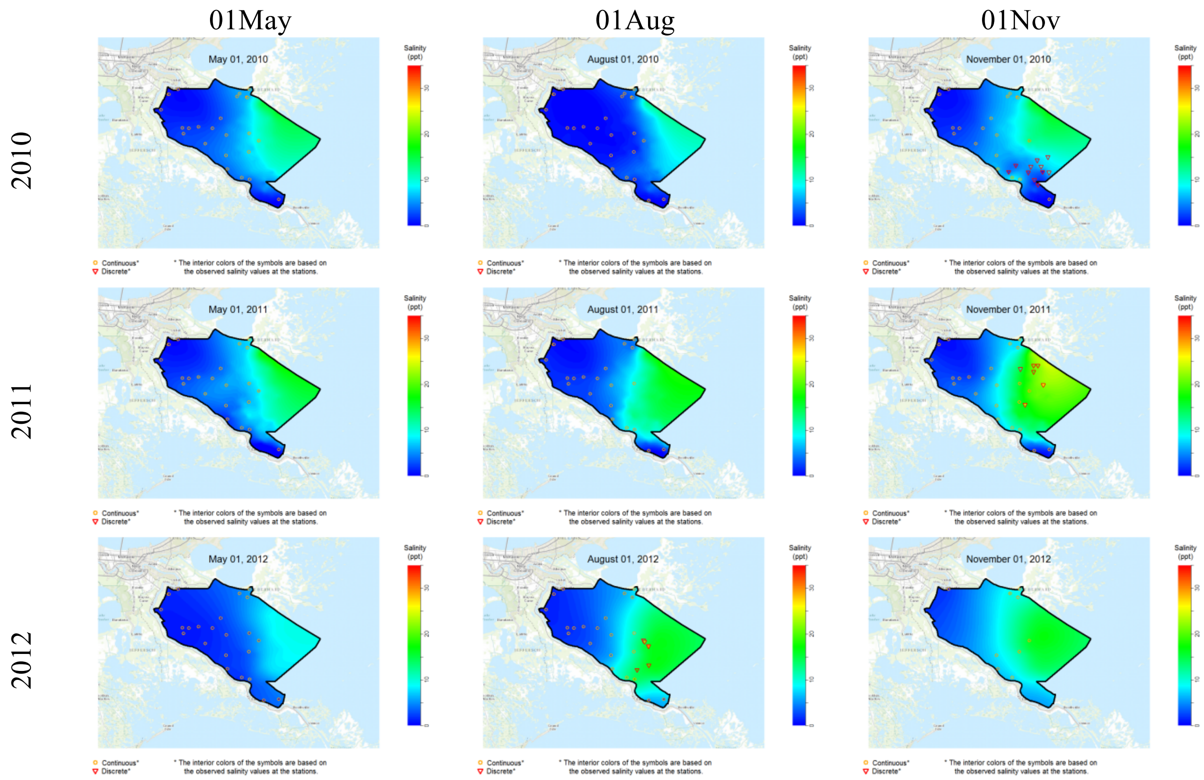

Figure S1.4. Estimated salinity in Caernarvon (CN) estimation area on three representative days in 2010, 2011, and 2012. Dots in the map indicate the location of salinity stations used to produce the estimated salinity values for the given day/year.

