

Further Analysis and Synthesis of Narragansett Bay (RI/MA USA) Oxygen, Chlorophyll, and Temperature

Daniel L. Codiga, Sole Proprietor

Academic Affiliation: University of Rhode Island,
Graduate School of Oceanography

**NARRAGANSETT BAY
ESTUARY PROGRAM**



FINAL REPORT

June 30, 2020

Narragansett Bay Estuary Program

Funded by the Environmental Protection Administration
through
New England Interstate Water Pollution Control Commission

NEI Job Code: 318-002
Project Code: S-2018-006

NBEP Contribution Number: NBEP-20-231A
EPA Grant Number: USEPA CE00A00004

Full citation:

Codiga, D.L. 2020. Further Analysis and Synthesis of Narragansett Bay (RI/MA USA) Oxygen, Chlorophyll, and Temperature. NBEP Technical Report. NBEP-20-231A. URL: <https://figshare.com/s/7d51f2540df6638a4552>. DOI: 10.6084/m9.figshare.12547676.

Abstract

Several related analyses of Narragansett Bay oxygen, chlorophyll, temperature, and influencing factors (nitrogen load, river flow, salinity, stratification, physical drivers) through 2017 are reported to help improve understanding of hypoxia and eutrophication. They build on and address identified gaps in the findings, on years through 2015, of the State of Narragansett Bay and Its Watershed (Narragansett Bay Estuary Program, 2017). Daily-resolution 2001-2017 bay-wide loads of total nitrogen from wastewater treatment facilities and surrounding watersheds are estimated and interpreted in a companion technical report. The seasonal Hypoxia Index results from fixed-site monitoring network near-bottom oxygen time series are updated to add, to the 10 sites previously treated, the shallow Phillipsdale Landing site in the estuarine headwaters of the tidal Seekonk River. At this location hypoxia is the most severe and occurs not only in years with high river flow but also, unlike other sites, in many dry years which may be related to longer water residence times. For both the relatively dry year 2016 and the relatively wet year 2017 hypoxia was among the least severe to date bay-wide. Weaker hypoxia in 2017 than prior years with comparable river flow is one of the strongest indications that bay hypoxia has responded to reduced nutrient loads. When river flow is intermediate both stratification and nitrogen load are generally intermediate, but either strong or weak hypoxia has occurred, suggesting that other factors are important under these conditions; temperature and physical drivers including wind speed and direction, tidal range, and non-tidal sea level difference between Providence and Newport are examined and found not to explain the variability. Oxygen and chlorophyll measurements from vessel-based spatial surveys are analyzed together with the fixed site time series to explore their complementary strengths, determine spatial and temporal decorrelation scales, and quantify correlations between spatial survey percent area metrics and time series metrics. The seasonal Chlorophyll Index from time series measurements, refined here to better capture regional patterns and changes, reveals inter-annual variability that is more independent from site to site and less tightly linked to river flow than oxygen. Spatial survey chlorophyll, from all areas of the bay both shallow and deep, shows a stronger decline during the past several years after load decreases than is seen in the Chlorophyll Index from time series observations, which are mostly from deeper locations; however, in 2017 both declined markedly. Long-term trends spanning all years of fixed site time series observations include warming of surface waters at a rate comparable to that seen in independent analyses region-wide, and warming of deep waters at a rate about twice as high. Salinities are decreasing, more strongly near the surface. Stratification is declining, due mainly to salinity and also weakly due to temperature, but at a rate unlikely to strongly impact hypoxia.

Executive Summary

To better understand Narragansett Bay hypoxia and eutrophication, several related analyses were completed using measurements of oxygen, chlorophyll, temperature, and influencing factors (nitrogen load, river flow, salinity, stratification, physical drivers) through 2017. These analyses help augment the findings, on years through 2015, reported by the Narragansett Bay Estuary Program (NBEP) in the State of Narragansett Bay and Its Watershed (NBEP, 2017; “SNBIW” hereafter) and address identified gaps.

Nitrogen loads. New estimates of total nitrogen load to the bay from wastewater treatment facilities (WWTFs) and surrounding watersheds were made. WWTF and river monitoring observations were obtained from the Rhode Island Department of Environmental Management (RIDEM), the Narragansett Bay Commission (NBC), and the Fall River WWTF. The budget treats 18 sources (as in SNBIW Chapter 8): 11 WWTFs with direct bay input (9 in Rhode Island, 2 in Massachusetts); 6 rivers, which include load from upstream WWTFs; and runoff direct to the bay from ungauged riparian areas. A companion report (“Daily-Resolution 2001-2017 Time Series of Total Nitrogen Load to Narragansett Bay from Bay-Wide Treatment Facility and Watershed Sources”; Codiga, 2020) presents these results (Task 2).

The new load time series are useful for analysis of hypoxic events, which have timescales of days to weeks, because they are the first with suitable temporal resolution. Annual averages agree acceptably well with the more accurate annual-mean load budgets used to assess reductions due to WWTF upgrades (SNBIW Chapter 8), although uncertainties are larger, particularly in the several earliest years when fewer measurements were available. The new results quantify bay-wide load reductions due to upgrades that occurred over several years and were mostly complete by 2012-2013; here the initial “post-reduction” year is taken to be 2014. They also highlight more pronounced post-reduction seasonality, associated with low May-October loads due to regulatory compliance.

Hypoxia. SNBIW Chapter 15 oxygen analyses used time series observations from the Narragansett Bay Fixed Site Monitoring Network (NBFSMN) maintained by the University of Rhode Island Graduate School of Oceanography (URI/GSO), RIDEM, NBC, and the Narragansett Bay National Estuarine Research Reserve. These analyses have been updated (Task 1) with two more years, through 2017, and expanded to add the shallow Phillipsdale Landing site in the estuarine headwaters of the tidal Seekonk River.

The seasonal Hypoxia Index (defined in SNBIW Chapter 15) results for Phillipsdale indicate it is the most severely hypoxic station (Task 4). In addition, inter-annual variability of hypoxia at Phillipsdale is not as closely linked to river runoff variations as at other sites. At Phillipsdale severe hypoxia occurred during years with relatively high river runoff, similarly to other sites, but in contrast also occurred during normal or dry years. This suggests hydrodynamic processes at Phillipsdale influence hypoxia differently, possibly by more strongly enhancing it during drier years due to longer Seekonk River residence times.

The 2016-2017 results improve our view of the post-reduction response of bay hypoxia. The SNBIW analysis treated years through 2015, reporting higher Hypoxia Index in 2013 when river flow was higher than typical and lower Hypoxia Index during the low river flow years 2014 and 2015. The present analysis appends the relatively dry 2016 and relatively wet 2017 (bay-wide river flow classified intermediate, relative to 2001-2017). For both years, Hypoxia Index results are among the lowest to date, except at the two shallowest embayment stations. Weaker hypoxia in 2017 than prior years with comparable river flow is one of the strongest indications that load decreases have reduced hypoxia.

Bay-wide river input was estimated (Task 6) using a newly developed two-stage method of scaling up individual rivers for ungauged area in their watersheds, then scaling up the sum of the rivers for ungauged areas draining directly to the bay. A new method, suitable for use in future years, was developed and applied to designate the resulting bay-wide river flow of individual years as wet, intermediate, or dry; this builds on the SNBIW, in which years were designated as wet or dry.

The relative importance of stratification and nitrogen load, both of which increase with river flow, to inter-annual variations in seasonal hypoxia was investigated at the bay-wide scale using stations outside the shallow embayments (Tasks 5 and 6). Stratification is correlated to river flow, modestly more strongly than the correlation of nitrogen load with river flow. With the exception of 2017 as just noted, high river flow brought strong stratification, high nitrogen load, and severe hypoxia; low river flow brought weak stratification, low nitrogen load, and weak hypoxia. When river flow is intermediate both stratification and nitrogen load are intermediate, but either strong or weak hypoxia has occurred, suggesting that other factors in addition to stratification and nitrogen load are important under these conditions. The potential roles of temperature and physical drivers (Task 3) including wind speed and direction, tidal range, and non-tidal sea level difference between Providence and Newport were explored. None explains the variability, suggesting the importance of other processes not examined.

Vessel-based spatial surveys of the bay led by Brown University, while less frequent than sampling of the NBFMSMN time series, complement them with better spatial coverage and resolution. Spatiotemporal variability in near-bottom oxygen is examined here in the first extensive analysis of the two datasets together (Task 7). Typical length scales and time scales of decorrelation are 7-8 km and 3-5 days respectively, with somewhat higher values in the deeper channelized areas. The percent hypoxic area, as determined on individual spatial surveys, correlates well with time series oxygen on the survey dates at deeper sites in the Providence River, Upper Bay, and to some extent Upper West Passage. Seasonal averages of percent hypoxic area also correlate with the seasonal Hypoxia Index from time series at these stations, though the correlation is weaker, likely due to the lower temporal resolution of the spatial surveys. Overall, these findings can help guide design of future monitoring efforts; while generally representative of each other and portraying similar spatiotemporal patterns of variability in deep oxygen, the two datasets have distinct strengths and weaknesses.

Chlorophyll. NBFMSMN near-surface chlorophyll time series were analyzed in SNBIW Chapter 16 with event-based index calculations using site-specific 80th percentile concentrations as thresholds. A group of phytoplankton experts, consulted (Task 10) to improve the index and thresholds, recommended (a) event-based analysis for three fixed bay-wide thresholds (20th, 50th, and 80th percentiles of time series from all years through 2017 at all sites) to complement the site-specific thresholds, and (b) adoption of “Chlorophyll Index”, instead of “Chlorophyll Bloom Index” as used in SNBIW, as a better descriptor.

SNBIW Chapter 16 results for site-specific thresholds were updated to include Phillipsdale and the years 2016 and 2017 (Task 9). As expected, Phillipsdale chlorophyll levels are among the highest of all stations. The Chlorophyll Index was then computed using the three new bay-wide thresholds, which are better suited to help identify patterns and trends across the bay (Task 10). For the highest threshold, the Chlorophyll Index is non-zero in most years only at sites in or near shallow embayments, which are most prone to blooms, toward the north and west; for the lowest threshold, the Chlorophyll Index is non-zero in most years at all sites except the one farthest south. Thus results from the bay-wide thresholds give a full portrayal of chlorophyll conditions at NBFMSMN sites. Characteristics common to results from the

three thresholds include the north-south gradient (e.g., SNBIW Chapter 16), and inter-annual variability that is independent from site to site and not tied strongly to river flow, unlike hypoxia. The Chlorophyll Index during post-reduction years is not strongly dissimilar to earlier years, until a bay-wide decline in 2017; as for hypoxia, the time series suggest 2017 is the first post-reduction year with marked decline.

Near-surface chlorophyll from vessel-based spatial surveys and the NBFSMN time series were analyzed together (Task 11), parallel to the deep oxygen analysis. Spatial and temporal decorrelation scales for chlorophyll are about 5 km and 2 days, both shorter than for oxygen. In both datasets there is a pattern of weakly negative correlations between areas north and south of the Providence River Estuary, suggesting a tendency for alternating timing of blooms and post-bloom declines in these two regions. The percent high-chlorophyll area, computed from spatial surveys using surface area with values above fixed thresholds, declined over at least the final several years of the record. As noted above, this is not seen as clearly in the time series, which mainly sample the deeper channelized areas, suggesting that chlorophyll declines due to load reductions may have occurred earlier and more strongly in shallower areas. Correlations between percent high-chlorophyll area from spatial surveys and Chlorophyll Index results from time series are generally weak, consistent with the short decorrelation scales.

In support of URI/GSO and NBC laboratory staff, progress was made (Task 8) toward potential post-deployment recalibration improvements to the NBFSMN chlorophyll time series, using results from acetone-extraction based laboratory analysis of “grab” field samples (SNBIW Chapter 16). More work remains before deciding whether recalibration is justified. If a recalibration is implemented, some of the findings here based on chlorophyll time series (Tasks 9, 10, 11) may need to be revisited.

Long-term trends in temperature, salinity, and stratification. Trends were analyzed using both near-surface and near-bottom NBFSMN time series observations (mid-May to mid-October, from 2001 for two sites and 2004-05 for other sites, through 2017; Task 12). Near-surface temperatures are warming, with average rate typical of an individual station 0.3-0.4 °C per decade over the years sampled. This is consistent with past work, including SNBIW Chapter 1, using independent datasets. An exception to the warming is the Mount Hope Bay station, where near-surface waters have cooled (average rate 0.5 °C per decade); this is likely due to changes in Brayton Point Power Station heat input, which was reduced then discontinued during roughly the second half of the record. Near-bottom temperatures, apparently investigated for the first time here, are warming bay-wide at about twice the rate of near-surface waters. This likely results from warming offshore, where deep water originates. Near-surface salinity has an increasing trend with typical average rate 1-2 PSU per decade, while near-bottom salinity is also increasing but roughly half as fast. Increasing salinities are likely due very low river runoff during several years late in the record. The salinity trend thus seems unlikely to continue; analyses of periods longer than treated here (e.g. SNBIW Chapter 2) show long-term precipitation increases due to regional-scale processes. The trend in stratification is decreasing (typical rate 1 kg m⁻³ per decade), mostly due to the difference between shallow and deep salinity trends, though the difference between shallow and deep temperature trends also decreases stratification weakly. Based on the relationship between hypoxia and stratification, the stratification trend magnitude suggests it will influence hypoxia weakly. The warming trend is likely more impactful to eutrophication and hypoxia, by increasing metabolic rates of chlorophyll and bacterial oxygen consumption, than the trends in salinity or stratification.

Products. Data files and documentation, for all datasets and analysis code developed, are provided online at <https://figshare.com/s/7d51f2540df6638a4552>.

Contents

Abstract.....	1
Executive Summary.....	2
List of Figures	7
List of Tables	9
Acronyms	10
Task 1 Phillipsdale Landing Time Series Data Reduction	11
1.1 Scope.....	11
1.2 Methods.....	11
1.3 Results.....	11
Task 2 Development of Nitrogen Load Time Series.....	13
2.1 Scope.....	13
2.2 Methods.....	13
2.3 Results.....	13
Task 3 Retrieval and Reduction of Wind, Tidal Range, and Sea Level Data	15
3.1 Scope.....	15
3.2 Methods.....	15
3.3 Results.....	15
Task 4 Update SNBIW Oxygen Results with Phillipsdale	16
4.1 Scope.....	16
4.2 Methods.....	16
4.3 Results.....	16
Task 5 Inter-Annual Variability of Hypoxia and Stratification.....	27
5.1 Scope.....	27
5.2 Methods.....	27
5.3 Results.....	27
Task 6 Drivers of Inter-Annual Hypoxia Variability	35
6.1 Scope.....	35
6.2 Methods.....	35
6.3 Results.....	38
Task 7 Time Series and Spatial Survey Datasets Together: Oxygen.....	48
7.1 Scope.....	48
7.2 Methods.....	48
7.3 Results.....	49
Task 8 Toward Improved Chlorophyll Time Series Using Grab Samples.....	78
8.1 Scope.....	78
8.2 Methods.....	78
8.3 Results.....	79
Task 9 Update SNBIW Chlorophyll Results with Phillipsdale	97
9.1 Scope.....	97
9.2 Methods.....	97
9.3 Results.....	97

Task 10 Refine Chlorophyll Index.....	103
10.1 Scope.....	103
10.2 Methods.....	103
10.3 Results.....	105
Task 11 Time Series and Spatial Survey Datasets Together: Chlorophyll	112
11.1 Scope.....	112
11.2 Methods.....	112
11.3 Results.....	113
Task 12 Long-Term Trends in Temperature, Salinity, and Stratification	139
12.1 Scope.....	139
12.2 Methods.....	139
12.3 Results.....	140
All Tasks: Data and Code Files Documentation.....	150
1. Phillipsdale Landing Time Series Data Reduction	150
2. Development of Nitrogen Load Time Series	151
3. Retrieval and Reduction of Wind, Tidal Range, and Sea Level Data	152
4. Update SNBIW Oxygen Results with Phillipsdale.....	152
5. Inter-Annual Variability of Hypoxia and Stratification.....	153
6. Drivers of Inter-Annual Hypoxia Variability	154
7. Time Series and Spatial Survey Datasets Together: Oxygen.....	154
8. Toward Improved Chlorophyll Time Series Using Grab Samples.....	155
9. Update SNBIW Chlorophyll Results with Phillipsdale	156
10. Refine Chlorophyll Index.....	156
11. Time Series and Spatial Survey Datasets Together: Chlorophyll	156
12. Long-Term Trends in Temperature, Salinity, and Stratification.....	157
References Cited	158
Acknowledgements.....	161

List of Figures

Figure 4-1. Hypoxia Index for individual sites in each group, relative to 2.9 mg L ⁻¹ threshold.	21
Figure 4-2. Hypoxia Index averaged over site groups, relative to threshold 2.9 mg L ⁻¹ day.....	22
Figure 4-3. Same as Figure 4-1 but Hypoxia Index relative to 1.4 mg L ⁻¹ threshold.	23
Figure 4-4. Same as Figure 4-2 but Hypoxia Index relative to 1.4 mg L ⁻¹ threshold.	24
Figure 4-5. Same as Figure 4-1 but Hypoxia Index relative to 4.8 mg L ⁻¹ threshold.	25
Figure 4-6. Same as Figure 4-2 but Hypoxia Index relative to 4.8 mg L ⁻¹ threshold.	26
Figure 5-1. Hypoxia-stratification diagram.	30
Figure 5-2. Same as Figure 5-1 but seasonal hypoxia index is relative to the 1.4 mg L ⁻¹ threshold (more severe hypoxia).	31
Figure 5-3. Same as Figure 5-1 but seasonal hypoxia index is relative to the 4.8 mg L ⁻¹ threshold (milder hypoxia).....	32
Figure 5-4. Same as Figure 5-1 but using the bay-wide group of stations BR-CP-NP-MV-QP-PP-TW-MH.	33
Figure 5-5. Same as Figure 5-1 but using the bay-wide group of stations only for stratification (and the BR-CP-NP-MV group for hypoxia).	34
Figure 6-1. Daily river flows 2001-2017, 10 rivers and ungauged area: basin-wide flow.....	43
Figure 6-2. Median daily basin-wide flow, June to September (black line).....	44
Figure 6-3. Hypoxia (red), river flow (blue), TN load (green), and stratification (black), 2005-2017.	45
Figure 6-4. Property-property plot of results in Figure 6-3.	46
Figure 6-5. Inter-annual variability of temperature and physical drivers.....	47
Figure 7-1. Spatial surveys station locations.	59
Figure 7-2. Rank of mean near-bottom DO concentration at spatial survey stations.....	60
Figure 7-3. Spatial survey correlation results.	61
Figure 7-4. Spatial surveys, decorrelation length scales.....	64
Figure 7-5. Same correlation results as in Figure 7-4 but ordered on the horizontal axis by relative rank (shown in Figure 7-2b) of station near-bottom DO using deeper 57 stations.....	65
Figure 7-6. Decorrelation time scale for deep DO concentration.	66
Figure 7-7. Correlations of spatial survey near-bottom DO concentration and daily mean of 15-minute temporal resolution time series on date of spatial survey.....	67
Figure 7-8. Correlation matrix between spatial surveys and time series oxygen.....	70
Figure 7-9. Lagged correlations between spatial survey and time series oxygen.	71
Figure 7-10. Percent hypoxic area from individual spatial survey dates, DO thresholds 4.8 mg L ⁻¹ (green), 2.9 mg L ⁻¹ (blue), and 1.4 mg L ⁻¹ (red).	72
Figure 7-11. Percent hypoxic area as function of time series oxygen.	73
Figure 7-12. Same as Figure 7-11 but horizontal axis is the daily deficit-duration relative to 1.4 (red), 2.9 (blue), and 4.8 (green) mg L ⁻¹ threshold.	74
Figure 7-13. Seasonal Hypoxia Index from time series vs seasonally-averaged percent hypoxia area from spatial surveys.....	75
Figure 7-14. Histograms to gauge representativeness of spatial survey dates.....	76
Figure 7-15. Histograms to gauge representativeness of time series locations.....	77
Figure 8-1. Phillipsdale. Sonde timeseries with grab sample results superposed.....	83
Figure 8-2. As Figure 8-1, Bullock Reach.	84
Figure 8-3. As Figure 8-1, Conimicut Point.	85
Figure 8-4. As Figure 8-1, North Prudence.....	86
Figure 8-5. As Figure 8-1, Greenwich Bay.	87
Figure 8-6. As Figure 8-1, Sally Rock.	88

Figure 8-7. As Figure 8-1, Mount View.	89
Figure 8-8. As Figure 8-1, Quonset Point.	90
Figure 8-9. As Figure 8-1, Poppasquash Point.	91
Figure 8-10. As Figure 8-1, T-Wharf.	92
Figure 8-11. As Figure 8-1, Mount Hope.	93
Figure 8-12. Phillipsdale. Scatterplots. See text for discussion.	94
Figure 8-13. Scatterplots using 15-min sonde values. Stations BR, CP, NP, GB, SR.	95
Figure 8-14. Scatterplots using 15-min sonde values. Stations MV, QP, PP, TW, MH.	96
Figure 9-1. Chlorophyll Index, individual sites, site-specific 80 th percentile thresholds (Table 9-2).	101
Figure 9-2. Chlorophyll Index, multi-site averages, site-specific 80 th percentile thresholds (Table 9-2).	102
Figure 10-1. Chlorophyll Index relative to (lowest) 4.9 $\mu\text{g L}^{-1}$ threshold.	108
Figure 10-2. Same as Figure 10-1 except (intermediate) 9.4 $\mu\text{g L}^{-1}$ threshold.	109
Figure 10-3. Same as Figure 10-1 except using (highest) 17.6 $\mu\text{g L}^{-1}$ threshold.	110
Figure 10-4. Chlorophyll Index, bay-wide thresholds, all three thresholds.	111
Figure 11-1. Rank of mean near-surface chlorophyll concentration at spatial survey stations.	121
Figure 11-2. Spatial survey correlation results.	122
Figure 11-3. Spatial surveys, decorrelation length scales.	125
Figure 11-4. Same correlation results as in Figure 11-3 but here they are ordered on the horizontal axis by relative rank (shown in Figure 11-1) of station near-surface chlorophyll.	126
Figure 11-5. Decorrelation time scale for surface chlorophyll concentration from lagged autocorrelations of time series measurements at individual fixed site locations.	127
Figure 11-6. Correlations of spatial survey surface chlorophyll concentration and daily mean of 15-minute temporal resolution time series on date of spatial survey.	128
Figure 11-7. Correlation matrix of spatial surveys and time series.	131
Figure 11-8. Lagged correlations between spatial surveys and time series.	132
Figure 11-9 Percent high-chlorophyll area from individual spatial survey dates, thresholds 4.9 $\mu\text{g L}^{-1}$ (green), 8 $\mu\text{g L}^{-1}$ (blue), and 27.5 $\mu\text{g L}^{-1}$ (red).	133
Figure 11-10. Percent high-chlorophyll area of spatial surveys vs time series.	134
Figure 11-11. Same as Figure 11-10 but using daily surplus-duration.	135
Figure 11-12. Seasonal Chlorophyll Index (time series) vs season-mean high-chlorophyll area (spatial surveys).	136
Figure 11-13. Histograms to gauge representativeness of spatial surveys.	137
Figure 11-14. Histograms to gauge representativeness of time series.	138
Figure 12-1. Raw time series, removal of mean seasonal cycle, linear regression.	145
Figure 12-2. Trend results for temperature.	146
Figure 12-3. Trend results for salinity. Presented as in Figure 12-2.	147
Figure 12-4. Trend results for stratification. Presented as in Figure 12-2.	148
Figure 12-5. Schematic summary of trends.	149

List of Tables

Table 1-1. Coverage of good quality data from Phillipsdale.	12
Table 4-1. Percent of time during mid-May to mid-October analysis period with valid near-bottom dissolved oxygen values at each of the eleven sites in each of the seventeen years. Asterisks (*) indicate records with less than 55 percent valid values, which were not included in the analysis. Blanks indicate sites for which data records were not obtained.	20
Table 6-1. Gauges used for river flow observations.	42
Table 6-2. Scale-up factors computed from long-term mean flow, from Table 7 of Ries (1990). The first 8 rows are the scale-up factors from the gauge (see left column of Table 1 above) to the mouth of each river. The last row is the scale-up factor from the sum of the 8 river mouth flows to the basin-wide flow.	42
Table 6-3. Threshold values, for median daily basin-wide flow between June and September based on 1990-2017 conditions, to designate individual years as wet, dry, or intermediate.	42
Table 6-4. Designation of bay-wide river flow conditions each year from 2001 to 2017 as “wet”, “intermediate”, or “dry”, using the criterion in Table 6-3.	42
Table 7-1. Ordering and ranks of spatial survey stations by increasing long-term mean oxygen concentration.	56
Table 7-2. Ordering and ranks of spatial survey stations by increasing long-term mean oxygen concentration, in two groups: deeper/shallower stations, where the median of the deepest-sampled depths is greater than or less than 4m.	57
Table 7-3. Regression r^2 between daily-mean fixed site bottom oxygen concentration and individual survey percent hypoxic area, corresponding to Figure 7-11. Bold values have $p < 0.05$	58
Table 7-4. Regression r^2 between seasonal Hypoxia Index from fixed site time series and seasonally-averaged percent hypoxic area from spatial surveys, corresponding to Figure 7-13. Bold values have $p < 0.05$. Missing values indicate the regression could not be computed because there were no non-zero values of percent hypoxic area.	58
Table 8-1. Basic statistics of grab sample datasets in the provided spreadsheets.	82
Table 9-1. Percent of time during mid-May to mid-October analysis period with valid near-surface chlorophyll values at each of the eleven sites in each of the seventeen years. Asterisks (*) indicate records with less than 55 percent valid values, which were not included in the analysis. Blanks indicate sites for which data records were not obtained.	99
Table 9-2. Chlorophyll thresholds for each site, computed as the 80th percentile of all years’ data between May 15 and October 14 from that site.	100
Table 11-1. Ordering and ranks of spatial survey stations by decreasing long-term mean chlorophyll concentration.	118
Table 11-2. Ordering and ranks of spatial survey stations by decreasing long-term mean chlorophyll concentration, in two groups: deeper/shallower stations, where the median of the deepest-sampled depths is greater than or less than 4m.	119
Table 11-3. Regression r^2 between daily-mean fixed site surface chlorophyll and individual survey percent high-chlorophyll area, corresponding to Figure 11-10. Bold if $p < 0.05$	120
Table 11-4. Regression r^2 between seasonal Chlorophyll Index from fixed site time series and seasonally-averaged percent high-chlorophyll area from spatial surveys, corresponding to Figure 11-12. Bold values have $p < 0.05$. The missing value indicates the regression could not be computed because there were no non-zero values of the index.	120

Acronyms

BR	Bullock Reach site of NBFSMN
CP	Conimicut Point site of NBFSMN
DO	Dissolved Oxygen
GB	Greenwich Bay site of NBFSMN
GRBY	Greenwich Bay group of NBFSMN stations
MH	Mount Hope Bay site of NBFSMN
MV	Mount View site of NBFSMN
NARR	North American Regional Reanalysis
NBFSMN	Narragansett Bay Fixed Site Monitoring Network
NBEP	Narragansett Bay Estuary Program
NBNERR	Narragansett Bay National Estuarine Research Reserve
NEIWPC	New England Interstate Pollution Control Commission
NP	North Prudence Island site of NBFSMN
PD	Phillipsdale Landing site of NBFSMN
PP	Poppasquash Point site of NBFSMN
PRUB	Providence River and Upper Bay group of NBFSMN stations
QP	Quonset Point site of NBFSMN
RIDEM	Rhode Island Department of Environmental Management
SNBIW	State of Narragansett Bay and Its Watershed, published 2017 by NBEP
SR	Sally Rock site of NBFSMN
TW	Prudent Island T-Wharf site of NBFSMN
URI/GSO	University of Rhode Island, Graduate School of Oceanography
UEP	Upper East Passage group of NBFSMN stations
UWP	Upper West Passage group of NBFSMN stations

Task 1 Phillipsdale Landing Time Series Data Reduction

1.1 Scope

Objectives. Obtain time series data from Phillipsdale Landing (Seekonk River); apply the same reduction steps as for other Narragansett Bay Fixed Site Monitoring Network (NBFSMN) stations; append results to dataset. **Methodology.** Time series (15-minute resolution, all years through 2017) will be obtained from Heather Stoffel (URI/GSO) and transferred to the standardized format (outliers removed, gaps flagged, density computed from temperature and salinity) of other stations. **Expected outputs and outcomes.** These reduced data will facilitate completion, for Phillipsdale, of the calculations, figures, and interpretation done in the SNBIW using data from other stations—and of those to be done for new tasks below. As a shallow site near the northern limit of the estuary, Phillipsdale is severely eutrophic and therefore an important gauge of bay response to nutrient load reductions.

1.2 Methods

The “corrected” files (distinct from raw and edited files, as explained at <http://www.dem.ri.gov/programs/emergencyresponse/bart/netdata.php>) for 15-minute resolution data from near-surface and near-bottom sensors at the Phillipsdale Landing site were obtained from Heather Stoffel (URI/GSO) for the years 2005 through 2017. The corrected files are the result of QA/QC treatments applied to the edited files by Stoffel, per the project Quality Assurance Project Plan (RIDEM, 2014). They contain various auxiliary data in multiple sheets of the spreadsheet. The subset of the spreadsheet with the corrected Phillipsdale data was extracted and saved in a “modified corrected” file, where minor additional changes were made if needed for alignment. Then Matlab scripts, that were applied to other stations for the SNBIW, were run on the files. These scripts read the raw values for date/time, temperature, salinity, oxygen, and chlorophyll, compute density from temperature and salinity, compute stratification from density at near-surface and near-bottom depths, and align all the resulting variables on a common gap-free date/time grid.

See page 150 for documentation of the supporting data and code files.

1.3 Results

The data coverage for near-bottom oxygen and near-surface chlorophyll from May 15 to October 14 in each year is shown in Table 1-1. These results correspond to results for all the other stations, as shown in Table 2 of SNBIW Chapter 15 on oxygen (updated in Task 4 below, see Table 4-1), and Table 2 of SNBIW Chapter 16 on chlorophyll (updated in Task 9 below, see Table 9-1).

The resulting data files were used in numerous tasks below.

Table 1-1. Coverage of good quality data from Phillipsdale.

	Near-bottom Oxygen	Near-surface chlorophyll
2005	100	100
2006	97	99
2007	100	94
2008	100	99
2009	91	81
2010	70	55
2011	100	100
2012	100	89
2013	86	86
2014	100	100
2015	100	100
2016	100	98
2017	100	94

Task 2 Development of Nitrogen Load Time Series

2.1 Scope

Objective. Develop nutrient load time series for the later analysis tasks. **Methodology.** All available data for nutrient concentrations and flow from wastewater treatment plant facilities (WWTFs) will be obtained from Heidi Travers (RI DEM). Data from results in SNBIW Chapter 8, for both treatment plants and rivers, will also be obtained from NBEP. Temporal resolution is generally monthly at best, so the main effort will be creating estimated time series at higher frequency (weekly or daily), using higher-frequency river flow data and relationships between it and less-frequent WWTF data. Methods will follow prior work, and guidance from C. Oviatt (URI/GSO), J. Krumholz (McLaughlin), M. Brush (VIMS) and J. Vaudrey (UConn) will be sought, as they have calculated similar load datasets. **Expected outputs and outcomes.** Reduced WWTF nutrient loads are a major influence on bay ecology. However, as noted in the SNBIW (NBEP, 2017), underlying processes are not fully understood. The loads data this task generates will play a crucial role in key analysis tasks, described below, designed to help improve knowledge of these processes.

2.2 Methods

This task became sufficiently involved that it made sense to document its methods and results in a stand-alone companion report, “Daily-Resolution 2001-2017 Time Series of Total Nitrogen Load to Narragansett Bay from Bay-Wide Treatment Facility and Watershed Sources” (Codiga, 2020).

See page 151 for information about supporting code and data files, and page 150 for how to use them.

2.3 Results

The abstract to the companion report (Codiga, 2020) is as follows:

Time series for the load of total nitrogen (TN) to Narragansett Bay, from 18 wastewater treatment facility (WWTF) and land-based runoff sources bay-wide, have been estimated at daily resolution for 2001-2017. The motivation for daily resolution is to enable including nitrogen load as a candidate influence (with others such as river flow, temperature, tidal conditions, etc; e.g., Codiga et al 2009) in statistical analyses investigating drivers of hypoxic events at the short timescales of days to weeks on which events are known to vary. Time series were estimated back to 2001 because such analyses rely on Narragansett Bay Fixed Site Monitoring Network time series oxygen observations, which began that year. The WWTF and riverine sources treated are generally the same as those in the annual-budget analyses in the State of Narragansett Bay and Its Watershed (NBEP 2017). The 18 sources include 11 WWTFs (nine in Rhode Island and two in Massachusetts) discharging directly to the bay; 6 rivers where they enter the bay, which include load from WWTFs located upstream on them; and runoff direct to the bay from ungauged riparian areas. The observations were obtained from the Rhode Island Department of Environmental Management, the Narragansett Bay Commission, the Fall River treatment facility, and the US Geological Survey. Load was computed as the product of concentration and flow. For concentration, linear interpolation was necessary as the nominal frequency of observations ranged from weekly or biweekly, for most of the largest sources, to monthly. However, temporal variations in load are dominated by flow variability, which spans multiple orders of magnitude while concentration variations are less pronounced, and daily flow measurements were available for many of the largest

sources. During periods when TN was not directly measured, it was computed as the sum of other measured concentrations (e.g., total Kjeldahl nitrogen plus nitrate plus nitrite), or using a correlation between TN and other constituents during periods when both were measured, or as the long-term mean seasonal cycle of TN computed from sampled years. As a check on the reasonableness of the results, the annual-mean loads from the new time series, computed by averaging the daily values, were compared to results of earlier budgets reported in NBEP (2017) for three periods: 2000-2004 (Nixon et al 2008), 2007-2010 (Krumholz 2012), and 2013-2015 (NBEP 2017). The differences are notable, as expected given the divergent methods, but the agreement is acceptable for the intended use of the new daily time series; for ascertaining long-term changes in loading, the NBEP results are more appropriate. It is recognized that the estimated daily TN loads are approximate, particularly for the several earliest years when fewer concentration measurements were made. A number of suggestions for how to improve the methods are given. The new time series were used to investigate hypoxia in a companion report (Codiga 2020) and are expected to be of use for various other studies. They are available for download in spreadsheet (.xlsx) and Matlab (.mat) file formats, with documented supporting code, at <https://figshare.com/s/7d51f2540df6638a4552>.

Task 3 Retrieval and Reduction of Wind, Tidal Range, and Sea Level Data

3.1 Scope

Objective. Retrieve wind, tidal range, and sea level data up through 2017 and append it to existing records, making possible the “driving factors” analysis tasks described below. **Methodology.** Methods as detailed in Codiga et al 2009 will be followed. Winds are from the North American Regional Reanalysis; sea level is from the Providence and Newport stations, partitioned in to tidal and non-tidal components; and tidal range is based on the Newport station. **Expected outputs and outcomes.** These environmental data support tasks below investigating influences of driving factors on bay conditions.

3.2 Methods

Winds. The 10-m elevation wind from the North American Regional Reanalysis (“NARR”; Mesinger, 2006), a meteorological model product that assimilates observed atmospheric conditions, is used because it has no missing-data gaps, in contrast to local wind gauges. Its grid has ~32 km resolution, coarse relative to the bay, so it does not capture details of small-scale local wind patterns. This trade-off is considered acceptable to gain the benefits of gap-free records for the entire period. The years 1999 through 2017 from the over-land NARR gridpoint nearest to Narragansett Bay, just north of Greenwich Bay but representative of region, are used. They are instantaneous winds, with a value each 3 hours. They were retrieved remotely from the NOAA server where very large netcdf files, containing data from all gridpoints are provided; the small desired subset (from one grid point) of these files was extracted using built-in Matlab functionality for data access protocols.

Tidal Range. The tidal range data is generated using tidal predictions. The predictions are available at the NOAA CO-OPS website accessible via web services. The hourly predictions for the Newport station (<https://tidesandcurrents.noaa.gov/noaatidepredictions.html?id=8452660>) are obtained and used to compute the difference between successive high and low tides. The daily tidal range is computed as the mean of all such differences that day. The tidal range cubed, a measure of mixing energy, is also computed. Values are interpolated to a 12 hour grid, to match the low-passed winds and sea levels.

Sea Level. Sea level observations which include both the tidal component and the non-tidal components (i.e. weather-band variations) were obtained. Raw hourly observations were obtained, along with atmospheric pressure data in order to do the inverse barometer correction, from NOAA servers via OpenDap, for both the Providence (<https://tidesandcurrents.noaa.gov/stationhome.html?id=8454000>) and Newport (<https://tidesandcurrents.noaa.gov/stationhome.html?id=8452660>) locations. The atmospheric pressure is used to correct the sea level for the inverse barometer effect. The hourly water level at each station, the inverse-barometer adjusted hourly water level at each station, the difference between Providence and Newport hourly water level, and the difference between Providence and Newport hourly adjusted water level are computed. The 12-hour subsampled low-pass filter (25-hr halfwidth triangle-weight) of all the same quantities is computed as well.

See page 152 for information about supporting code and data files, and page 150 for how to use them.

3.3 Results

The products generated are used as indicators of physical drivers in calculations below.

Task 4 Update SNBIW Oxygen Results with Phillipsdale

4.1 Scope

Methodology. Same as in SBNIW. **Expected outputs and outcomes.** Updated versions of figures and tables presenting fixed-site results in the SNBIW oxygen chapter, now including Phillipsdale; and an explanation of Phillipsdale results in the context of other stations, paralleling the SNBIW narrative.

4.2 Methods

The methodology is the same as described in the SBNIW Chapter 15. Regarding the data from the Bullocks Reach site during 2017, these data are included in the analysis, but Heather Stoffel indicated several serious issues had to be addressed when creating the “corrected” files during the quality control and quality assurance process. There was a higher occurrence of outliers, and many more periods with differing offsets, particularly in the salinity and temperature records, than for measurements from more typical years.

See page 153 for information about supporting code and data files, and page 150 for how to use them.

4.3 Results

Table 4-1 below is an updated version of Table 2 from SNBIW Chapter 15 showing percent time of valid data between mid-May and mid-October. It is updated to include Phillipsdale station (new first column) and to include years 2016 and 2017 (all stations) as two new rows at bottom. The first year of corrected data files for Phillipsdale obtained is 2005. The percent time of valid data at Phillipsdale is generally very high, at or near 100%, with the exception of 70% in 2010 and 86% in 2013. In 2016 and 2017 the percentages at most stations were generally high, with nearly all at 87% or higher, and the lowest at 77%. For most stations in 2016 the max was 87% due to the deployment start in early June. For some stations in 2017 the max was 94% due to the deployment start in late May.

For purposes of the analysis of groups of multiple sites together, Phillipsdale (PD) is the only station in its group. (This is similar to the Mount Hope Bay station, which is the only station in its group.) This is justified because its location is substantially farther up-estuary than all other sites, and is shallower than nearly all other sites. Thus the station groups in Table 1 of SNBIW Chapter 15 (PRUB: Providence River and Upper Bay; UWP: Upper West Passage; UEP: Upper East Passage; GRBY: Greenwich Bay) are unchanged, except to add Phillipsdale (PD) as a new group, containing only the PD site.

The river flow analysis described in SNBIW designated five years (2003, 2006, 2009, 2011, and 2013) as “wet”, meaning precipitation and river runoff that year were higher than normal. An updated method for designating the river flow during a given year as wet, intermediate, or dry has been developed in Task 6 below. It identifies those same 5 years as wet, 7 years as “intermediate”, and 5 years as “dry”. These “wet”, “intermediate”, and “dry” designations are indicated with background bars in three shades of increasingly lighter gray in the figures below. These designations differ from the figures in SNBIW, for which only the years designated “wet” (not “normal” nor “dry”) were shown as gray bars.

The following description is structured closely following that in SNBIW Chapter 15, updated to include results for Phillipsdale. As in the SNBIW first focuses on 2013-2015 results but results for 2016 and 2017 are now also included. It also presents and discusses the results for the 1.4 mg L⁻¹ and 4.8 mg L⁻¹ thresholds (though not included in the SNBIW, these were provided to NBEP as the SNBIW was being developed).

The Hypoxia Index is the season-cumulative (mid-May to mid-Oct) deficit-duration, or sum of the areas swept out beneath the threshold by the oxygen concentration curve during all sub-threshold events that season. Its units are mg L⁻¹ day. Its meaning and interpretation are explained in SNBIW Chapter 15, in particular its pages 279-281 and its Figure 2.

Figure 4-1 (showing individual station results) and Figure 4-2 (showing multi-site mean results) below are updated versions of Figures 4 and 5 from SNBIW Chapter 15. They present results for the Hypoxia Index relative to the 2.9 mg L⁻¹ threshold. They have been updated to include Phillipsdale station, and to include the years 2016 and 2017.

Figure 4-3 and Figure 4-4 below are in the same format as Figure 4-1 and Figure 4-2, but show results for Hypoxia Index relative to 1.4 mg L⁻¹.

Figure 4-5 and Figure 4-6 below are in the same format as Figure 4-1 and Figure 4-2, but show results for Hypoxia Index relative to 4.8 mg L⁻¹.

The individual site results from 2013 to 2015 reveal two dominant aspects that are generally applicable to other years. The first aspect is pronounced inter-annual variability. The Hypoxia Index was higher in 2013 than in 2014 or 2015 at all sites, and substantially so at most sites (Figure 4-1) indicating lower overall dissolved oxygen concentrations in 2013 than in 2014 or 2015. This inter-annual variability is notably far larger than uncertainties in the Index values due to sampling gaps. The higher 2013 Index values over those of 2014 and 2015 were generally similar at all sites, indicating they result from a process with bay-wide influence. Precipitation is a major factor responsible for this variability, as hypoxia was more severe in 2013, which had wet river runoff conditions, compared to the dry years of 2014 and 2015. For one station, Phillipsdale (PD), there are several exceptions to this pattern. At PD, during wet years the Hypoxia Index was generally elevated, but many of the highest values occurred during normal or dry years, for example 2008, 2012, and 2015-2016.

The second dominant characteristic was a down-Bay gradient, in which the Hypoxia Index generally decreased from north to south (Figure 4-1 and Figure 4-2). This was a persistent feature, with very few exceptions in any given year, and the gradient was stronger in wet years. The Hypoxia Index decreased from north to south from Phillipsdale, through the Providence River and Upper Bay (PRUB) sites, and down through the Upper West Passage (UWP) (Figures 1 and 2). Bottom water dissolved oxygen increased with distance southward.

Using 2.9 mg L⁻¹ as the dissolved oxygen threshold, the Hypoxia Index (season-cumulative deficit-durations) were typically less than 30 mg L⁻¹ day with a maximum of about 65 mg L⁻¹ day (Figure 4-1). During the dry years of 2014 and 2015, the Index was effectively zero everywhere except at PD and Greenwich Bay (GB). During the wet year of 2013, the Index at PD was comparable to other years, but it was elevated for GB, PRUB, Mount Hope Bay (MH), and the Mount View site (MV) in UWP, while the Index at other sites in the Bay was zero or very near to zero (oxygen concentrations remained above the

2.9 mg L⁻¹ threshold). During 2017, in contrast to prior years with comparable river flow (see Figure 6-2), hypoxia was nearly zero at all stations except for PD and GB, and the level of hypoxia at GB was substantially lower. The 2017 results suggest that, with the exception of those two sites which are shallower and farther inland than the others, the nutrient load reductions to the bay due to treatment facility upgrades have resulted in reduced hypoxia severity even during years with intermediate river flow. Whether this is the case for wet conditions remains to be seen.

The main results for Hypoxia Index relative to the 1.4 mg L⁻¹ threshold (Figure 4-3 and Figure 4-4) and the 4.8 mg L⁻¹ threshold (Figure 4-5 and Figure 4-6) are similar, in most respects, to those just described for the 2.9 mg L⁻¹ threshold.

The Hypoxia Index for the 1.4 mg L⁻¹ threshold (Figure 4-3 and Figure 4-4) reflects only hypoxia events that reach very low oxygen levels. These do not occur at eastern stations (PP, TW, MH) or the southernmost station in the west (QP). At mid-bay stations these events occur rarely (MV, in one year, 2003) or only reach small magnitudes of up to a few mg l⁻¹ day (BR, CP, NP). The stations where the index has reached higher values on multiple occasions include PD, GB, and SR. The maximum Hypoxia Index value was about 18 mg l⁻¹ day, at GB in 2003. At SR, and to a large extent GB, higher values of the Index are generally associated with wet years; at PD they occurred in some wet years and some normal and dry years. Since 2013, even at PD and GB, there have been no events, even in 2017 during intermediate flow conditions. This suggests the severity of the most intense bay hypoxia (i.e. relative to the lowest threshold of 1.4 mg L⁻¹) has been reduced in association with the treatment plant upgrades.

The Hypoxia Index for the 4.8 mg L⁻¹ threshold (Figure 4-5 and Figure 4-6) includes the influence of mildly depressed oxygen levels. Such conditions have occurred commonly at all stations except TW, where they are rare. The patterns of inter-annual variability and down-bay gradient are generally similar to those described above for the 2.9 mg L⁻¹ threshold (Figure 4-1 and Figure 4-2). The highest values have occurred at PD, where the Index is typically between 100 and 200 mg l⁻¹ day, and the maximum of about 220 mg l⁻¹ day occurred in 2008. Values nearly as high as those have occurred commonly at BR, GB, and SR, with lesser values seen at CP, NP, MV, and the lowest (other than TW) at PP and MH. Hypoxia relative to the 4.8 mg L⁻¹ threshold since 2013 shows evidence of reductions at all stations, including the 2017 with intermediate flow, except for PD. At other stations, even in 2017 hypoxia severity was strongly reduced, except at GB where it was more modestly weaker that year.

The nature of inter-annual variability in hypoxia at PD clearly differs from that at nearly all other stations. At PD, elevated hypoxia can occur during wet, normal, or dry years, and the most severe hypoxia occurs in dry years more often than in wet years. This may be associated with the PD site being shallow, located farthest up-estuary nearest to major nutrient loading sources, and with circulation conditions that are potentially different from other sites. It is possible that during dry years the PD site flushes more poorly, has longer water residence times, and is thus able to support more phytoplankton growth and decay, leading to enhanced hypoxia compared to other sites. The GB site has some similar characteristics, with high Index in some dry years, suggesting similar processes may be responsible.