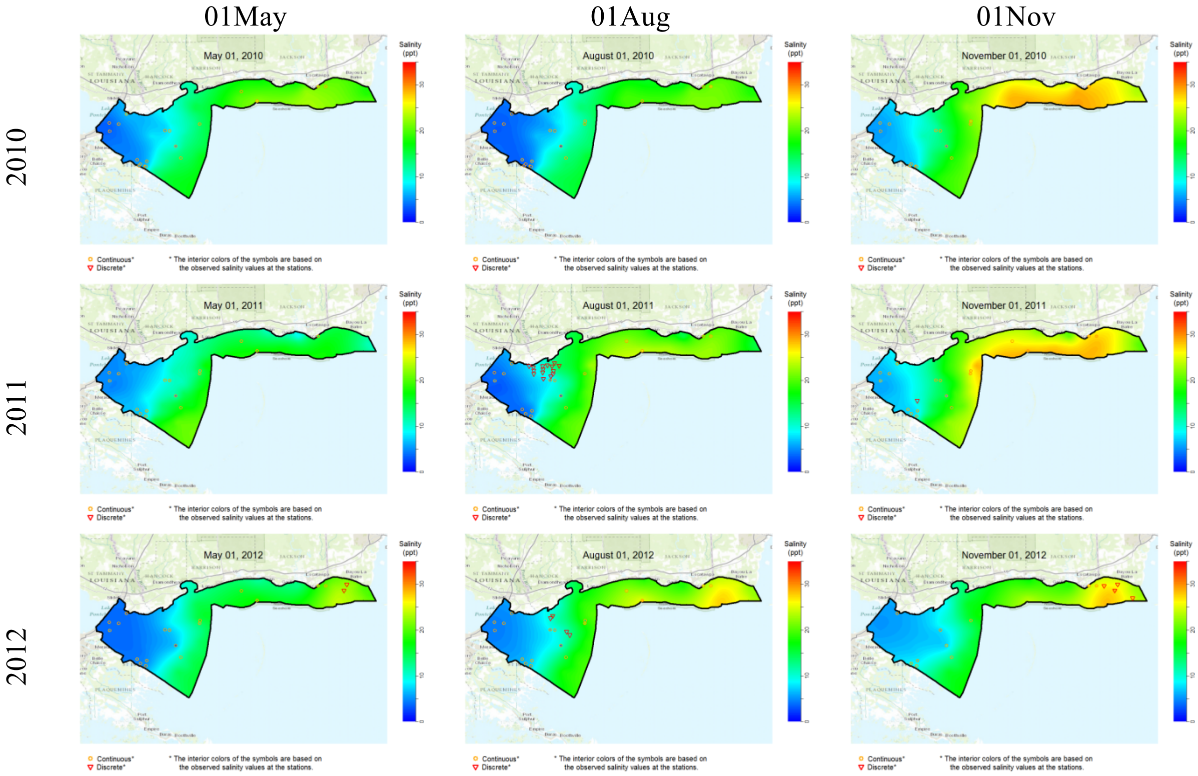

Figure S1.5. Estimated salinity in Bonnet Carré (BC) estimation area on three representative days in 2010, 2011, and 2012. Dots in the map indicate the location of salinity stations used to produce the estimated salinity values for the given day/year.