Trends extending over multiple years have been difficult to discern due to the relatively large inter-annual variability. Wet years (2003, 2006, 2009, 2011, and 2013) appeared as local maxima in the Hypoxia Index, generally at all sites, with the notable exception of PD as just discussed. Here 2014 is treated as the first “post-reduction” year. Hypoxia in 2013 was comparable to earlier wet years. Since 2013 there have been 3 dry years, during which the Hypoxia Indices relative to all three thresholds have been notably lower than in the past, except at PD where they are also lower but more modestly than at

the other stations for the 2.9 mg L^{-1} and 4.8 mg L^{-1} thresholds. Conditions in 2017, a year with intermediate flow, are the first indication that bay hypoxia is responding positively to the nutrient load reductions. Hypoxia in 2017 was markedly less widespread and less severe than in earlier years with comparable river flow. While at PD there is little indication of improvements, and at GB the changes are modest, at other stations 2017 was the first year on record with intermediate flow for which there was little to no hypoxia. Continued monitoring and analysis of future years with intermediate and wetter than normal flow will be necessary to determine whether these apparent improvements are sustained.

Table 4-1. Percent of time during mid-May to mid-October analysis period with valid near-bottom dissolved oxygen values at each of the eleven sites in each of the seventeen years. Asterisks (*) indicate records with less than 55 percent valid values, which were not included in the analysis. Blanks indicate sites for which data records were not obtained.

	PD	BR	CP	NP	MV	QP	PP	TW	GB	SR	MH
2001		100		62							
2002		75		81							
2003		79	65	73				80	63		
2004		86		82	65		64	88	99		
2005	100	89	76	91	66	*	*	90	83		60
2006	97	79	79	72	94	78	98	89	98		98
2007	100	66	94	78	93	55	93	92	93		93
2008	100	96	96	89	93	97	92	86	88	96	83
2009	91	100	100	80	83	80	87	88	96	81	87
2010	70	85	77	78	82	88	82	97	96	83	84
2011	100	85	90	90	91	85	83	72	100	91	73
2012	100	98	87	96	96	74	96	100	99	94	96
2013	86	99	91	98	80	90	92	98	92	93	100
2014	100	94	96	93	92	96	85	100	100	96	96
2015	100	77	81	86	91	82	91	100	100	79	86
2016	100	79	87	77	87	87	87	100	90	94	87
2017	100	90	82	94	94	89	94	100	97	95	92

Figure 4-1. Hypoxia Index for individual sites in each group, relative to 2.9 mg L⁻¹ threshold.

Years with bay-wide river flow during May to September designated as wet, intermediate, and dry are marked by gray vertical bars of three increasingly lighter shades, respectively. To improve clarity by reducing overlap of lines and symbols, symbols from each site are systematically offset a small distance horizontally relative to those of other sites in the frame. Error bars indicate uncertainties due to gaps in the time series because of sensor malfunction or data not meeting quality standards, and in most cases are smaller than the symbols.

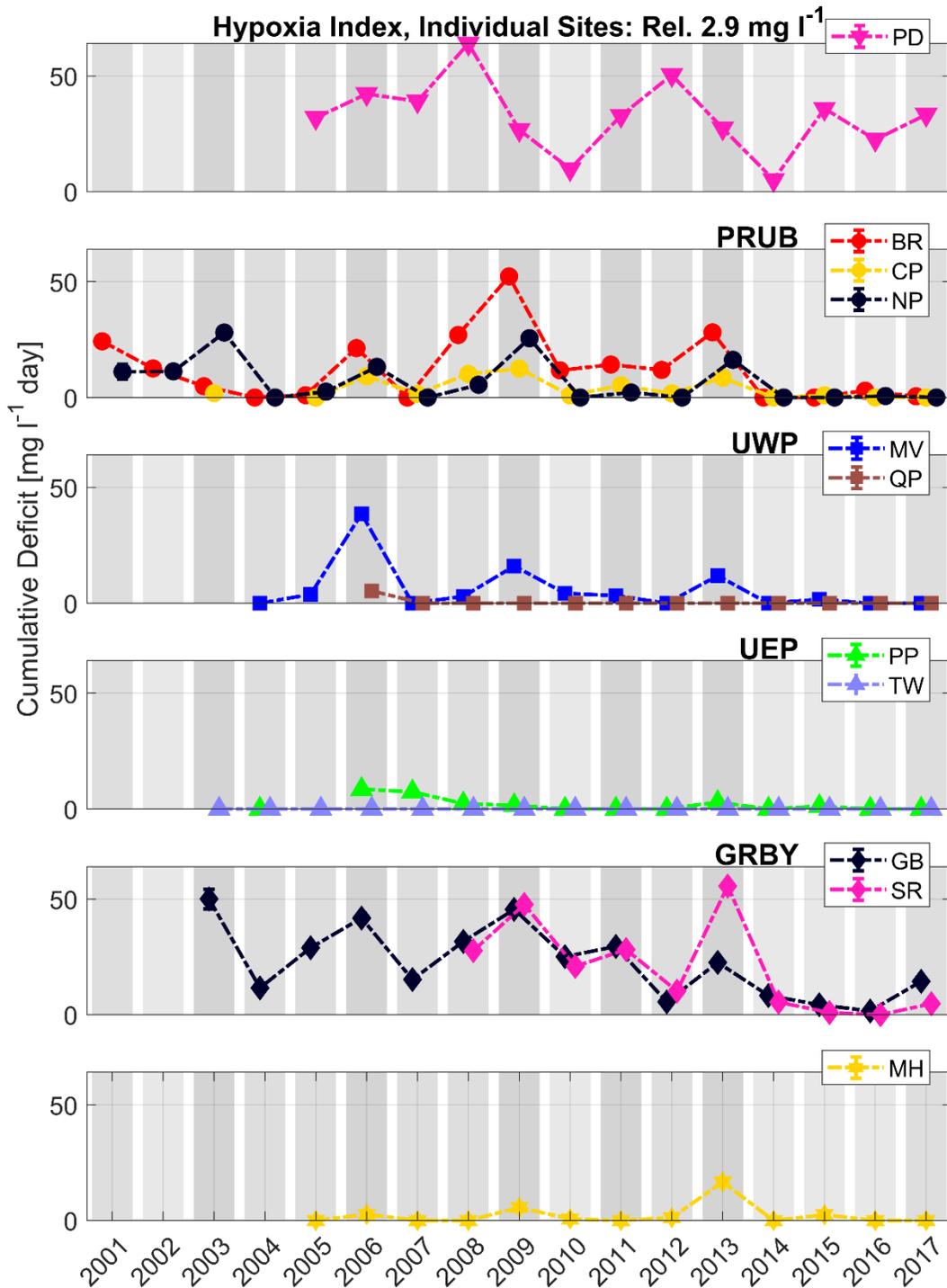


Figure 4-2. Hypoxia Index averaged over site groups, relative to threshold 2.9 mg L⁻¹ day.

Error bars indicate the mean uncertainty, among the stations in the group, due to gaps in the time series.

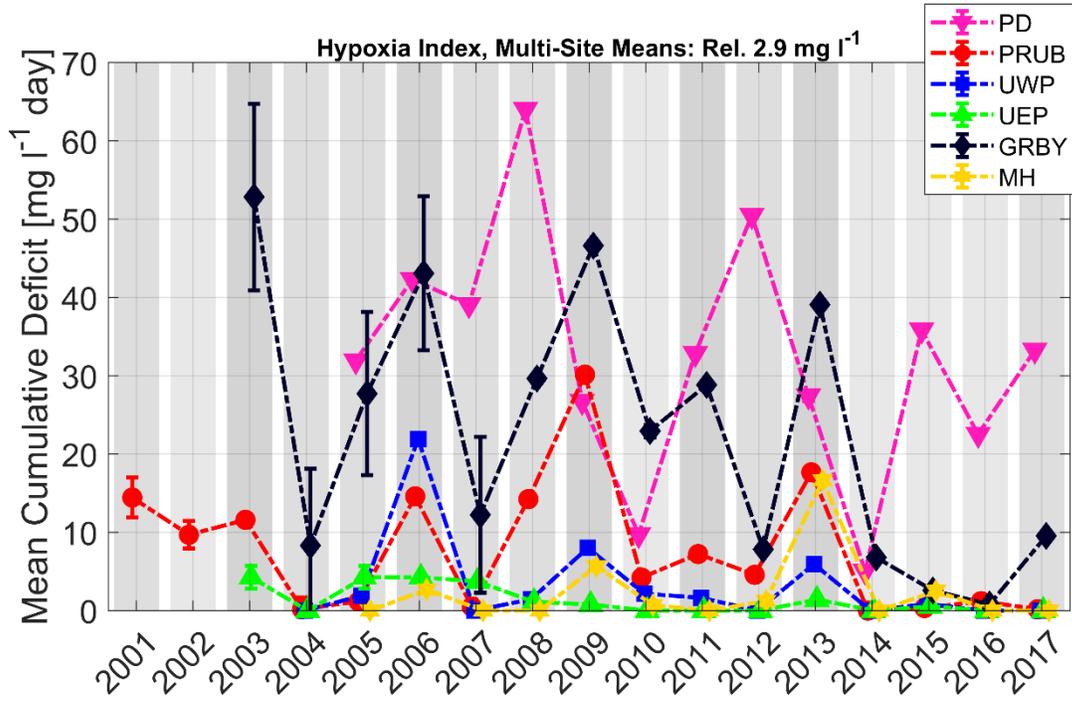


Figure 4-3. Same as Figure 4-1 but Hypoxia Index relative to 1.4 mg L⁻¹ threshold.

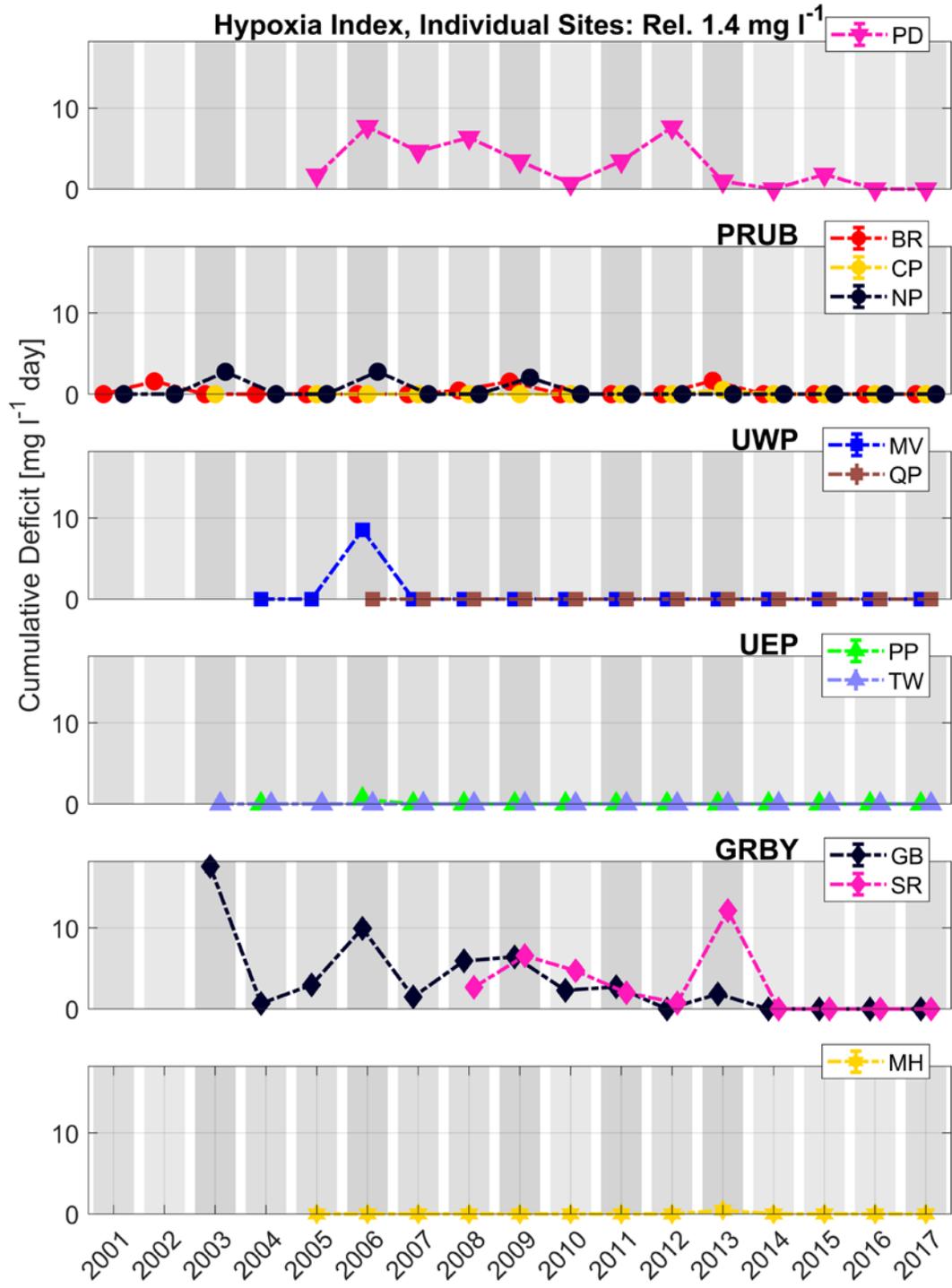


Figure 4-4. Same as Figure 4-2 but Hypoxia Index relative to 1.4 mg L⁻¹ threshold.

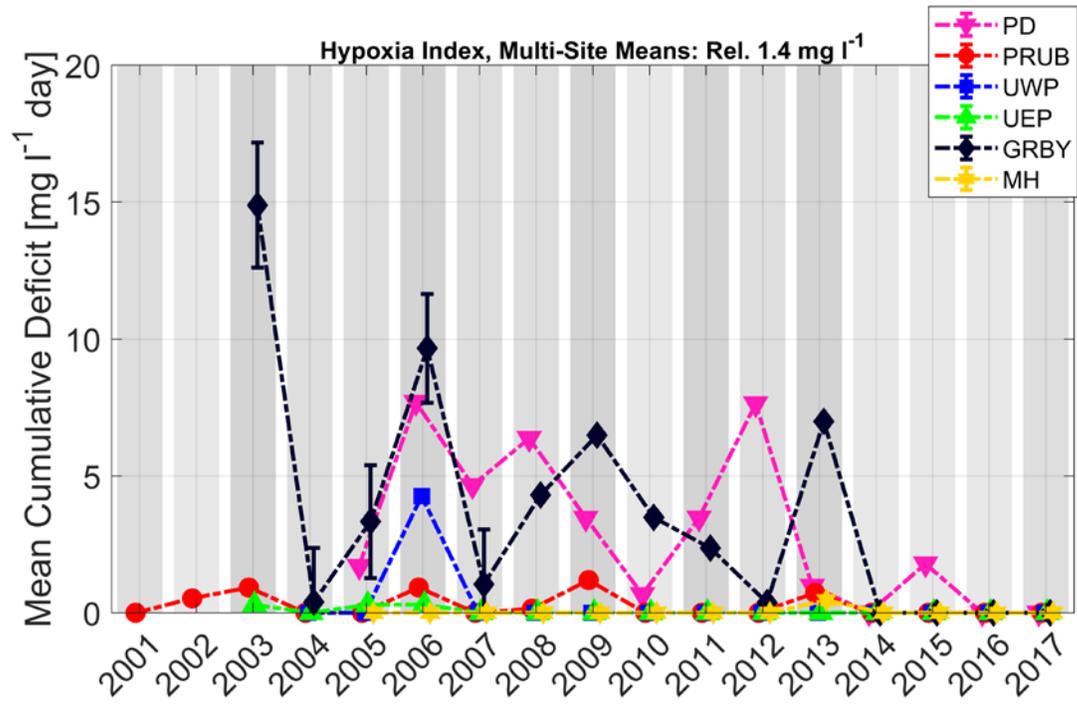


Figure 4-5. Same as Figure 4-1 but Hypoxia Index relative to 4.8 mg L⁻¹ threshold.

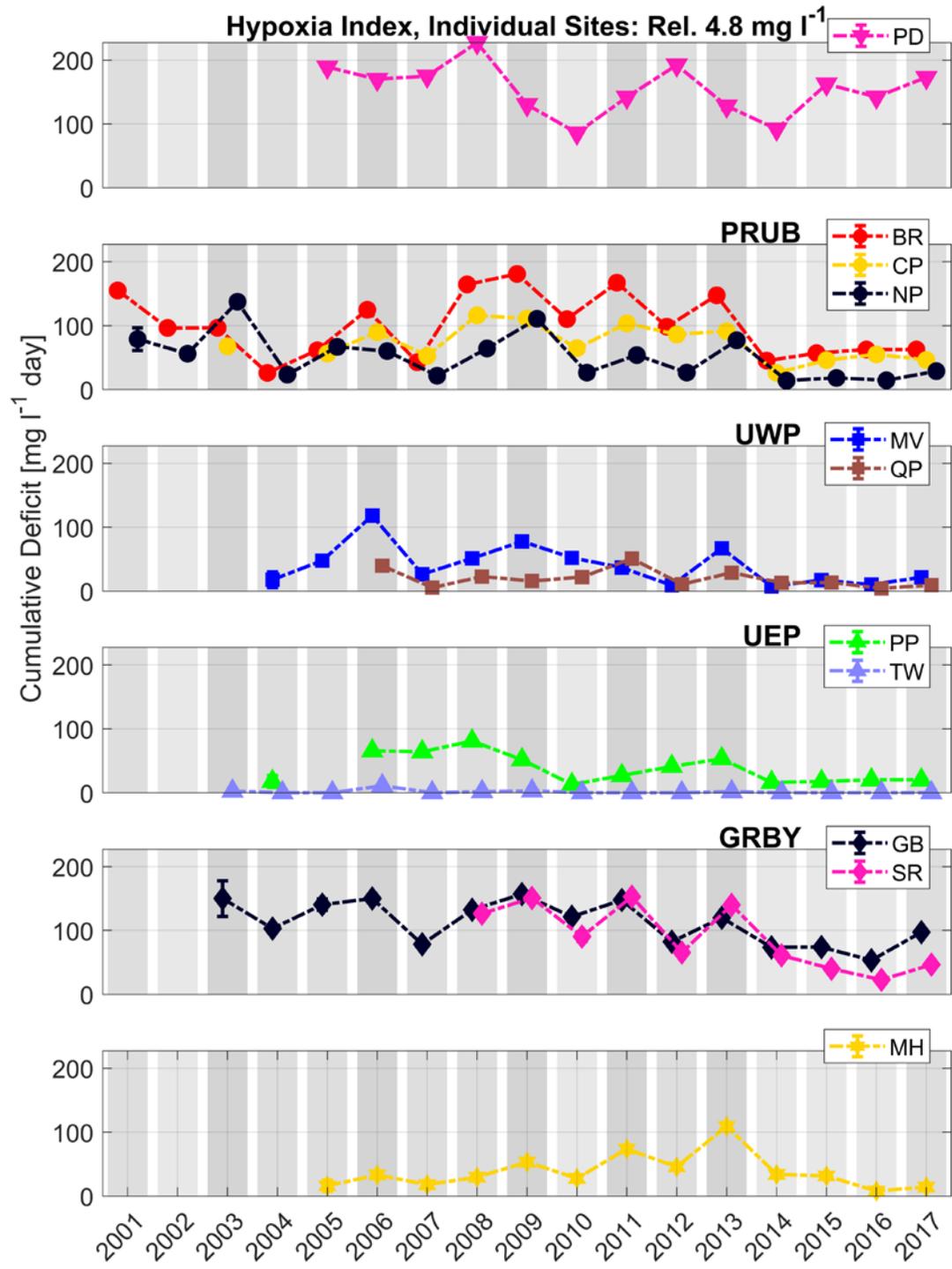
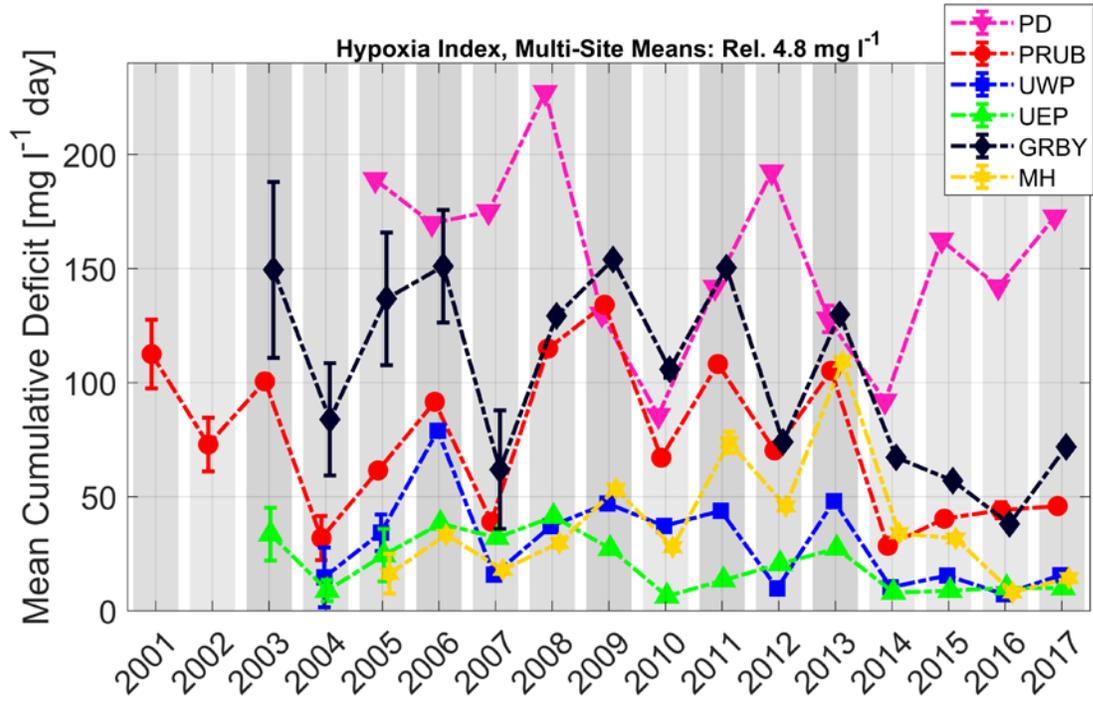


Figure 4-6. Same as Figure 4-2 but Hypoxia Index relative to 4.8 mg L⁻¹ threshold.



Task 5 Inter-Annual Variability of Hypoxia and Stratification

5.1 Scope

Methodology. Methods of Codiga (2012) to explore inter-annual variations in hypoxia and density stratification will be expanded to: include Phillipsdale; include all years through 2017; and examine new combinations of NBFSMN stations. **Expected outputs and outcomes.** Codiga (2012) used data up through 2009 to show that inter-annual variations in stratification and hypoxia were correlated, but pointed out limits to the relationship; the analysis will be updated. Emphasis will be on comparisons between hypoxia severity in years after nutrient load reductions were completed, for a given stratification, and hypoxia severity during earlier years that had comparable stratification. In this way the impacts of load reductions can be better isolated, independent of the influence of stratification.

5.2 Methods

The near-surface and near-bottom 15-minute resolution time series for temperature and salinity from each station were used to compute density using the seawater equation of state. Then the stratification was computed (also 15 minute resolution) as the difference between the near-bottom density and the near-surface density. Diagnostic plots of density and stratification were inspected and revealed that for all records other than Phillipsdale, over nearly all of the record duration the near-bottom density was higher than the near-surface density, as expected, indicating good data quality. In some records there were short periods of outliers but they were infrequent and only present for very short durations.

In contrast, in the Phillipsdale record there was noticeable drift in temperature and/or salinity during some parts of the record, leading to unacceptable drift in the densities, and therefore stratification values that were negative. This is an unphysical result indicating poor data quality for density stratification. On consultation with the data originators, in this analysis of stratification, Phillipsdale was therefore excluded. Stratification, as the difference between densities at two depths, is more sensitive to sensor drift than temperature, salinity, density or oxygen. Therefore it is consistent to exclude Phillipsdale from the analysis of stratification, while including it in other analyses (for example hypoxia).

See page 154 for information about supporting code and data files, and page 150 for how to use them.

5.3 Results

For context, some of the key findings of Codiga (2012) are summarized first: At individual stations, the relationship between stratification and the seasonal hypoxia index is weak. This is because hypoxic events are intermittent and stratification varies strongly on timescales of days to weeks, but the starting and ending times of hypoxic events are not closely synchronous with increases or decreases in stratification. When using averages across multiple stations, this variability is less extreme and a pattern emerges, though relatively weak, in which more severe hypoxia corresponds to stronger stratification. This relationship is strongest when the stratification used is the average over the late spring period only, as opposed to when averaging across the entire season.

The present analysis supports the earlier conclusions. In Codiga (2012) the results of the strongest hypoxia-stratification relationship were presented. They were based on averaging across the 4 stations

where hypoxia has been most common in the deeper parts of the bay from year to year, namely BR-CP-NP-MV, and on averaging the stratification during the Jun 15 to Jul 15 period. So here we start with results for averaging (both hypoxia and stratification) over the same group of stations, using the seasonal hypoxia index based on the same 2.9 mg L^{-1} threshold (Figure 5-1).

To reveal the nature of how the hypoxia-stratification relationship depends on the time period over which stratification is averaged, Figure 5-1 superposes results for four periods: the first one month sampled (5/15-6/15), two first two months (5/15-7/15), the first three months (5/15-8/15) and all five months (5/15-10/15). In addition, results for 2001-2013 in Figure 5-1 (open symbols) are distinguishable from those for 2014-2017 (filled symbols), the latter period being “post-reduction” years.

A general tendency for higher seasonal hypoxia index (vertical axis) corresponding to higher stratification (horizontal axis) is clear in Figure 5-1, albeit with a high degree of scatter. For a given year the stratification result varies substantially depending on the period used for averaging (compare the black, blue, green, and red symbols from individual years). In many years, the typical timing of peak stratification is late spring (Codiga 2012), so it is increasingly weaker for the one-month to five-month averaging periods used (that is, the black, blue, green, and red symbols tend to appear in order from right to left; for example, the year 2001). In other years this is not the case, for example 2009, when the timing of peak stratification was later than in a typical year.

The post-reduction years (2014 onward) include a few weakly stratified years, with 2015 and 2016 having the weakest measured stratification to date. They also include one year with moderate stratification (2017), and no years with strong stratification. While this lack of a post-reduction year with stratification comparably as strong as the most-stratified earlier years makes it difficult to identify a change in the hypoxia-stratification relationship since reductions have occurred, it is nonetheless clear (Figure 5-1) that in nearly all recent years with both weak (2014, 2015, 2016) and moderate (2017) stratification, the seasonal hypoxia index has been as low as or markedly lower than in earlier years with stratification in the same range. Note that in 2013, most load reductions were complete, moderate stratification occurred, and hypoxia was comparable to earlier years with similar stratification.

The nature of these patterns generally also applies to results for using the lower (1.4 mg L^{-1}) and higher (4.8 mg L^{-1}) thresholds when computing the seasonal hypoxia index. For the most severe hypoxia, threshold of 1.4 mg L^{-1} (Figure 5-2), for weak and moderate stratification almost all years have very low seasonal hypoxia index. The sensitivity to the period over which stratification is averaged (different colors/symbols for give years), and the nature of conditions before (open symbols) and after (filled symbols) load reductions, are similar to those described for the 2.9 mg L^{-1} threshold (Figure 5-1). For milder hypoxia, threshold of 4.8 mg L^{-1} (Figure 5-3), as expected the hypoxia index is non-zero for any level of stratification. This means the distribution of points across the plot is slightly more suggestive of a hypoxia-stratification relationship, though scatter is still substantial. Sensitivity to stratification averaging period, and differences between pre/post load reduction results, for the mild hypoxia threshold results (Figure 5-3) are similar to the other thresholds (Figure 5-1 and Figure 5-2).

As noted above, the hypoxia-stratification relationship is weaker for individual stations than for groups of stations. So the result for using all stations bay-wide, excluding shallow embayments, is of interest (Figure 5-4). This bay-wide case uses stations BR-CP-NP-MV-QP-PP-TW-MH. It includes the Mount Hope Bay station, a deeper site where hypoxic events occur similarly to the other stations in the group, and excludes the Phillipsdale, Greenwich Bay, and Sally Rock stations where hypoxia is generally much more persistent and severe. Comparing this bay-wide result (Figure 5-4) to the results using the four stations

BR-CP-NP-MV (Figure 5-1) reveals that the main difference is that hypoxia index values are all in a lower range. This is expected, as hypoxia is typically milder or rare at some of the further down-bay stations added. Otherwise, the details of the relationship between hypoxia and stratification are altered in relatively minor ways, but the general patterns are the same and support the same conclusions.

The final example uses the bay-wide set of stations for the stratification calculation and the four relatively deeper stations where hypoxia is more frequent (BR-CP-NP-MV) for the seasonal hypoxia index calculation (Figure 5-5). This could be expected to tighten up the hypoxia-stratification relationship, to the extent such a relationship applies to stratification computed using the largest spatial scales and hypoxia computed using the stations where it is most common. And in fact the pattern is slightly more clear than in the prior calcs (compare to Figure 5-1 to Figure 5-4). However the relationship is still weak, in the sense that there can be dramatically different hypoxia index results in years which have only modestly differing stratification.

In summary, the earlier findings of Codiga (2012) are supported by these new calculations, which include all 17 years, and have explored some new combinations of stations and time periods over which to average the results. In addition, a comparison of years 2001-2013 to years after load reductions were complete has been made, to help assess potential long-term trends. In the context of stratification during 2001-2013, stratification in post-reduction years has been either weak (2014, 2015, 2016) or moderate (2017). Hypoxia during the recent weakly-stratified years has been very low, comparable to earlier weakly stratified years. Hypoxia during the moderately-stratified 2017 was markedly weaker than in most earlier moderately-stratified years, which as noted in Task 4 is an indication that load reductions are causing the intensity of hypoxia to lessen.

Figure 5-1. Hypoxia-stratification diagram.

Results for both hypoxia and stratification are averages over the deeper stations where hypoxia is most common (BR-CP-NP-MV). The seasonal hypoxia index is relative to the 2.9 mg L⁻¹ threshold. The numbers labeling each symbol correspond to the years 2001 (“01”) through 2017 (“17”). See text for further explanation.

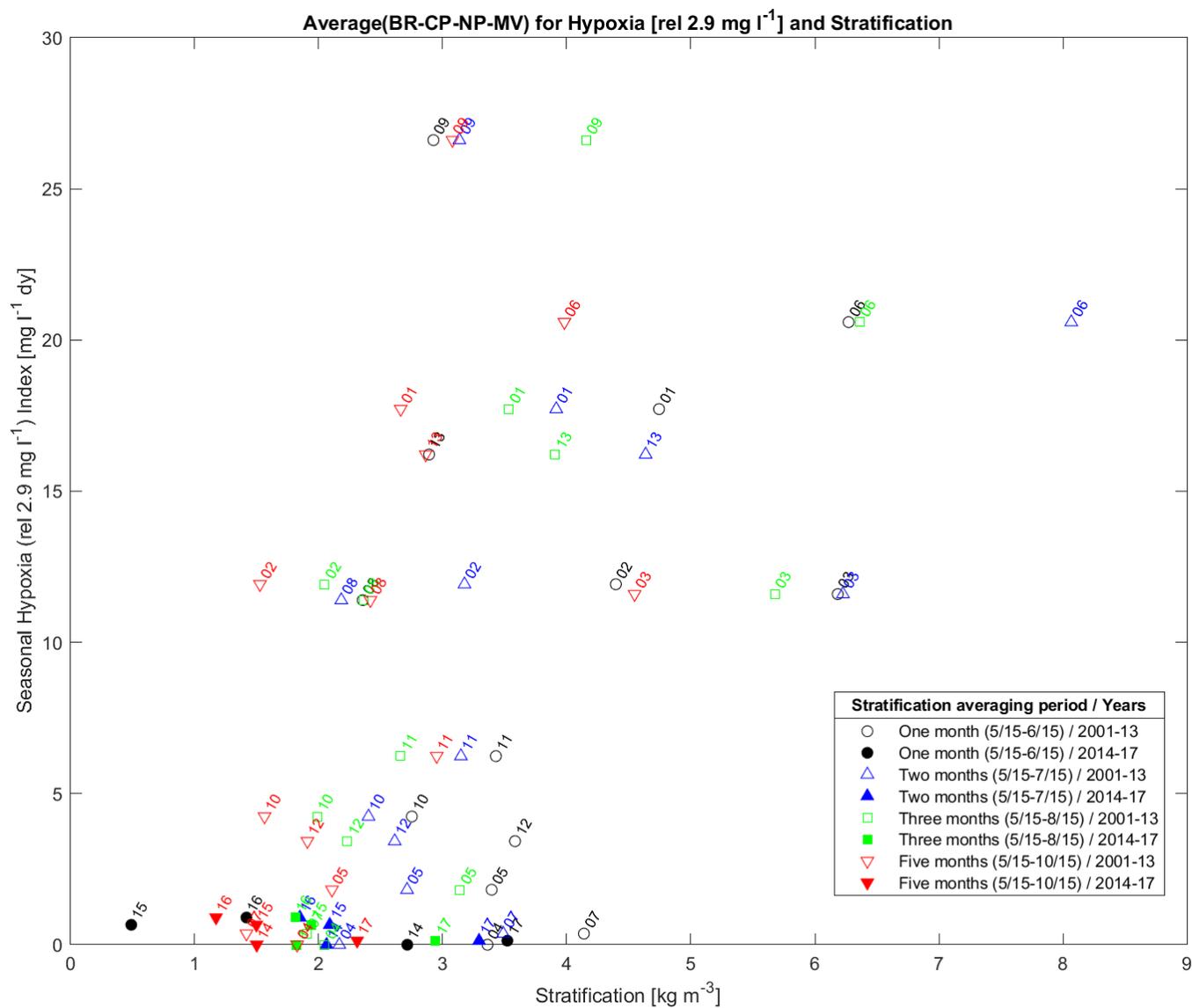


Figure 5-2. Same as Figure 5-1 but seasonal hypoxia index is relative to the 1.4 mg L⁻¹ threshold (more severe hypoxia).

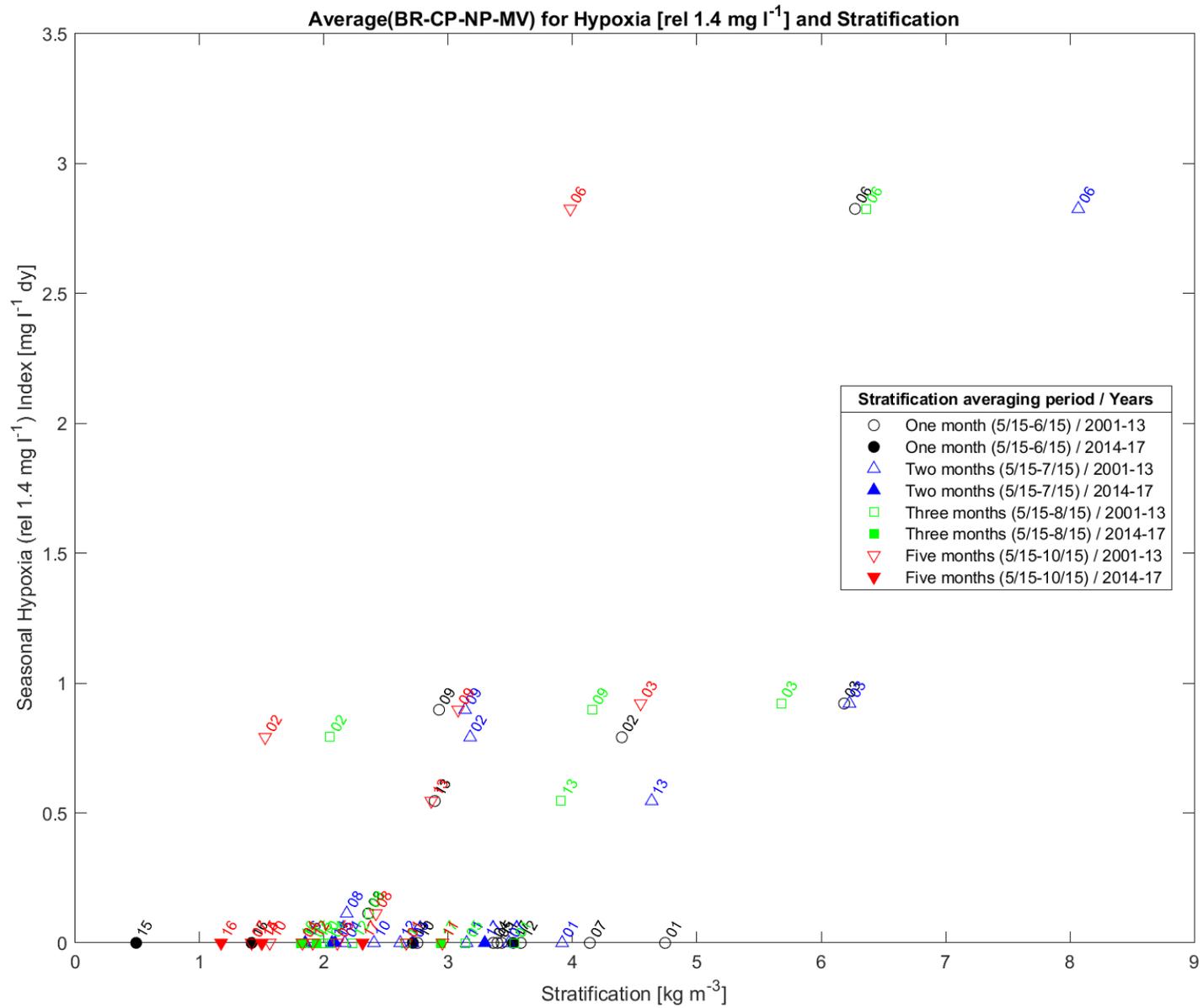


Figure 5-3. Same as Figure 5-1 but seasonal hypoxia index is relative to the 4.8 mg L⁻¹ threshold (milder hypoxia).

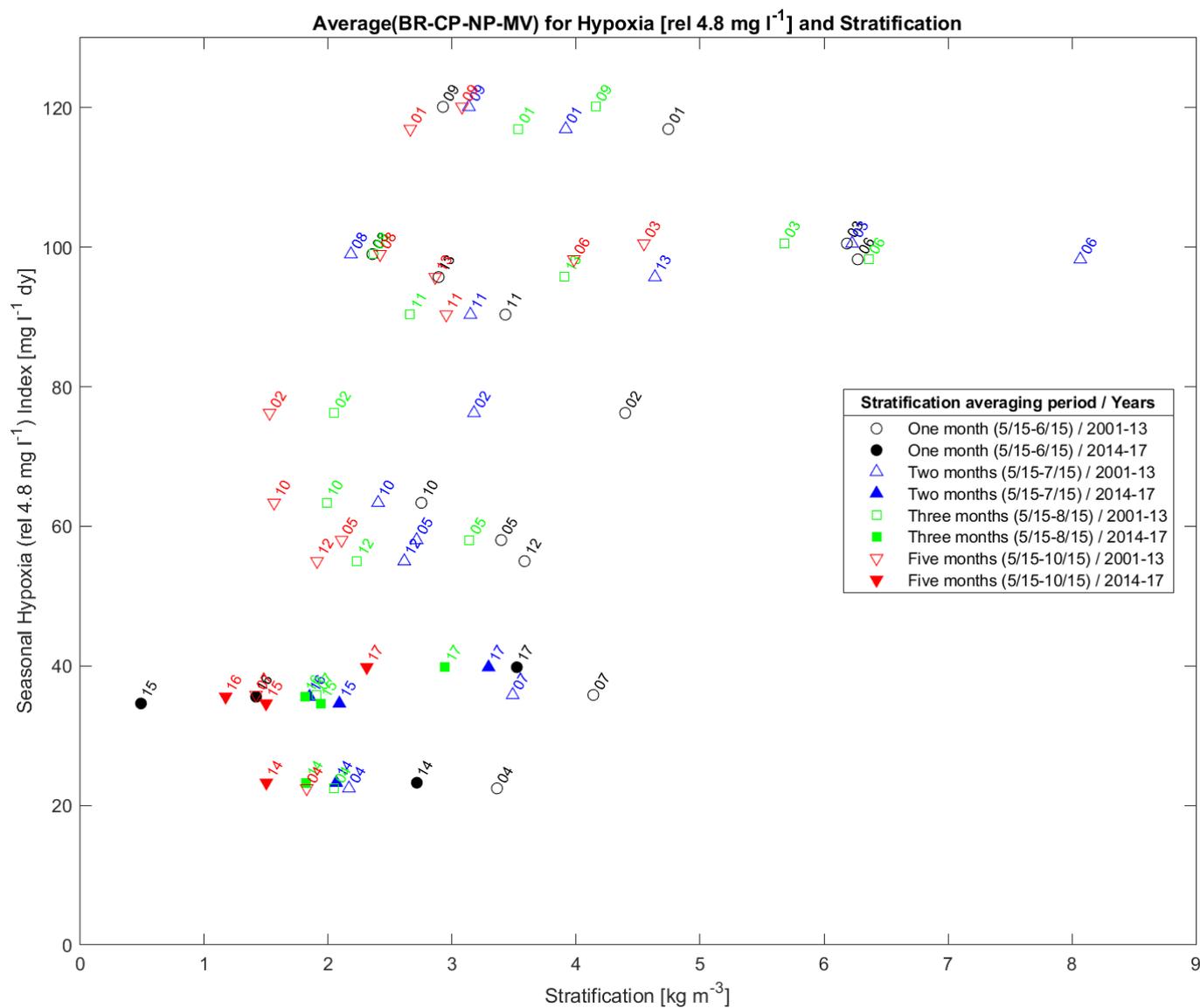


Figure 5-4. Same as Figure 5-1 but using the bay-wide group of stations BR-CP-NP-MV-QP-PP-TW-MH.

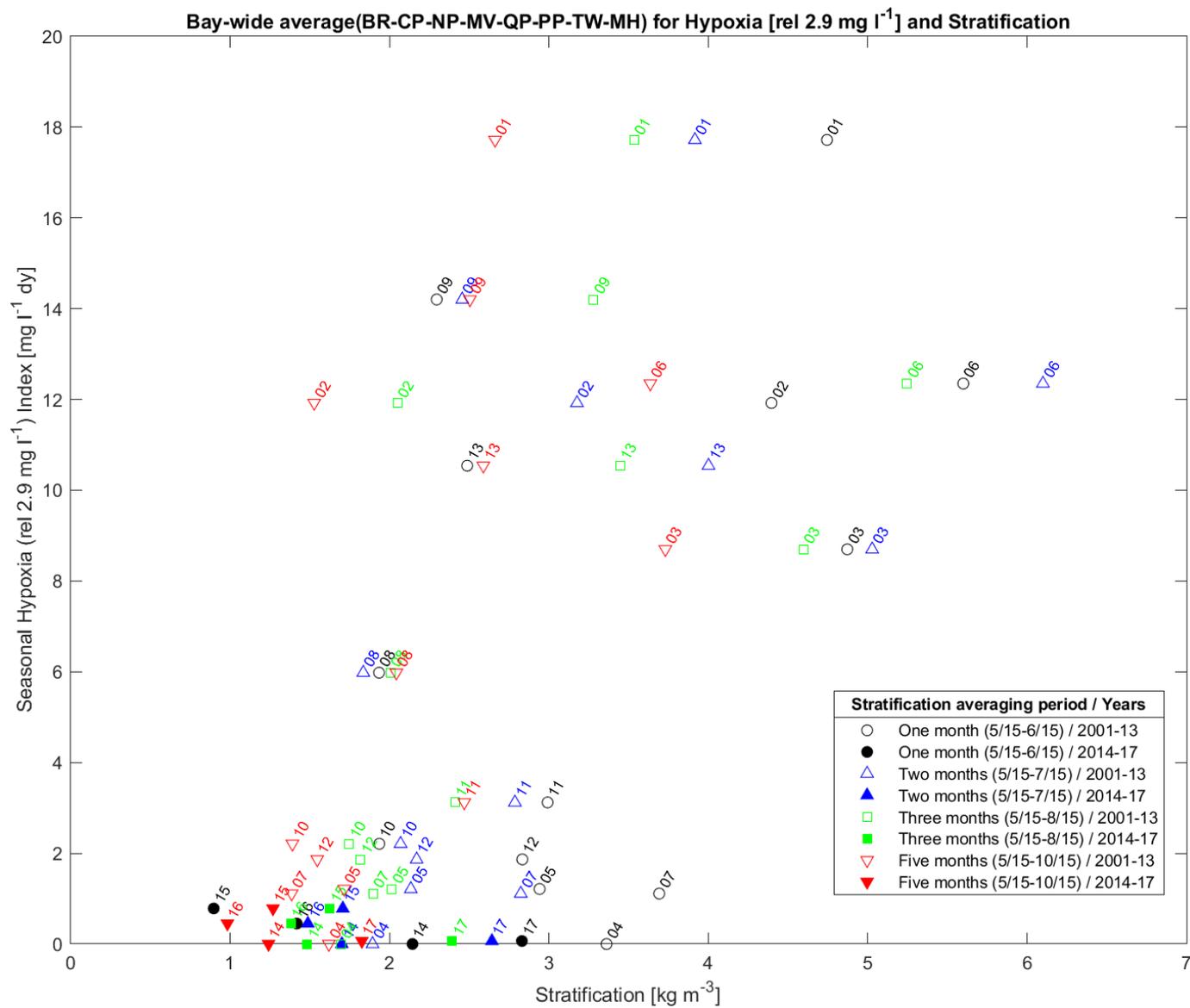
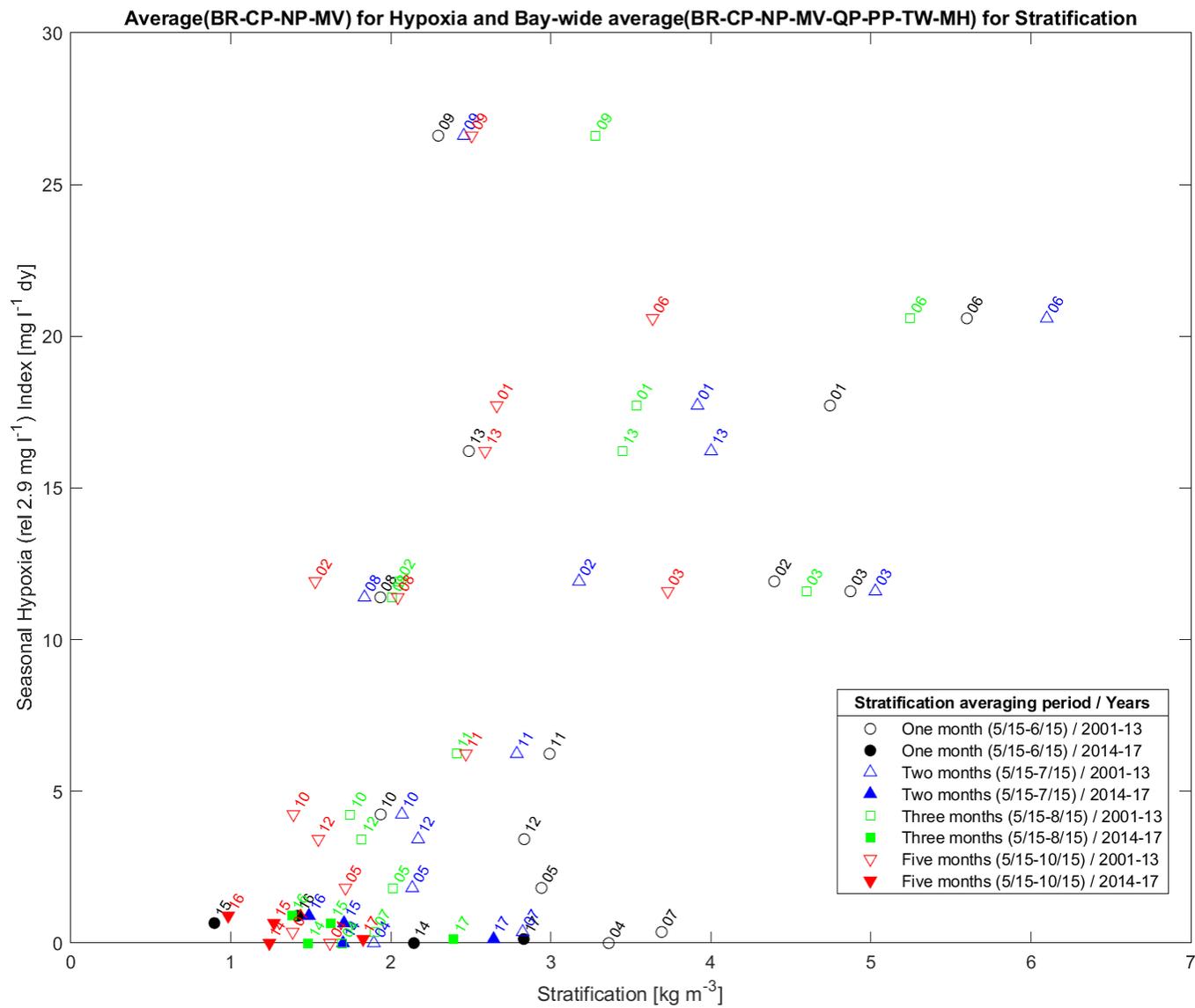


Figure 5-5. Same as Figure 5-1 but using the bay-wide group of stations only for stratification (and the BR-CP-NP-MV group for hypoxia).