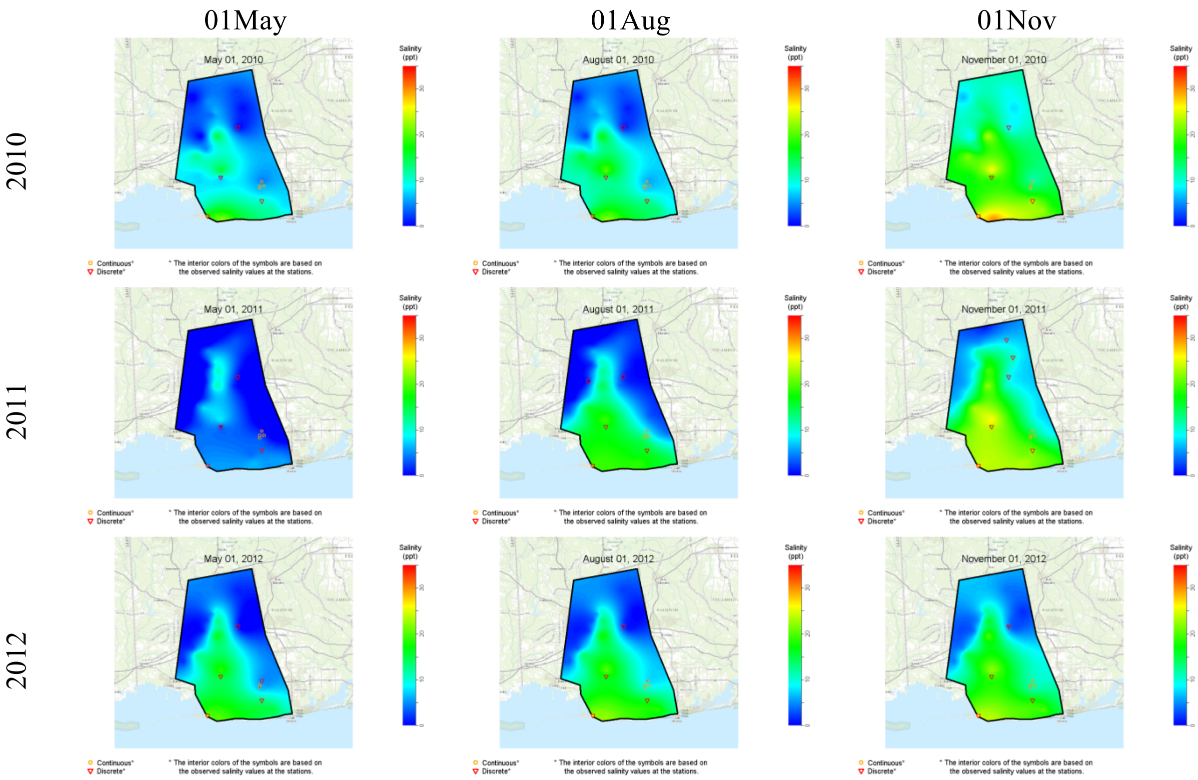

Figure S1.6. Estimated salinity in Mobile Bay (MB) estimation area on three representative days in 2010, 2011, and 2012. Dots in the map indicate the location of salinity stations used to produce the estimated salinity values for the given day/year.

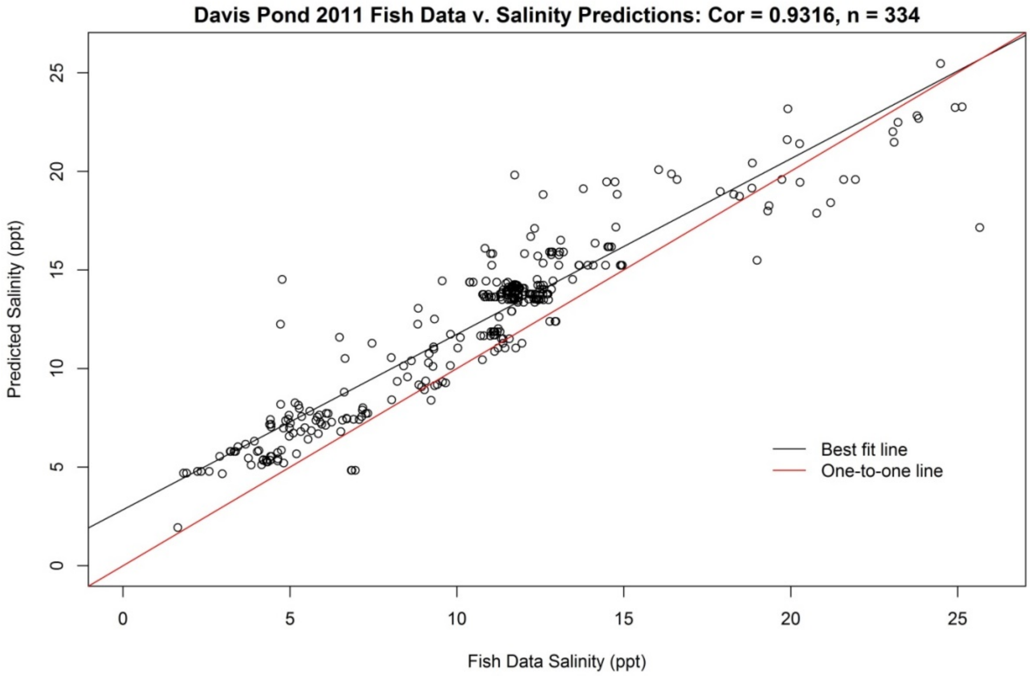

Figure S1.7. Scatter plot comparing salinity predictions in the Davis Pond area to discrete measurements at the same place and time. All measurements were taken in 2011