Task 6 Drivers of Inter-Annual Hypoxia Variability

6.1 Scope

Objective. Build on prior multivariate analyses to discern which driving factors, among a wide range of biological and physical candidates, are most closely related to hypoxia. **Methodology.** Multiple linear regressions (e.g. Codiga et al 2009, which treated data up through 2006) and similar methods will be used on data through 2017. They will include nutrient load as a potential driver, an important advance over earlier work (which treated river flow, chlorophyll, temperature, stratification, tidal range, non-tidal sea level difference, and wind). River flow data and statistics will be provided by Q Kellogg (URI), who recently completed extensive analyses with them (Kellogg, 2017). (In addition, work proposed here has been planned in coordination with Dr Kellogg, to avoid potential overlap with her ongoing NBEP/SNEP-funded research using NBFSMN data.) **Expected outputs and outcomes.** In years with high river flow, hypoxia is more severe (e.g. Codiga et al 2009, Oviatt et al 2017). Furthermore, density stratification is stronger in years with high runoff (Codiga 2012). An important question to be addressed here, as noted in the SNBIW, is whether the increased hypoxia in wet years is due primarily to river-driven increases to stratification or to nutrient loads.

6.2 Methods

River flow: Obtaining observations and filling gaps. Q Kellogg was consulted, and offered helpful advice. Her previous analysis focused on a few river gauges, selected because they have been sampled for many decades, as decadal variability was the focus of her work. It was determined that her results were not applicable to the present analysis, for which a different set of gauges is important. So the analysis here began with obtaining the river flow observations from USGS for all the appropriate gauges, rather than using the Kellogg results.

Available daily flow observations were obtained from the USGS website for all gauges in Table 6-1 from 1990 to the end of 2017. For the Blackstone River the primary gauge used is the Pawtucket gauge, not the Woonsocket gauge as has been used in many prior analyses, because the Pawtucket gauge is closer to the mouth of the river where it enters the bay.

Not all records spanned all years and were gap-free. Gaps less than 3 days long were filled by linear interpolation. Three rivers had gaps longer than 3 days long, which were filled by regression against a nearby river (right column of Table 6-1, showing the correlation coefficients). Gap-filled records from the 8 gauges in the left column of the table are used in the remainder of the analysis.

River flow: Determining basin-wide and ungauged flows. The aggregate freshwater runoff to the Narragansett Bay basin is taken to consist of 11 sources: 8 gauged rivers, 2 ungauged rivers both estimated by correlation to a gauged river, and the remainder of the ungauged flow. The ungauged Hardig Brook and Maskerchugg River are included as sources, in addition to the 8 gauged rivers. They are included in order to capture flow entering Greenwich Bay, and because of their proximity to sites of long term time series observations to be analyzed; their flows are estimated from regressions against the Hunt River flow, to which they are highly correlated, using coefficients from Ullman et al (2019) where the method is explained.

The aggregate basin-wide runoff, including ungauged areas, is computed from the 8 gauge observation records by a two-stage scale-up process using the results of Ries (1990) based on drainage areas and historical records. This approach is similar to that of Kellogg (2018), who used a one-stage scale up from the gauges to the basin-wide flow and, due to their emphasis on several decades-long conditions, used three gauges with complete long records (Taunton Bridgewater, Blackstone Woonsocket, and Pawtuxet Cranston).

Here, the first stage is to scale up the 8 gauged flows to the flows at the mouth of the river each gauge is on. A unitless scale-up factor equal to or greater than 1.0 is applied to the gauge flow, to account for flow entering the river from drainage areas downstream of the gauge, between the gauge and the river mouth. If the gauge was located at the mouth, the factor would be 1.0. For rivers that have the gauge located near the mouth, or that have little flow entering the river from the drainage area between the gauge and the mouth, the factor will exceed 1.0 by a small amount. The highest scale-up factor is for the Taunton River, because the gauge used is far upstream from the mouth.

For the gauged rivers except the Hunt, 12 long-term mean monthly river mouth scale-up factors (Table 6-2) are computed as the ratio of the long-term mean monthly river mouth flow and the corresponding gauge flow, from Table 7 in Ries (1990). For the Hunt, Ries did not consider mouth flow separately, so the value of one is used for the river mouth scale-up factors. The monthly scale-up factors are assigned to the 15th of every month and interpolated to daily resolution in order to avoid unrealistic step functions from month to month. These daily values are then multiplied by the daily gauge observations to yield daily records for the 8 mouths.

The second stage is to determine basin-wide flow by scaling up the sum of the records for the 8 mouth flows. Twelve monthly values of the basin-wide scale-up factor (last row of Table 6-2) are computed as the ratio of the long-term mean monthly basin-wide flow, which includes ungauged areas, and the sum of the 8 mouth flows (again from Table 7 in Ries, 1990). The basin-wide scale-up factors are assigned to the 15th of every month and interpolated to daily resolution in order to avoid unrealistic step functions from month to month. These daily values are then multiplied by the sum of the mouth flows, to yield the daily basin-wide flow.

Finally, the ungauged flow is computed as the difference of the basin-wide flow result and the sum of mouth flows plus the Hardig and the Maskerchugg; by definition this ungauged flow corresponds to ungauged areas outside of the riversheds of all 10 rivers. Figure 6-1 shows the results, for the period from 2001-2017.

River flow: Classifying individual years as wet, intermediate, or dry. The basin-wide flow, computed as just described, for the years 1990 through 2017 is used. This range spans at least a few cycles of multiyear variability, while also not being overly influenced by conditions prior to 1990. Due to decadal variability (Kellogg, 2018), conditions prior to 1990 are less representative of the 2001-2017 period, which is of primary interest here as this analysis focuses on hypoxia using fixed-site time series observations that were begun in 2001. Years earlier than 1990 were also not included because their records are not available for all of the 8 gauges used (Table 6-1), and the records that are available include more substantial gaps. This method for designating individual years as wet/intermediate/dry can be applied unmodified to future years outside the 1990-2017 range.

The median daily flow during June through September each year from 1990 to 2017 is computed and used to determine its 66.7th and 33.3rd percentiles (Table 6-3). For any individual year, if the median June to September daily flow is higher than 52.43 m³ s⁻¹ it is designated “wet”, those lower than 34.53 m³ s⁻¹ are designated “dry”, and those between the two thresholds are designated “intermediate”.

The method used here for assigning wet/intermediate/dry years is generally similar to that used by SNBIW Chapter 15 for assigning wet/dry years, but differs in one key way. In the SNBIW each year was designated either wet or dry, and no years were designated intermediate. This has the undesired effect that for two years with very similar flow, neither one particularly wet nor dry, one will be designated wet and the other dry. Using the present method, both such years are designated intermediate.

The thresholds computed here from 1990-2017 data (Table 6-3) can be used to designate any year, including future years. Here they are used to assign designations to 2001-2017 (Table 6-4, Figure 6-2). A consistent basis for the designations in future years beyond 2017 will result from continuing to use these same thresholds. This will have the advantage, as compared to re-computing them with inclusion of each additional year as it becomes available, that past years’ designations are not subject to change. If long-term trends in conditions are pronounced in coming years, it potentially could become helpful to compute new thresholds to account for them, but probably only at least a decade from now.

It is important to note that for an individual year, the designation of wet, intermediate or dry will be assigned even though the flow conditions may be very near the boundary between two of these classifications. A prominent example is 2017, which is at the very high end of the intermediate range, with median of daily June to September flow just lower than the 52.43 m³ s⁻¹ threshold defining a wet year. Exploration of sensitivity to use of the May to September period (instead of the May to September period), with all other aspects of the method unchanged, indicated that 2017 would in fact be designated wet in that case. In this light, while the ability to classify years as wet, intermediate, or dry is useful and the method presented here is defined in a clear and justifiable way, it should nonetheless be borne in mind that distinctions between wet and intermediate, or intermediate and dry, can be minor so use of the numerical results (Figure 6-2) is more informative.

Hypoxia. The emphasis is on understanding inter-annual variations in the influence on hypoxia of river flow, stratification, and nutrient load. The seasonal timescale is examined, using bay-wide parameters. This has the advantage that multiple stations are averaged, and river flow and nutrient loads used are collective across all sources, so results are less likely to reflect local processes not representative of the bay overall.

The original scope for this task mentions event-based analysis and multiple linear regressions (MLRs), along the lines of those in Codiga et al. (2009). However, the event-based approach cited does not address inter-annual variability. And because the focus here is inter-annual variations, the number of years of observations available is barely sufficient to do effective MLRs. For these reasons, those specific analysis techniques are not reported on here.

For hypoxia and stratification the bay-wide results were computed as an average of results for all fixed-site stations except Greenwich Bay, Sally Rock, and Phillipsdale Landing. The reason these three stations were excluded is that, as evidenced in the earlier analyses in the SNBIW and prior tasks, they are generally shallower and have hypoxic conditions with severity and temporal variability that are notably different than the other stations, apparently as driven more by local processes than for the rest of the

bay. By excluding these stations the investigation is focused on the 8 deeper stations where hypoxia is less severe and less frequent, with inter-annual variability that tends to be more similar among stations and linked to river flow and associated nutrient load and/or stratification.

The years 2005-2017 are treated, because until 2005 the fixed site network did not yet include many of the present sites.

For hypoxia the metric used is the Hypoxia Index (Task 4), the season-cumulative hypoxia deficit-duration relative to thresholds of 1.4 mg L^{-1} , 2.9 mg L^{-1} , and 4.8 mg L^{-1} .

For river flow, the metric used is the basin-wide flow, computed as described in this section above.

For nutrient load, the metric used is the basin-wide load of Total Nitrogen (TN), computed as described in Task 2 above.

For stratification, three metrics were examined: the mean, the median, and the season-cumulative surplus-duration. For the latter, the surplus-duration was computed using the same algorithm as the deficit-duration for hypoxia, where the events are positive anomalies in stratification, and the thresholds used were 2, 3, and 4 kg m^{-3} .

Four choices for the range of months during the year to use in the calculation for a seasonal value (mean or median) of river flow, TN load, and stratification are used: May-Jun, Jun-Jul, May-Sep, and Jun-Sep. These enable conditions during different parts of the year (for example, late spring vs entire spring-summer period) to be examined.

See page 154 for information about supporting code and data files, and page 150 for how to use them.

6.3 Results

A time series plot (Figure 6-3), using representative results from hypoxia threshold 2.9 mg L^{-1} and mean stratification, captures most aspects of the results. Each parameter (hypoxia, river flow, nutrient load, stratification) is normalized relative to its maximum value during 2005-2017.

The hypoxia index (red line, Figure 6-3) shows pronounced inter-annual variability with clear peaks in 2006, 2009, 2013 and the lowest minima in 2005, 2007, 2010, 2012, and 2014-2017.

River flow (blue lines, Figure 6-3) varies inter-annually with peaks and minima similar to those in hypoxia. The pattern of highs and lows in river flow and hypoxia paralleling each other extends from 2005 through 2013. In contrast, as discussed in Task 5 above, the years 2014-2017 show hypoxia markedly lower while river flow is similar to past years, though not including a year as wet as the wettest past years.

The river flow results demonstrate the sensitivity to the month range used. For example, in 2011, the late springtime river flow was modest and the late summer river flow was higher than usual. As a result the May-Jun (solid blue line) and May-Jul (dashed blue line) river flow is notably lower than for May-Sep (dotted blue line) and Jun-Sep (dash-dot blue line). In contrast, for 2006, the late springtime river flow dominated, so results for May-Jun, May-Jul, May-Sep, and Jun-Sep are all similar.

Stratification (black lines, Figure 6-3) varies inter-annually with overall range similar to river flow. Its variations for different month ranges are also similar to those for river flow.

TN load (green lines, Figure 6-3) shows patterns also comparable in some ways to those in river flow, including the nature of year-to-year variability. TN load differs from river flow in that the overall range from highest to lowest values is smaller, and the variability in results when using different month ranges is less because loads are more uniform than river flow during the months included. In addition, in TN load there is a long-term trend of declining values, due to treatment plant upgrades.

Note that in 2013, the TN load was a local maximum, comparably high to the previous years with peaks, and not particularly sensitive to the month range used. This indicates that, although many of the major plant upgrades were completed by 2012-2013, the main impact of the associated reduction to TN loads did not occur until 2014. In light of the fact that plant upgrades were completed in 2013 and yet 2013 had strong hypoxia, there has been discussion among researchers about a possible lag of the bay response to reduced TN loads.

It is informative to examine the dependence of these same normalized hypoxia, TN load, and stratification results on river flow using a parameter-parameter plot (Figure 6-4). The tightest relationship is between stratification (black symbols) and river flow, a strong linear correlation that is not particularly sensitive to the range of months used (square, asterisk, point, triangle symbols). For TN load (green symbols, Figure 6-4) there is a similar relationship, though it is not as tight as between stratification and river flow. For hypoxia (red symbols, Figure 6-4) the relationship to river flow is the weakest. For example, when river flow is near the middle of its range, hypoxia spans the range from minimal to peak values.

The same plots (as Figure 6-3 and Figure 6-4) have been made and examined (not shown) using the other hypoxia thresholds (1.4 mg L^{-1} and 4.8 mg L^{-1}) and the other stratification metrics (median; surplus-duration). While there are some changes in the details, none of the general characteristics of the relationships among hypoxia, river flow, TN load, and stratification are dramatically different from the above descriptions. This is not unexpected. Earlier results for hypoxia (Task 4) have shown that patterns are generally similar using different thresholds. Insensitivity of stratification results to use of mean, median, or surplus-duration reflects that its variations are not strongly nonlinear or skewed.

Discussion. It is well understood that years with more severe hypoxia occur when there is more river flow. A challenging remaining question is whether this is due to higher nutrient load and increased productivity fueling excess organic matter that decays, or due to stronger stratification that more effectively isolates deep waters and limits re-oxygenation by vertical mixing. Both of these processes are driven by river flow. In effect, the driving factors (nutrient load and stratification) are correlated with each other, making it difficult to isolate their relative importance.

One conceptual model is that nutrient load is the more important driver of inter-annual variability in hypoxia. In this view, stronger river flow brings more nutrients which is sufficient to drive more severe hypoxia, while stratification may increase hypoxia severity as well but plays a secondary role and is not necessary. An alternative conceptual model is that stratification is the more important driver because there is a minimum threshold for nutrients, above which hypoxia is supported, which is always exceeded. Stated differently, in this view there are always sufficient nutrients to drive hypoxia, but hypoxia is only severe in years when there is strong stratification that isolates deep water.

This investigation has not conclusively identified one of these conceptual models as more applicable than the other. The ability of the empirical descriptive approach taken here to distinguish relative influences of TN load and stratification is limited by the correlation between them, as they both originate from river flow. Better success in isolating TN load and stratification effects will require other approaches that are able to discern more directly the mechanisms by which each influences hypoxia.

For intermediate river flow (normalized Q about 0.4-0.65), the results suggest that neither nutrient load nor stratification can explain the inter-annual variability in hypoxia, meaning neither of the two conceptual models noted is very applicable. This is made most clear by Figure 6-4. The inter-annual variability in hypoxia as a function of river flow is more pronounced than for either TN load or stratification. Stated differently, when river flow is intermediate, the TN load and stratification are intermediate as well, but hypoxia can be either very strong or very weak. For example, from Figure 6-4 the years 2005, 2007, 2008, 2009, and 2017 had river flow, stratification, and nutrient load conditions that were intermediate for at least some of the averaging intervals; however, while 2009 had strong hypoxia and 2008 had intermediate hypoxia, the other three years had weak hypoxia.

This suggests that in years with moderate river flow, factors other than TN load and stratification determine how severe hypoxia will be. These other factors could be related to any number of different processes that vary inter-annually. There are many candidates for such processes, biological, chemical, and physical.

Temperature is an obvious consideration, as it sets metabolic processes. Physical processes include circulation-related conditions such as flushing time and mixing conditions; specific related parameters are the wind speed and direction, tidal range cubed, and non-tidal sea level difference between Providence and Newport. Previous analyses have shown that inter-annual variations in wind and tidal range conditions are not strongly linked to stratification (Codiga 2012). However, in a year with moderate river flow and moderate stratification, strongly differing winds or tidal range during the spring-fall months could potentially determine whether hypoxia is weak or strong though their influence on advection, for example by strengthening or weakening the vertically sheared estuarine circulation and changing residence times. Sea level difference between Providence and Newport reflects the large scale pressure gradient between the northern and southern bay and is coarsely indicative of residual circulation strength; earlier analyses suggested it may play a role in hypoxic events (Codiga et al., 2009).

For these reasons the hypothesis that temperature, winds, tides, or sea level differences are responsible is explored. Temperatures from all NBFMSN stations, near-surface and near-bottom, were averaged from mid-May to mid-October. Winds, tidal range, and sea level differences are as tabulated in Task 3 above. Winds are from the North American Regional Reanalysis data-assimilative meteorological model. Tidal range is computed from sea level records at Newport. Sea level differences are from inverse-barometer adjusted non-tidal water level observations, Providence station less Newport station. As noted in Task 3 the winds and sea level were low-passed with 25-hour half-width triangle weight and all parameters were sub-sampled to 12 hour resolution, and averaged over each of the same month ranges as in Figure 6-4. For each parameter, the anomaly relative to the mean of 2005-2017 is normalized by the maximum anomaly magnitude (1.19° temperature; 0.960 m s⁻¹ wind speed; 39.4° wind direction; 0.23 m³ tidal range cubed; 1.23 cm sea level difference).

Results (Figure 6-5) indicate that none of these parameters can explain the differences, during moderate river flow years, between hypoxia that is weak (2005, 2007, 2017), moderate (2008), or strong (2009).

For legibility Figure 6-5 only shows the wind, tidal, and sea level difference results for the May-June averaging period, which is representative of the other averaging periods. Patterns of extreme values that tend to have opposite sense in 2005, 2007, and 2017 as compared to 2009, and moderate 2008 values, would indicate support for the hypothesis and are not apparent. Further work is needed to better understand the factors that influence whether hypoxia is weak or strong during years with moderate river flow when stratification and nutrient loads are therefore moderate.

Table 6-1. Gauges used for river flow observations.

River used in analysis	River used for regression and gap-filling
Taunton (Bridgewater 01108000)	Threemile (North Dighton 01109060) $r^2 = 0.940$
Blackstone (Pawtucket 01113895)	Blackstone (Woonsocket 01112500) $r^2 = 0.993$
Pawtuxet (Cranston 01116500)	
Ten Mile (East Providence 01109403)	
Woonasquatucket (Centerdale 01114500)	
Moshassuck (Providence 01114000)	
Palmer (Reed St, South Rehoboth 01109220)	Ten Mile (East Providence 01109403) $r^2 = 0.968$
Hunt (East Greenwich 01117000)	

Table 6-2. Scale-up factors computed from long-term mean flow, from Table 7 of Ries (1990). The first 8 rows are the scale-up factors from the gauge (see left column of Table 1 above) to the mouth of each river. The last row is the scale-up factor from the sum of the 8 river mouth flows to the basin-wide flow.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Taunton	2.29	2.20	2.20	2.23	2.23	2.49	2.03	2.17	1.91	2.27	2.13	2.29
Blackstone	1.07	1.07	1.07	1.07	1.07	1.07	1.07	1.07	1.07	1.07	1.07	1.07
Pawtuxet	1.21	1.21	1.19	1.19	1.20	1.20	1.20	1.21	1.20	1.21	1.17	1.20
Ten Mile	1.02	1.02	1.02	1.03	1.03	1.03	1.03	1.03	1.03	1.03	1.02	1.02
Woonasquatucket	1.37	1.36	1.36	1.36	1.36	1.36	1.36	1.36	1.36	1.36	1.37	1.36
Moshassuck	1.47	1.45	1.48	1.50	1.49	1.53	1.59	1.56	1.57	1.53	1.46	1.44
Palmer	1.02	1.02	1.02	1.02	1.02	1.02	1.02	1.02	1.02	1.02	1.02	1.02
Hunt	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Basin-wide	1.30	1.29	1.25	1.27	1.33	1.37	1.64	1.79	1.72	1.59	1.45	1.35

Table 6-3. Threshold values, for median daily basin-wide flow between June and September based on 1990-2017 conditions, to designate individual years as wet, dry, or intermediate.

Period of year	“Wet” above (66.7 th percentile):	“Dry” below (33.3 rd percentile):
June to September	52.43 m ³ s ⁻¹	34.53 m ³ s ⁻¹

Table 6-4. Designation of bay-wide river flow conditions each year from 2001 to 2017 as “wet”, “intermediate”, or “dry”, using the criterion in Table 6-3.

Wet	2003, 2006, 2009, 2011, 2013
Intermediate	2001, 2004, 2005, 2007, 2008, 2012, 2017
Dry	2002, 2010, 2014, 2015, 2016

Figure 6-1. Daily river flows 2001-2017, 10 rivers and ungauged area: basin-wide flow.

Inset shows 2016 only.

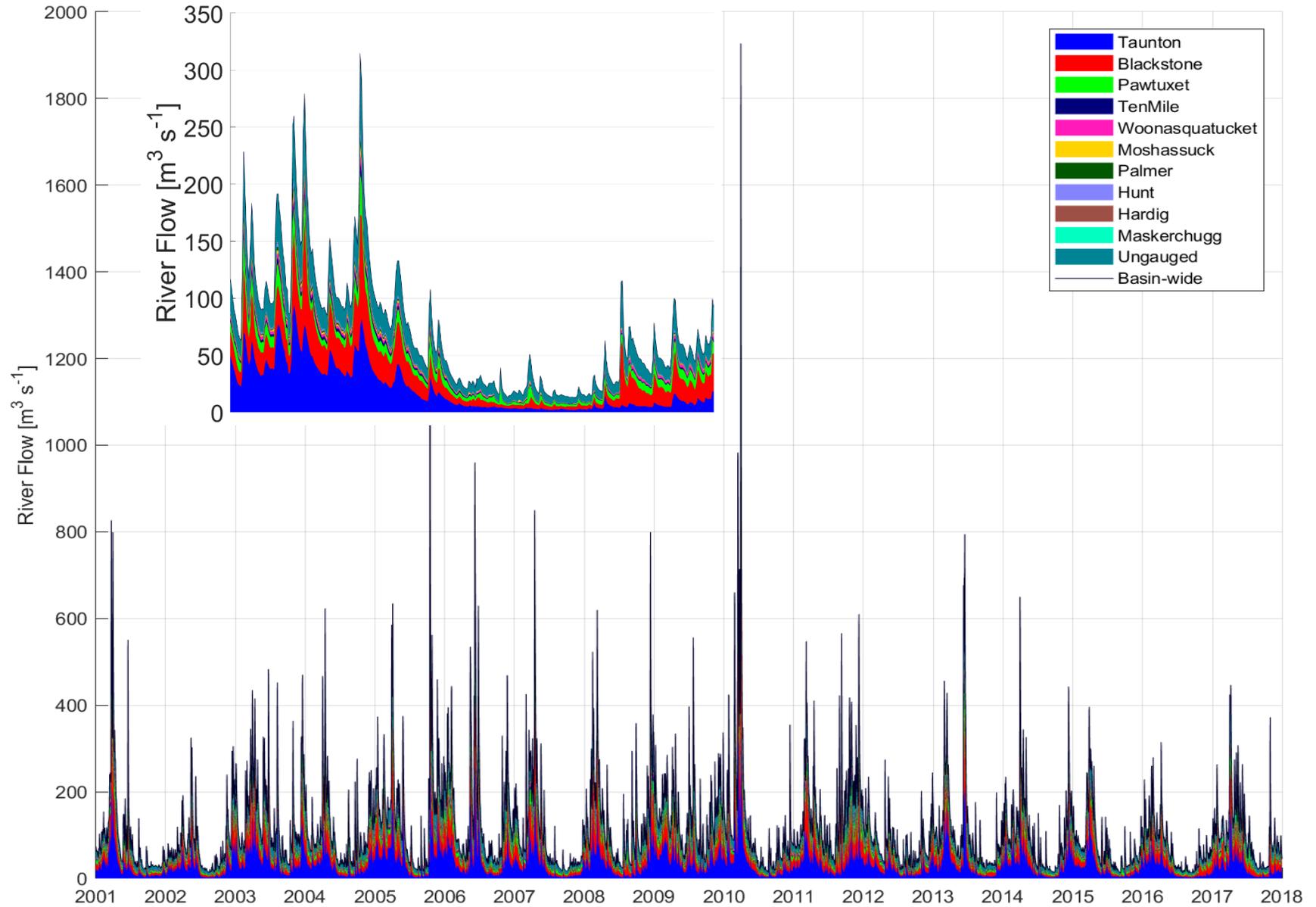


Figure 6-2. Median daily basin-wide flow, June to September (black line).

From the two-stage scale-up method, with the new thresholds shown (red lines), and the resulting designations of wet (blue squares), intermediate (green triangles) and dry (orange circles) years as in Table 6-4. The gray background bars (darkest wet, medium intermediate, lightest dry) demonstrate how wet, intermediate, and dry flow conditions are indicated on plots of other quantities in other sections of this report.

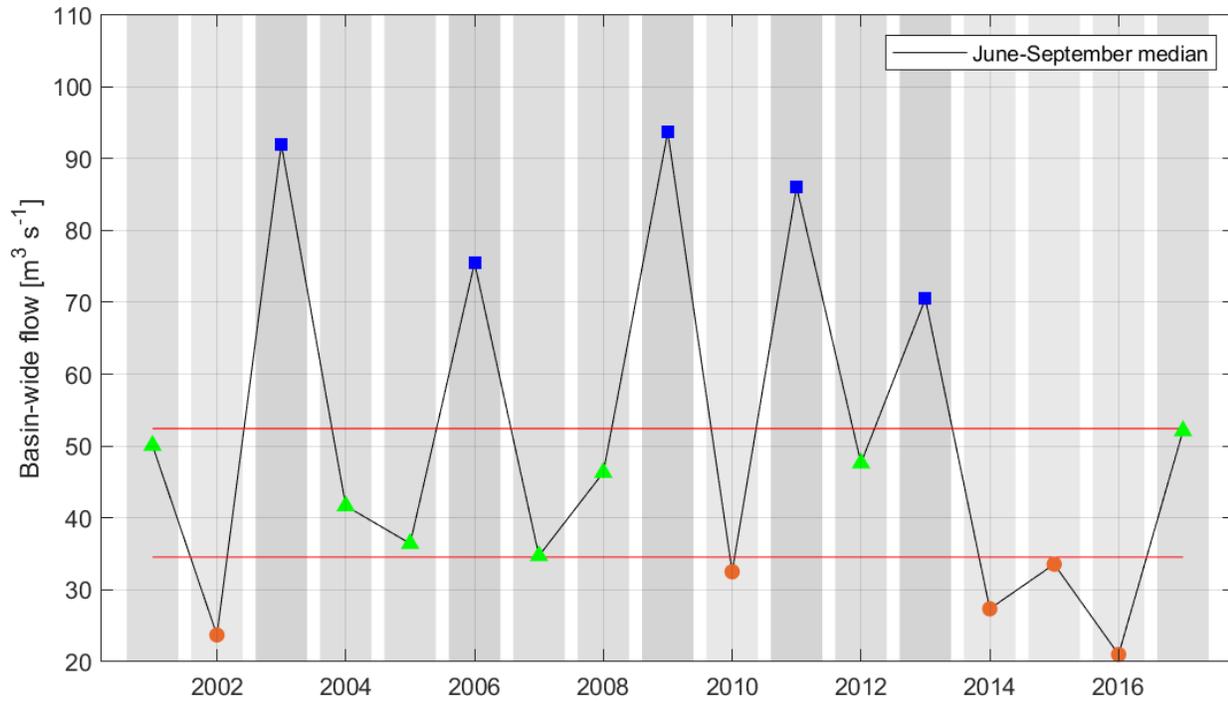


Figure 6-3. Hypoxia (red), river flow (blue), TN load (green), and stratification (black), 2005-2017.

Each is normalized by its maximum value. Hypoxia is the season-cumulative deficit-duration relative to 2.9 mg L^{-1} . All are bay-wide (hypoxia and stratification averaged over stations BR-CP-NP-MV-QP-PP-TP-MH; river flow and TN load scaled from all sources). For river flow, TN load, and stratification, four different month ranges are used: May-Jun (solid lines), Jun-Jul (dashed lines), May-Sep (dotted lines), and Jun-Sep (dash-dot lines).

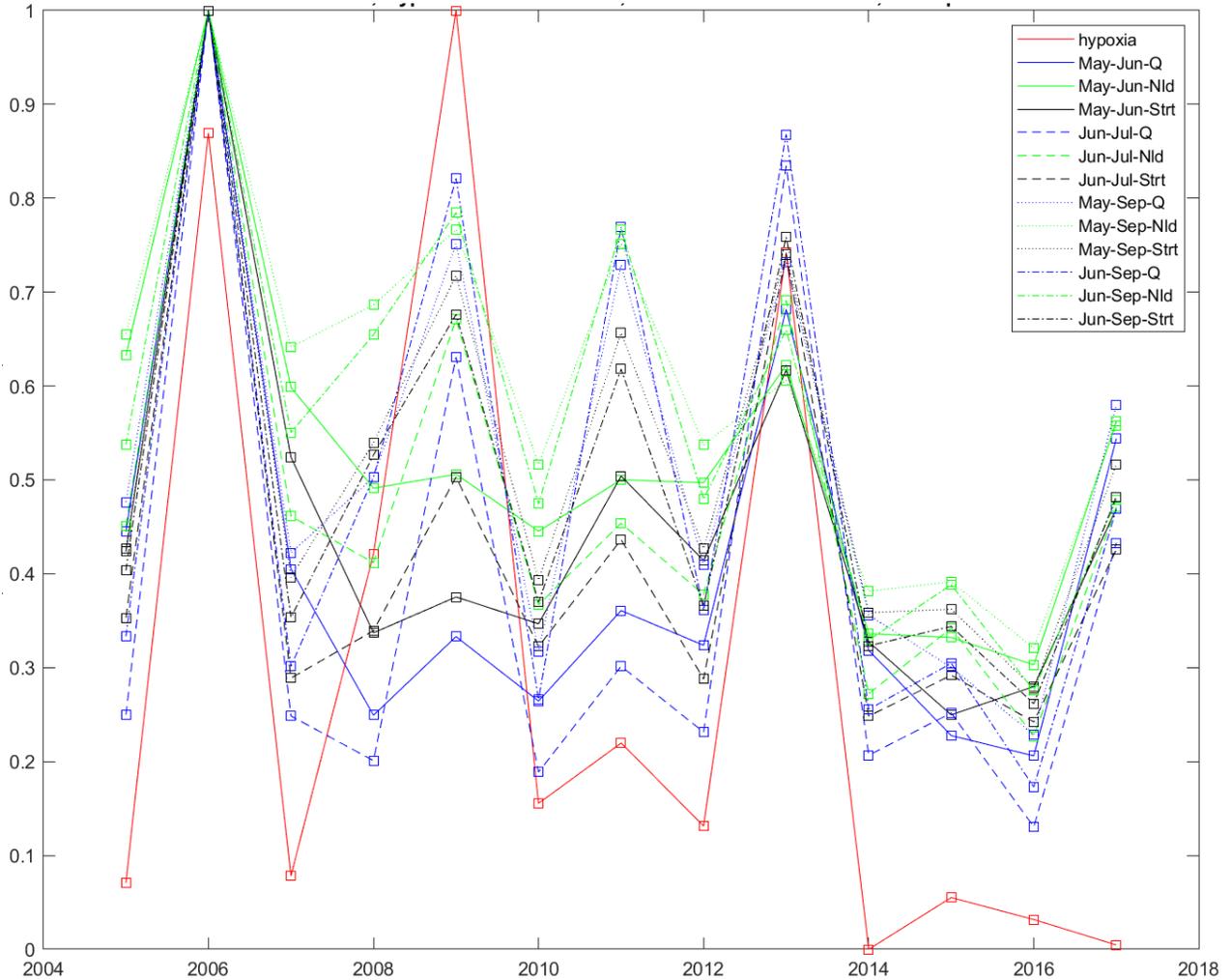


Figure 6-4. Property-property plot of results in Figure 6-3.

Shows dependence of hypoxia (red), TN load (green), and stratification (black) on river flow Q. Results from using different month ranges are shown as different symbols (see legend inset).

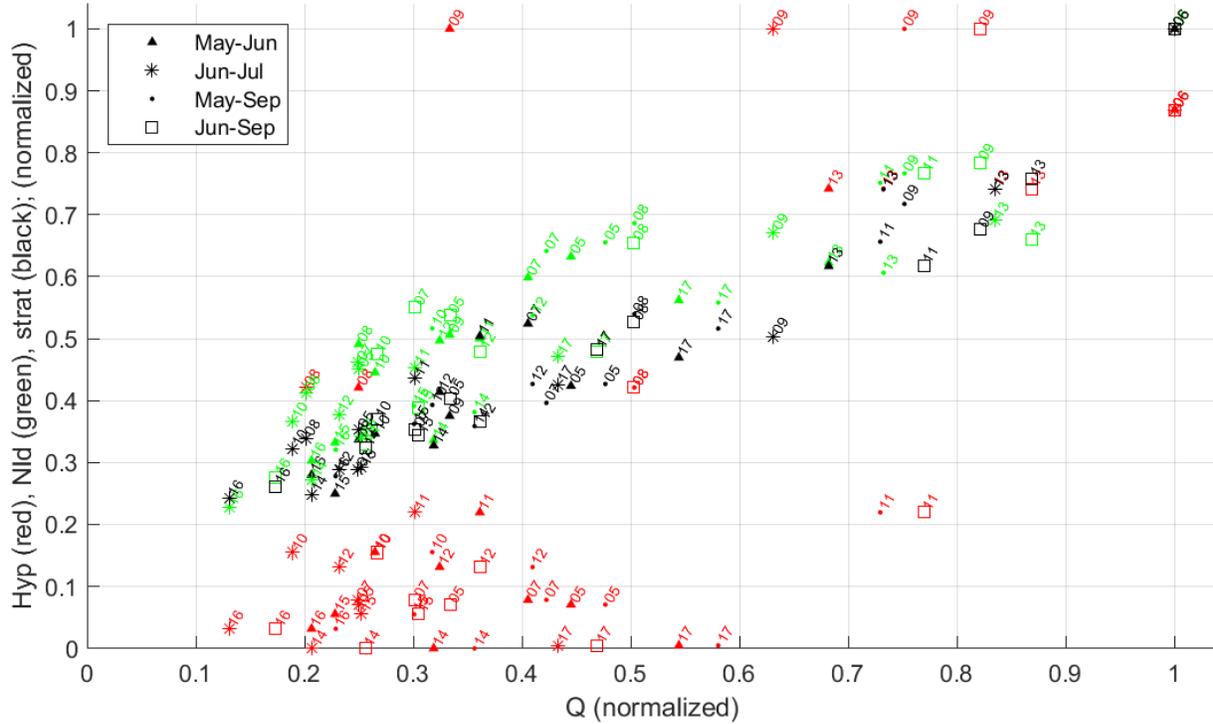
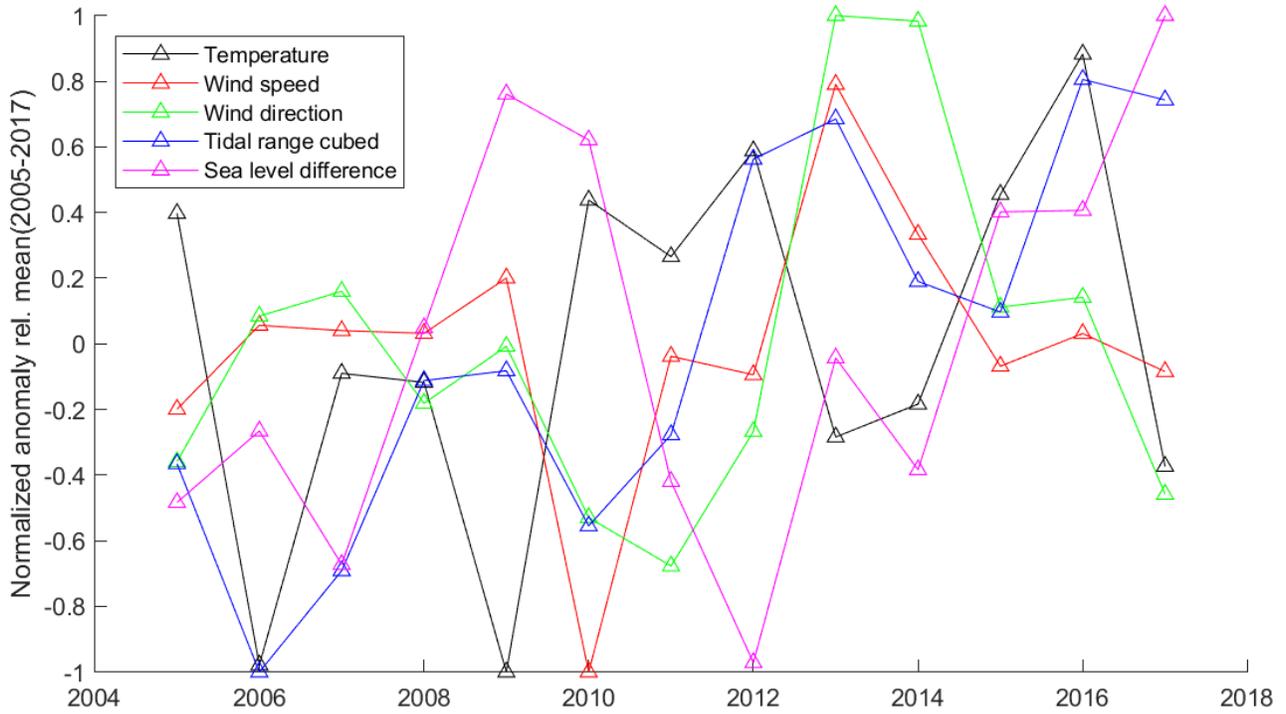


Figure 6-5. Inter-annual variability of temperature and physical drivers.

Averaged over the May-June period, which is representative of results using the other monthly periods shown in Figure 6-4.



Task 7 Time Series and Spatial Survey Datasets Together: Oxygen

7.1 Scope

Objective. Improve knowledge of hypoxia spatio-temporal characteristics. Shed light on the representativeness and uncertainty in the spatial information from the NBFSMN time series dataset, and in the temporal information from the spatial surveys dataset. **Methodology.** The spatial surveys dataset (Prell et al, 2016; <http://www.geo.brown.edu/georesearch/insomniacs/>) will be provided by Warren Prell, Brown University, including the percent hypoxic area metrics presented in the SNBIW. A new, daily-timescale, deficit-duration metric will be created from the time series data. Decorrelation timescales and lengthscales will be used to characterize the two datasets and determine relationships between them. **Expected outputs and outcomes.** The time series data provide dense temporal coverage at few sites; the spatial surveys achieve better spatial coverage and resolution, but occur infrequently. This analysis will better quantify their complementary strengths and weaknesses. This will be useful to improve understanding of the spatial and temporal characteristics of hypoxia. It will also be helpful for cost-effective planning of future monitoring strategies.

7.2 Methods

Spatial surveys. Warren Prell provided the raw data file “DO MASTER 2005-2017.xlsx” containing survey measurements (Prell et al., 2015, 2016) at stations located across most of the bay (Figure 7-1, taken from Prell et al 2015) updated to include all years through 2017. See Prell et al. (2015, 2016) for the station names and other information about the dataset. A table was constructed containing the following values, taken from each individual profile, from all surveys and all stations: DO concentration at deepest depth, minimum DO concentration, DO percent saturation at deepest depth, minimum DO percent concentration.

When the 77 stations are ranked in order of increasing long-term mean near-bottom oxygen concentration, the highest-ranked (lowest mean DO) stations tend to be the shallowest and furthest to the northern (Providence River estuary) and western (Greenwich Bay) areas, as expected (Table 7-1; Figure 7-2a). Some important exceptions to this pattern include stations in northern areas for which the bathymetric depth (included in Table 7-1) is shallow enough to be within the surface mixed layer, such that they are ranked among the lowest (e.g., PRT02, PRT03, PRT04). Per guidance from Prell, because variability at these shallower stations is driven by more strongly by surface mixed layer processes that are not the focus here, stations were divided in to two groups: those with median of the deepest-sampled depth, across all surveys, greater than or less than 4 m (Table 7-2). The spatial pattern of ranks of only the deeper group of stations is more smoothly varying (Figure 7-2b). When presenting results of certain calculations, ordering stations using ranks within the two groups separately (Table 7-2) can yield more structured results as compared to ordering them by, for example, rank among all 77, latitude, or station depth.

Inter-station distances (in km), between each pair of survey stations and between each survey station and each fixed-site monitoring network station, were tabulated. These are direct-line distances regardless of whether the path between stations is wholly in-water, so some of them are along a line that passes over land. Nonetheless they give useful information about spatial scales.

Prell also provided “percent hypoxic area” results for each spatial survey, for thresholds 1.4, 2.9, and 4.8 mg L⁻¹, in a spreadsheet. As explained in the SNBIW, the method by which they are generated is based on GIS analysis and consists of first estimating a spatially continuous field of DO values by applying specified correlation length scales in perpendicular directions, along and across an axis aligned to match the orientation of the bay as a whole, to the measurements at spatial survey locations. Using the resulting continuous field, on a pixel by pixel basis, the percent hypoxic area is computed from the ratio of the area with DO values less than the threshold to the total sampled area. Both the individual survey percent hypoxic areas and their annual-means are used.

Narragansett Bay Fixed Site Monitoring Network. For the date and time of each individual spatial survey profile, values from the time series of deep DO concentration and percent saturation at each of the NBFSMN sites were tabulated. This was done using (a) the individual 15-minute resolution time series measurement closest in time to the profile time, (b) the hourly-mean of the time series data centered on the profile time, and (c) the midnight to midnight daily mean of the time series data on the day of the profile.

The daily-mean DO concentration was computed for the entire time period treated in the SNBIW, all days from May 15 through Oct 14 of each year.

New daily-resolution deficit-duration values (units mg L⁻¹ day) were computed for each fixed site time series, relative to the three thresholds (1.4, 2.9, 4.8 mg L⁻¹). Other than the day-long duration limit, these deficit-duration results are the same as used in prior calculations (e.g., on a season-cumulative basis they are the Hypoxia Index). If no values that day are below the threshold the daily deficit-duration is zero; if some are below the threshold, there is a non-zero deficit-duration, proportional to how long the values are sub-threshold and how far below the threshold they extend, because the deficit-duration is a measure of the area swept out below the threshold.

See page 154 for information about supporting code and data files, and page 150 for how to use them.

7.3 Results

Background context. Understanding of spatial and temporal variations of oxygen in Narragansett Bay derives mainly from analysis of two datasets: vessel-based spatial surveys and fixed-site mooring time series. The datasets are known to be complementary, and have been used together only in limited ways. This analysis uses them together and further quantifies the relationships between them.

For the spatial surveys the main strengths are (a) sampling coverage that spans a larger geographic area of the bay, including both deeper and shallower areas, with finer spatial resolution (77 stations), and (b) sampling throughout the water column, with fine resolution (0.5 m) of vertical structure. The main weakness is how infrequently the surveys occur (5-7 times per season), which limits their ability to capture temporal variability.

For the fixed site time series the main strengths are the span of temporal coverage throughout the late spring to early fall season (mid-May to mid-October), and the high frequency 15-minute resolution sampling. The main weakness is that there are relatively fewer stations (11), located mostly in the deeper channelized areas of the bay, and the sensors are at near-surface and near-bottom depths which do not fully resolve the vertical structure of the water column.

In preparation to use the spatial survey and fixed site measurements together, it is useful first to investigate (a) spatial structure using solely the surveys, and (b) temporal structure using solely the fixed site time series.

Spatial surveys only. Using the spatial surveys, correlations (ρ , Pearson's coefficient; n value between 62 and 71, the number of surveys across all sampled years) were computed between individual values of near-bottom DO from all surveys and all years at each of 12 reference stations and corresponding individual values from other reference and non-reference stations having bathymetric depth more than 4 m. The 12 reference stations were chosen to include a range of depths and locations within the deeper channel areas and the shallower areas outside of channels.

Maps present ρ values for correlation between one of the 12 reference stations and the other 56 stations deeper than 4m (Figure 7-3). These are, in effect, updated and extended versions of Fig 11.7 from Saarman et al 2008. They present distinct spatial structure characterizing oxygen variability in the bay. For example, ρ values for reference station UPB10, which is a deeper station in the channel in the upper bay, are elevated for distances of up to about 10 km or more along the channel in both the northern and southern directions. In contrast, ρ values for SR02, which is a shallow station at the most northern reach of the bay, fall off within a few km. In general, the shallower stations are not as strongly correlated to each other, which can be understood to be due to the influence of local effects on oxygen conditions, such as winds and the changing mixed layer depth, that are not as important to near-bottom oxygen at deeper stations. Interestingly, for some stations such as WPS09 which is in the central Upper West Passage, ρ values are relatively high across large areas of the bay, including both shallower areas and deeper channel areas.

The same ρ values can be plotted as a function of distance between each station pair (Figure 7-4), to help quantify the decorrelation length scales. For simplicity the distances used are direct point-to-point lengths between stations, whether the straight line path between them is wholly in water or not, and it should be borne in mind that for some station pairs the line passes over portions of land. As expected, a high degree of scatter exists in the results. The decorrelation length (taken to be where ρ decreases to 0.7) ranges from about 1 to 16 km, with most values between about 4 and 11 km, and an average of about 7-8 km. Decorrelation scales between stations with similar bathymetric depths tend to be larger (solid symbols in Figure 7-4; bathymetric depths not more than 5 m different). For deeper stations located in the channel (e.g., UPB10) the decorrelation length scale lies at the higher end of this range, and vice-versa for shallower stations in embayments (e.g., GRB04), as illustrated in Figure 7-4 by a solid line connecting results from each of these two stations.

Finally, the results can also be presented with the stations ordered by the rank of their long-term mean near-bottom DO concentration (Figure 7-5). Somewhat more organized patterns are present, in comparison to the distance-based results of Figure 7-4, due to the general geographic proximity of similarly-ranked stations to each other (Figure 7-2). The rank-based results reinforce the conclusions from Figure 7-3 and Figure 7-4: that deeper stations in or near the channels are more strongly correlated with each other, that stations with similar bathymetric depths are more strongly correlated, and that shallower stations are not as strongly correlated with each other.

Fixed site time series measurements only. Decorrelation temporal scales were examined using lagged autocorrelations of daily-mean deep DO concentration from time series records at individual fixed sites (Figure 7-6). ρ decreases to 0.7 over timescales between about 2 and 10 days, with most values

between about 3 and 5 days. The shorter timescales apply to the shallower stations such as Phillipsdale and Greenwich Bay. The longer timescales apply to the deeper stations such as T-Wharf and Bullocks Reach.

Combining spatial surveys and time series. Now the two datasets are used together, in context of knowledge that based on the spatial surveys the decorrelation length scale for bottom DO concentration averages 7-8 km and based on the time series its decorrelation time scale is about 3-5 days.

Time series measurements on the day of each spatial survey can be used to further assess decorrelation spatial structure, using the map style presentation, and all 77 spatial survey stations (shallower and deeper than 4m). The results using the daily mean (midnight to midnight) time series deep DO concentration, on the date of the spatial survey, show many similarities (Figure 7-7) to the maps presented in Figure 7-3. The maps of Figure 7-7 are, in effect, updated versions of Fig 11.6 in Saarman et al 2008. Rho values are highest at spatial survey stations located nearest to the time series station, and decrease substantially at distances of between a few km and up to about 10 km.

The rho values for all pairs of spatial survey stations and fixed site stations fill a 77x11 matrix (Figure 7-8, left). The 77 spatial survey stations are ordered by their rank all-survey mean DO, in the two groups (median of deepest-sampled depth greater than or less than 4 m); the deeper stations are ranks 1-57 and the shallower stations are ranks 58-77. The order of the fixed site stations is the same as in Figure 7-7: Phillipsdale (PD), Bullocks Reach (BR), Conimicut Point (CP), North Prudence (NP), Greenwich Bay (GB), Sally Rock (SR), Mount View (MV), Quonset Point (QP), Poppasquash Point (PP), T-Wharf (TW), and Mount Hope (MH). The general pattern for higher rho values at the deeper group of spatial survey stations (vertical axis, 1-57) relative to the shallower group of spatial survey stations (vertical axis, 58-77) is clear. In addition, rho values for the spatial survey stations with the lowest mean DO (highest ranks, vertical axis 1 to about 10), which are mostly very shallow stations at the northernmost reach of the bay or in Greenwich Bay, are notably lower at fixed site stations (horizontal axis, higher than about 5) not located very close to them.

To investigate sensitivity of these results to the use of daily-mean time series results, they were repeated, but using the 15-minute resolution time series measurements nearer in time (instead of their daily mean) to the time of each individual spatial survey profile. Two cases were treated: the hourly mean, using the hour centered on the time of the profile; and the individual 15-minute value nearest to the time of the profile. Results for the hourly mean case show limited bias relative to the daily-mean results (Figure 7-8, right); results for the individual 15-min case (not shown) are very similar to the hourly-mean results, as expected given that variability on timescales as short as an hour is not particularly pronounced. Bias of the hourly mean results relative to the daily mean results (Figure 7-8, right) is strongest at PD (1 on horizontal axis), where rho for hourly means is lower than daily means, and magnitudes of the differences typically reach about 0.2. There also appears to be some bias toward lower rho at TW (10 on horizontal axis), and toward higher rho at PP, and these are weaker differences than at PD. The overall pattern of weak bias demonstrates that variability on timescales shorter than a day is not prominent, and is consistent with the above decorrelation time scale estimates being longer than a day.

Finally, to explore both spatial and temporal variability while also using both datasets together, lagged correlations were computed between daily-mean time series measurements and spatial survey results. Results are complex with substantial scatter, with the BR fixed site station representative (Figure 7-9).

They show rho values at zero lag that are either larger than for nonzero lags or not substantially different from them, and rho that generally decreases as the magnitude of the lag increases up to 5 days, beyond which there is less change. These patterns are weak, and there are exceptions, but in general they corroborate results from spatial correlations and lagged temporal autocorrelations presented above.

Results for profile-minimum (instead of near-bottom) DO, and for DO percent saturation. As noted by Prell et al (2015, 2016), a “mid-depth minimum” in oxygen is a characteristic feature of some spatial surveys. For most spatial survey profiles the value at the deepest depth sampled is the same as the minimum value in the profile; however, at many stations located in the deeper channel areas of the bay, during certain surveys the minimum value in the profile occurs above the bottom at a mid-profile depth.

To investigate the importance of the mid-depth minimum many of the calculations described above, completed using the deepest depth sampled (“near bottom”) in the profile, were repeated but using the minimum values of DO concentration in each spatial survey profile. The results are not presented here, but they were not substantially different than the near-bottom DO results shown. There were some small differences and they were as expected based on the nature of above-bottom DO minima as described by Prell et al (2015, 2016).

To investigate spatiotemporal structure in DO percent saturation, most of the above calculations were repeated but using DO percent saturation instead of DO concentration. These results are not presented here but, as expected, their patterns were very similar to those of DO concentration which have been presented and discussed. The minor differences that occur are as expected, due to the known relationship of DO percent saturation to DO concentration—namely, that DO concentration is affected by both biological production/consumption and physical influences of temperature/salinity on the DO saturation level, whereas DO percent saturation in effect varies dominantly only due to biological influences.

Percent hypoxic area. The percent hypoxic area results are shown in Figure 7-10 for each individual survey, at the three threshold levels of 1.4, 2.9, and 4.8 mg L⁻¹. This is, in effect, an updated version of Fig 9 of SNBIW Chapter 15, augmented to show the 1.4 and 4.8 mg L⁻¹ results, and results through 2017. Two characteristic features are clear: inter-annual variability, and a reduction to lower values over the past 5-10 years as treatment plant nutrient load reductions have taken effect. These two characteristics are similar in many ways to results of the fixed site analysis (for example, hypoxia index), as described in detail in many of the earlier sections of this report.

For the 2.9 mg L⁻¹ threshold the percent hypoxia area exceeds 40% in relatively wet years (2006, 2008, 2009, and 2013), and has not been higher than about 20% in recent years; main features of the 1.4 mg L⁻¹ threshold results are generally similar. Results for the 4.8 mg L⁻¹ threshold are different in that they are only infrequently as low as 10-20%, and appear to have a slightly less pronounced change of results after 2014 relative to prior years.

The extent to which the percent area results differ from the fixed site analysis results (see prior tasks) reflects the different nature of the two datasets: the percent area results capture conditions across both shallow and deep portions of the bay, whereas the time series results are from fixed site stations that are mostly at relatively deeper locations in or near the channel areas. In this context the fact that percent hypoxic area for the 4.8 mg L⁻¹ threshold appears to differ the most from the time series results

can be understood to be due to its dependence on more extensive areas of the bay, both shallow and deep, due to the higher threshold value.

Regressions of percent hypoxic area from individual surveys with the daily-mean time series results for the date of each survey help quantify this relationship (Figure 7-11). In individual frames of Figure 7-11, for each spatial survey there is a group of three symbols aligned at the same x-axis value, the daily-average deep DO at that station on the date of the spatial survey. For clarity of the figures, the regression lines are not shown. At a given station, the r^2 value is generally lowest, intermediate, and highest for the 1.4, 2.9 and 4.8 mg L⁻¹ thresholds, with typical ranges of r^2 values across the 11 stations from about 0.1-0.3, 0.2-0.5, and 0.3-0.7, respectively (Table 7-3). The increasing proportions of zero or near-zero percent hypoxic area for the intermediate and low threshold make a linear model less suitable and contribute to their lower r^2 values. Among the 11 stations, the strongest regressions are with relatively deep stations, located in or near the channels (BR, CP, NP, MV, QP; and, for 4.8 mg L⁻¹, MH). This reflects the reduced variability of deep water in the bay, which is less influenced by local air-sea interactions compared to water nearer the surface. The extent to which the daily mean deep DO at deeper fixed site stations is representative of percent hypoxic area corresponds to a correlation r^2 of about 0.4-0.5 for threshold 2.9 mg L⁻¹ and about 0.6-0.7 for threshold 4.8 mg L⁻¹.

To investigate the potential utility of the new daily deficit-duration results, the regressions of Figure 7-11 were repeated but using the respective deficit-duration results on the horizontal axis (Figure 7-12) instead of the daily-mean DO. Some relationships have strength comparable to the above results, with r^2 values (not shown) as high as about 0.6. However, that is only true for the 4.8 mg L⁻¹ threshold. For the other thresholds there are increasing proportions of deficit-duration values that are zero, contributing to reduced correlation coefficients. The new daily-resolution deficit-duration does not enhance the strength of the relationships, compared to using the daily means.

To investigate the relationship between the two datasets at the time scale of entire hypoxia seasons, the average of the percent hypoxic area results from all spatial surveys each year were regressed (Figure 7-13, Table 7-4) against the seasonal hypoxia index computed from the time series (Task 4 above). These results are an extension of Fig 10 in SNBIW Chapter 15, but here showing all stations and including results for the 1.4 and 4.8 mg L⁻¹ thresholds as well as 2.9 mg L⁻¹. Consider first the results for the 2.9 mg L⁻¹ threshold (blue symbols in Figure 7-13, middle column of Table 7-4). There are relatively strong correlations (r^2 from 0.65 to 0.83 and $p < 0.05$) at three deeper channel stations in the Providence River and Upper Bay (BR, CP, NP), and at two stations (GB, SR) in and adjacent to the shallow embayment Greenwich Bay. There also is a relatively strong correlation at the Poppasquash Point station farther south, in the Upper East Passage, though with lower r^2 value 0.52. Phillipsdale is a shallow embayment station but in contrast to GB and SR its correlation results are weak, likely because its farther north location is more hydrodynamically isolated from the portions of the bay as discussed in other tasks above. Other stations have weak correlation results in part because they are farther south, where hypoxia occurs less, so a larger portion of the data values are zeros.

Next consider the results for the 4.8 mg L⁻¹ threshold, for which fewer data values are zeros (green symbols in Figure 7-13, right column of Table 7-4). The strengths of the correlations and the spatial patterns are generally similar to those described for the 2.9 mg L⁻¹ threshold. Some differences from the 2.9 mg L⁻¹ threshold results are that the BR, CP, and NP r^2 values are lower by between 0.04 and 0.17,

the r^2 values at GB and SR are respectively lower and higher, and the PP result p value is > 0.05 while those for TW and MH are < 0.05 .

Finally, the results for the 1.4 mg L^{-1} threshold (red symbols in Figure 7-13, left column of Table 7-4) include the highest proportion of zero values, particularly at stations other than the shallow embayments (PD, GB, SR). Correlation results from both a Providence River station (PD) and a Greenwich Bay station (GB) are reasonably strong. They also are reasonably strong at the NP, MV, and PP stations, but probably not meaningful as these deeper stations have such high proportions of zeros for this low threshold.

Representativeness of each dataset relative to the other. The representativeness of the spatial survey temporal sampling, relative to the temporal variability captured in the time series records, was investigated as follows. Using all years of data at each fixed-site station, the histogram of all daily-mean DO values in the time series was compared to the histogram from daily-mean values only on dates of the spatial surveys (Figure 7-14). The date range used in each year was from the earliest spatial survey to the latest. For most stations it is difficult to perceive substantial differences in the distributions. This suggests that despite how infrequent the spatial surveys are, the temporal variability they capture is not unrepresentative. The Kolmogorov-Smirnov test indicated the two distributions were statistically different at 3 of the 11 stations (Sally Rock, Poppasquash Point, and Mount Hope), with lower DO more prevalent using survey dates only as compared to using all dates. However, when the calculations were repeated using all dates from May 15 through Oct 14 each year, the distributions (not shown) were statistically different at all stations. The late spring and early fall parts of the season, before the earliest survey and after the latest survey, generally have higher DO values because they are before and after the peak timing of hypoxic events, and including them in the calculation shifts the distribution using all dates far enough toward higher values to change the outcomes of the test. This assessment of the representativeness of the spatial surveys is sensitive to whether dates before the earliest survey and after the latest survey are included, and when they are included the conclusion is that the distributions are different.

The representativeness of the spatial sampling in the time series, relative to the spatial variability captured in the spatial surveys, was investigated in a similar way (Figure 7-15). Histograms were constructed using spatial surveys measurements only. One uses results from all stations and the other only uses the 11 spatial survey stations located nearest to the 11 fixed site stations. The latter results are shifted toward lower DO values, as expected because the fixed site stations are located in deeper areas. The statistical test indicates the distributions are different.

The Kolmogorov-Smirnoff test indicates representativeness of each dataset for the other is not complete. This is to be expected. However, the differences are generally difficult to discern, as descriptive summary statistics and plots of each dataset reveal many similar patterns independently.

A separate way to assess the extent to which the spatial surveys capture temporal variability in the time series is to compare the statistical strength of the regressions of spatial survey results with (a) time series results from only the same days as the surveys (Figure 7-11 and Figure 7-12; Table 7-3) and (b) time series results from all dates (Figure 7-13, Table 7-4). The extent to which the latter has weaker statistical strength is an indication that the spatial surveys are not representative of temporal variability in the time series.

Summary. The spatial surveys dataset alone captures detailed spatial structure, with station-to-station decorrelation length scale of about 7-8 km on average. This length scale is longer (up to typically 11 km) for deeper stations in or near the channels and for station pairs with similar bathymetric depths to each other, and is shorter (down to typically 4 km) for shallower stations. The time series dataset alone captures detailed temporal structure, with decorrelation time scale of about 3-5 days on average; this time scale is longer/shorter (typical range 2-10 days) for deeper/shallower stations. The two datasets have been used together here in various additional ways to further explore and quantify these spatiotemporal patterns.

Percent hypoxic area results have been regressed against daily-mean fixed site time series results on the dates of the spatial surveys. The fixed site stations that are most strongly correlated are BR and CP, and to a lesser extent NP, MV, QP, and MH. These are all deeper stations in and near the channels. The correlation coefficients reach as high as about 0.6-0.7 when using the 4.8 mg L⁻¹ threshold and about 0.4-0.5 for the 2.9 mg L⁻¹ threshold. The correlations are weaker when using daily deficit-duration results instead of daily-mean oxygen concentrations.

At the timescale of the entire hypoxic season, the annual-mean percent hypoxic area is strongly correlated to the Hypoxic Index (Task 4) at certain stations. For the 2.9 mg L⁻¹ and 4.8 mg L⁻¹ thresholds, the stations with the strongest correlations ($r^2 > 0.65$, $p < 0.05$) are in deep channel areas of the Providence River and Upper Bay, or in and near Greenwich Bay. For the 1.4 mg L⁻¹ threshold both the Phillipsdale and Greenwich Bay stations show reasonably strong correlations.

The spatial surveys are separated in time by typically at least 1-2 weeks, well longer than the autocorrelation timescale from the time series dataset. Conversely, the time series station locations do not span shallow and deep areas of the bay to the same extent as the spatial surveys. The Kolmogorov-Smirnov test suggests the two datasets are not fully representative of each other in these expected ways. While these are initial evaluations of the representativeness of the datasets for capturing variability of their counterpart, there are many other methods possible. Despite the statistical result confirming differences, they are hard to discern, and the main attributes of spatiotemporal variability are seen in both datasets independently.

With respect to design of future monitoring, these results can be used to help choose what types of monitoring to prioritize, given the goals for temporal and spatial characteristics to be monitored.

Table 7-1. Ordering and ranks of spatial survey stations by increasing long-term mean oxygen concentration.

Column 1: Station name
 Column 2: Median deepest sampled depth [m]
 Column 3: Average (all years, all surveys) deepest-sampled bottom oxygen concentration [mg L⁻¹]
 Column 4: Rank within 77 stations (higher rank is higher oxygen concentration)

PRC07	6.5	1.19	1	UPB10	14.5	4.07	28	WPS11	6.5	4.77	55
PRN03b	13.0	2.12	2	GRB_A	4.0	4.08	29	WPS13	6.0	4.77	56
PRN07	8.0	2.17	3	WPS10	9.0	4.08	30	BH07	5.0	4.79	57
SRO1	3.5	2.37	4	UPB09b	5.0	4.09	31	WPS09	6.5	4.80	58
PRN01	12.5	2.45	5	GRB03	3.5	4.11	32	PRT02	2.5	4.82	59
SRO2	5.5	2.58	6	GRB09	11.5	4.13	33	EPN03	14.5	4.87	60
PRC05b	14.0	2.85	7	UPB11	13.5	4.15	34	WPS14	7.0	5.04	61
PRN08	7.5	2.86	8	BBT2	6.5	4.15	35	RWU03	7.0	5.04	62
SRO3	6.5	2.88	9	BRT1	5.0	4.20	36	EPN04	8.0	5.06	63
GRB04	4.5	2.96	10	GRB_D	3.5	4.21	37	PRT01	2.5	5.07	64
GRB05	3.0	3.07	11	UPB08	14.0	4.27	38	PRS04	3.0	5.10	65
PRC02	13.5	3.17	12	WPS07	10.0	4.29	39	GRB10	3.0	5.11	66
BRT3	8.0	3.36	13	UPB04b	6.0	4.29	40	PRT04	2.0	5.13	67
PRS07	14.0	3.41	14	UPB01	18.5	4.30	41	EPN02	19.0	5.20	68
BRT4	6.5	3.48	15	WPS05	10.5	4.31	42	PRT03	2.5	5.31	69
BRT2	13.5	3.52	16	BH10	5.0	4.33	43	WPS01	12.5	5.43	70
BBT1	5.0	3.58	17	UPB06	5.5	4.38	44	BRT6	2.5	5.46	71
GRB02	2.5	3.59	18	UPB05	15.0	4.39	45	RWU05	19.0	5.50	72
UPB13	6.5	3.66	19	EPN08b	7.0	4.49	46	WPS02	11.5	5.54	73
UPB03	12.5	3.70	20	GRB_B	2.5	4.53	47	GRB12	3.0	5.60	74
PRS03	13.0	3.72	21	EPN06	14.0	4.53	48	EPS01	21.5	5.81	75
GRB01	3.0	3.79	22	BH05	4.5	4.56	49	UPB12	3.0	5.89	76
GRB07	8.0	3.88	23	WPS04	10.5	4.58	50	WPS08	2.5	6.22	77
GRB06	4.5	3.90	24	UPB02	7.0	4.58	51				
BBT3	9.0	3.96	25	UPB07b	8.5	4.66	52				
GRB_C	2.5	3.99	26	RWU03b	7.0	4.67	53				
PRC08b	3.5	4.07	27	EPN05	13.8	4.74	54				

Table 7-2. Ordering and ranks of spatial survey stations by increasing long-term mean oxygen concentration, in two groups: deeper/shallower stations, where the median of the deepest-sampled depths is greater than or less than 4m.

Column 1: Station
 Column 2: Average (all years, all surveys) deepest-sampled bottom oxygen concentration [mg L⁻¹]
 Column 3: Deeper group: Rank within 57 stations in deeper group
 Shallower group: 57 + rank within 20 stations in shallower group

Deeper			Shallower					
			BH05	4.56	39	UPB12	5.89	76
PRC07	1.19	1	WPS04	4.58	40	WPS08	6.22	77
PRN03b	2.12	2	UPB02	4.58	41			
PRN07	2.17	3	UPB07b	4.66	42			
PRN01	2.45	4	RWU03b	4.67	43			
SR02	2.58	5	EPN05	4.74	44			
PRC05b	2.85	6	WPS11	4.77	45			
PRN08	2.86	7	WPS13	4.77	46			
SR03	2.88	8	BH07	4.79	47			
GRB04	2.96	9	WPS09	4.80	48			
PRC02	3.17	10	EPN03	4.87	49			
BRT3	3.36	11	WPS14	5.04	50			
PRS07	3.41	12	RWU03	5.04	51			
BRT4	3.48	13	EPN04	5.06	52			
BRT2	3.52	14	EPN02	5.20	53			
BBT1	3.58	15	WPS01	5.43	54			
UPB13	3.66	16	RWU05	5.50	55			
UPB03	3.70	17	WPS02	5.54	56			
PRS03	3.72	18	EPS01	5.81	57			
GRB07	3.88	19						
GRB06	3.90	20						
BBT3	3.96	21	SR01	2.37	58			
UPB10	4.07	22	GRB05	3.07	59			
WPS10	4.08	23	GRB02	3.59	60			
UPB09b	4.08	24	GRB01	3.79	61			
GRB09	4.13	25	GRB_C	3.99	62			
UPB11	4.15	26	PRC08b	4.07	63			
BBT2	4.15	27	GRB_A	4.08	64			
BRT1	4.20	28	GRB03	4.11	65			
UPB08	4.27	29	GRB_D	4.21	66			
WPS07	4.29	30	GRB_B	4.53	67			
UPB04b	4.29	31	PRT02	4.82	68			
UPB01	4.30	32	PRT01	5.07	69			
WPS05	4.31	33	PRS04	5.10	70			
BH10	4.33	34	GRB10	5.11	71			
UPB06	4.38	35	PRT04	5.13	72			
UPB05	4.39	36	PRT03	5.31	73			
EPN08b	4.49	37	BRT6	5.46	74			
EPN06	4.53	38	GRB12	5.60	75			

Table 7-3. Regression r^2 between daily-mean fixed site bottom oxygen concentration and individual survey percent hypoxic area, corresponding to Figure 7-11. Bold values have $p < 0.05$.

Station	1.4 mg L ⁻¹	2.9 mg L ⁻¹	4.8 mg L ⁻¹
PD	0.0193	0.0957	0.2117
BR	0.2724	0.5183	0.6695
CP	0.2390	0.4743	0.6809
NP	0.3115	0.5402	0.5510
GB	0.1666	0.2809	0.2799
SR	0.1649	0.3352	0.3872
MV	0.3091	0.4978	0.4634
QP	0.2575	0.4462	0.5263
PP	0.0747	0.1751	0.3090
TW	0.2678	0.3425	0.3563
MH	0.1396	0.3365	0.5340

Table 7-4. Regression r^2 between seasonal Hypoxia Index from fixed site time series and seasonally-averaged percent hypoxic area from spatial surveys, corresponding to Figure 7-13. Bold values have $p < 0.05$. Missing values indicate the regression could not be computed because there were no non-zero values of percent hypoxic area.

Station	1.4 mg L ⁻¹	2.9 mg L ⁻¹	4.8 mg L ⁻¹
PD	0.3300	0.0300	0.0000
BR	0.1400	0.8300	0.7900
CP	0.0000	0.7800	0.6200
NP	0.8700	0.6500	0.4900
GB	0.7900	0.7000	0.5700
SR	0.2600	0.7400	0.8700
MV	0.5600	0.0400	0.1200
QP			0.0400
PP	0.5500	0.5200	0.2700
TW		0.1800	0.4500
MH	0.0000	0.1600	0.4500

Figure 7-1. Spatial surveys station locations.

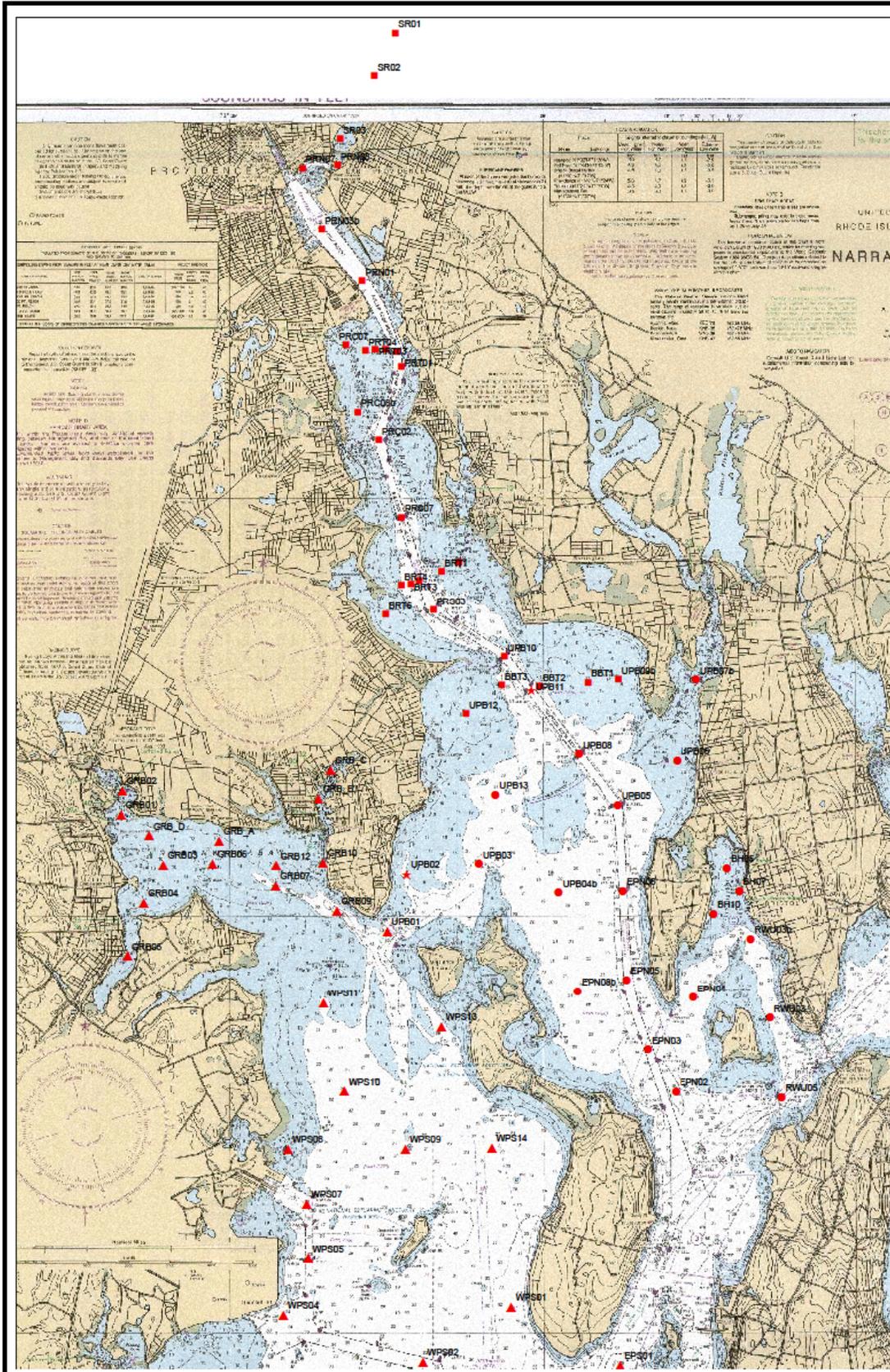


Figure 7-2. Rank of mean near-bottom DO concentration at spatial survey stations.

Lower rank is lower DO concentration. a. All 77 stations (see Table 7-1). b. Subgroup of 57 stations with median near-bottom sample depth greater than 4m (see Table 7-2).

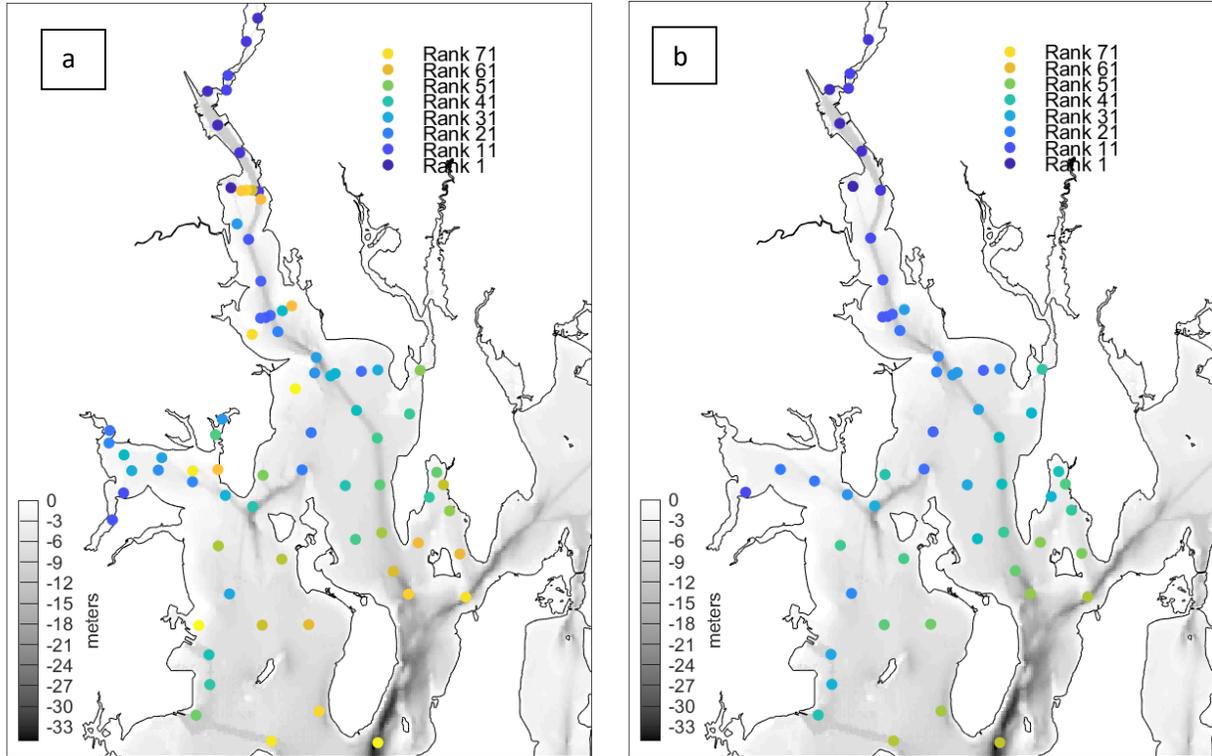


Figure 7-3. Spatial survey correlation results.

(a) Reference stations SR02, PRN01, PRS07, UPB10.

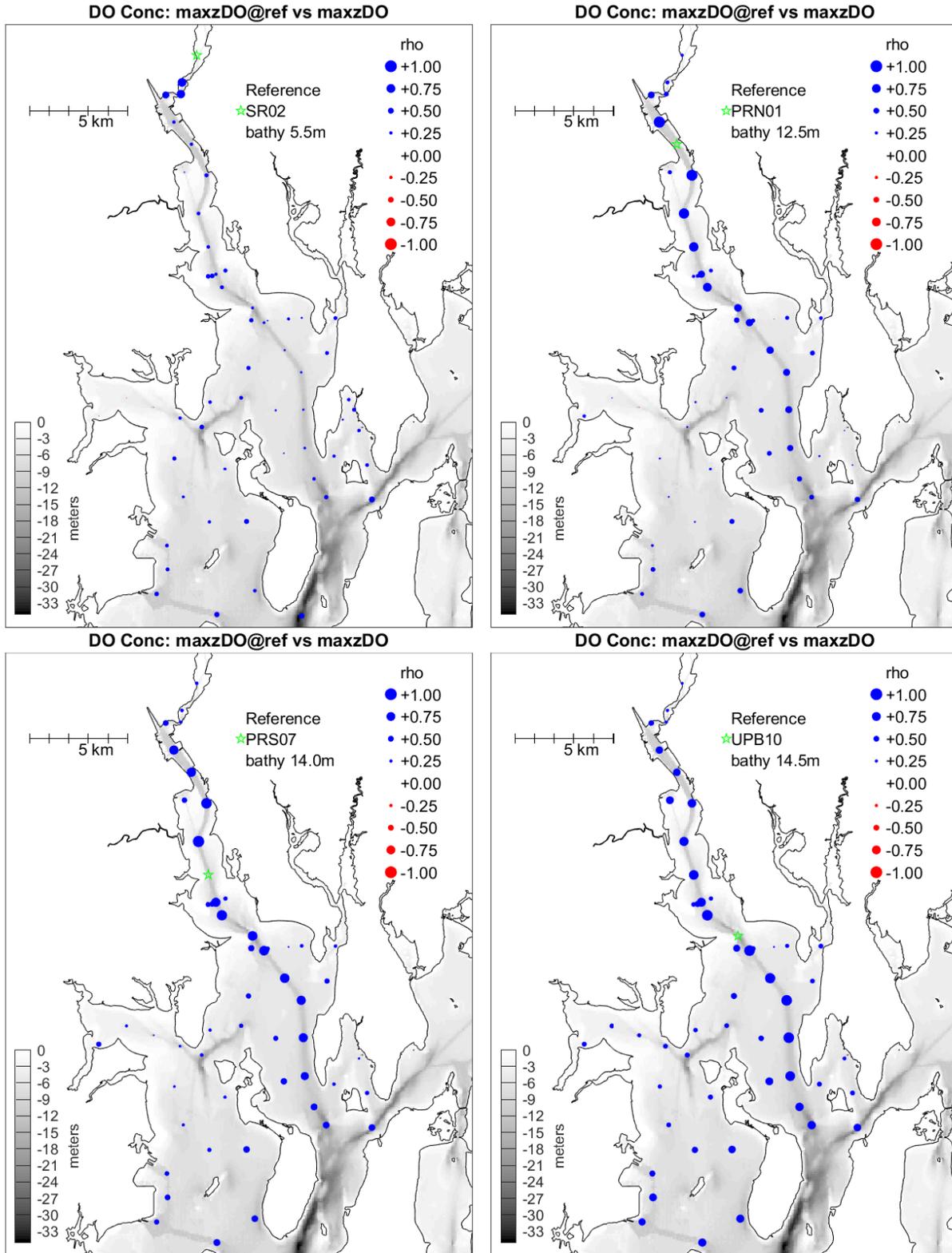


Figure 7-3 (b) Reference stations UPB08, GRB09, GRB04, WPS09.

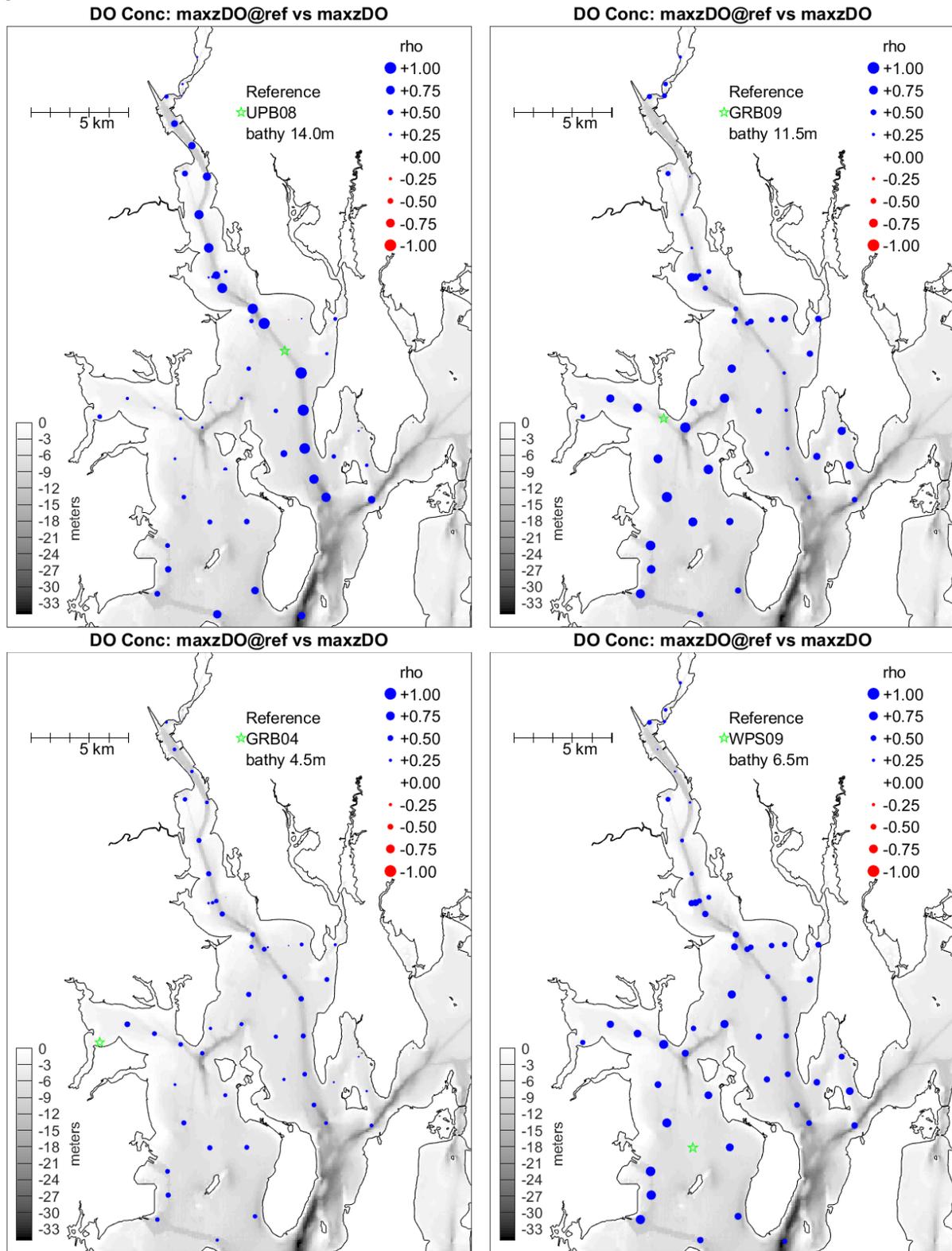


Figure 7-3. (c) Reference stations UPB06, EPN06, EPS01, RWU05.

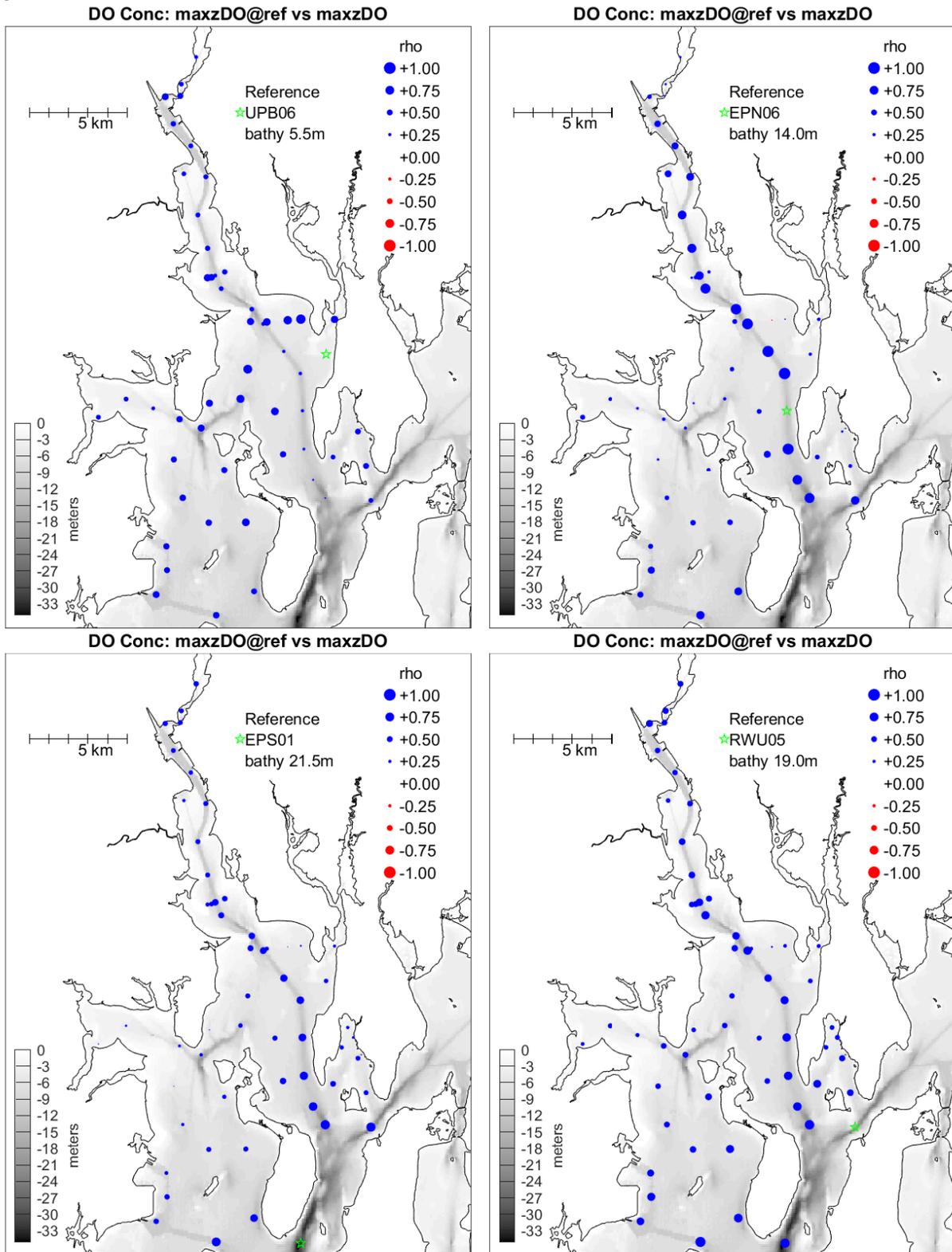


Figure 7-4. Spatial surveys, decorrelation length scales.

Correlation of temporal variability (all surveys, all years) at 12 reference spatial survey stations (listed in legend, with their bathymetric depth) to all other spatial survey stations, as a function of distance between stations in the pair. At zero distance all values are 1. Solid symbols are station pairs for which the difference in bathymetric depths is less than 5 m. Lines connect solid symbols for stations UPB10 and GRB04, which are representative of deeper channel locations and shallower embayment stations, respectively.

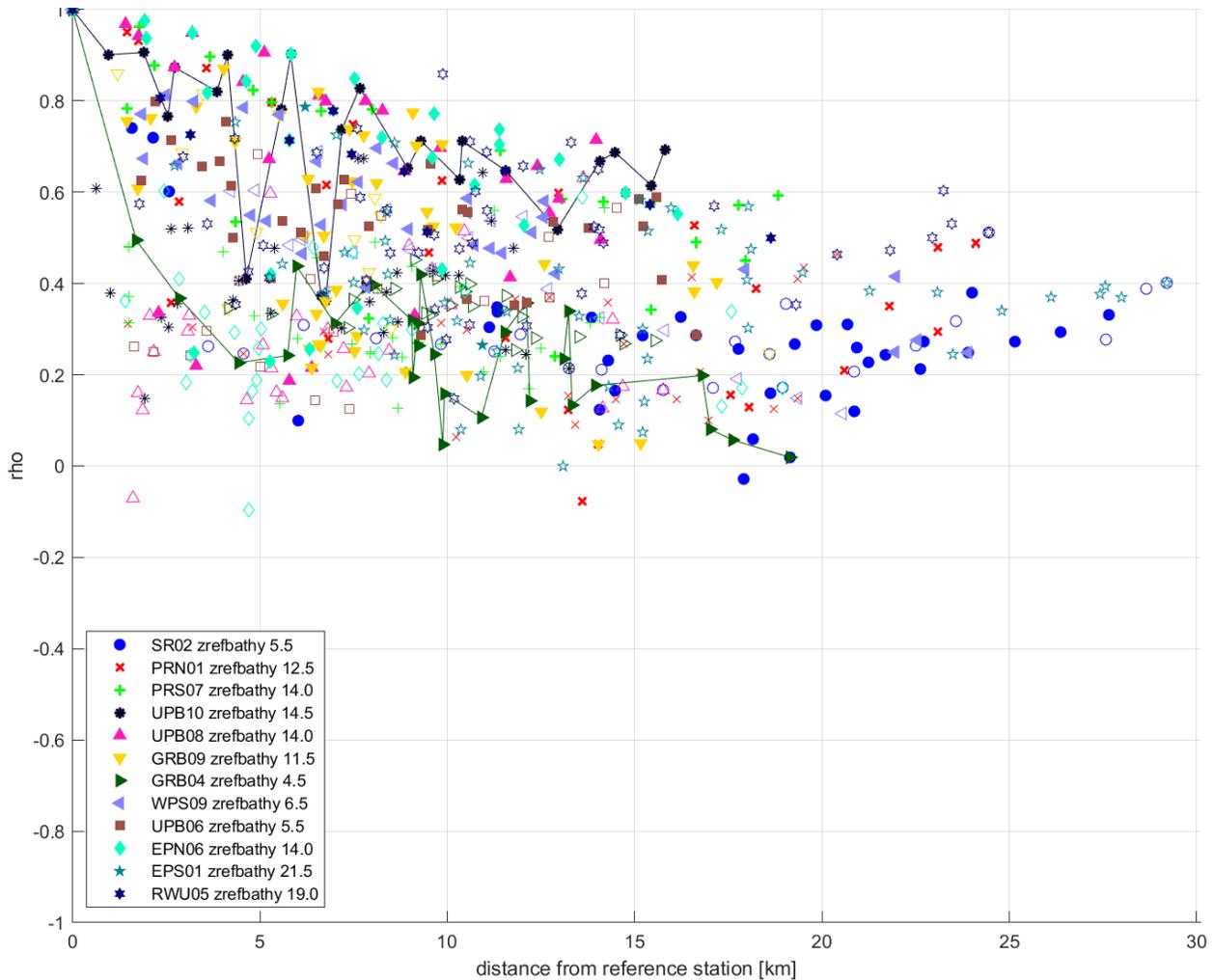


Figure 7-5. Same correlation results as in Figure 7-4 but ordered on the horizontal axis by relative rank (shown in Figure 7-2b) of station near-bottom DO using deeper 57 stations.

(Solid symbols are again station pairs for which the difference in bathymetric depths is less than 5 m.) The horizontal axis is the difference between the near-bottom DO rank of each of the other 56 stations and the near-bottom DO rank of the reference station. For example, reference station SR02 (blue circles) has near-bottom DO rank of 5 (Table 7-1) so there are 4 results that are on the negative portion of the horizontal axis; the station with the most extreme near-bottom DO rank of 57 (WPS08) appears farthest to the right at 52 (its rank difference relative to that of SR02) on the horizontal axis.

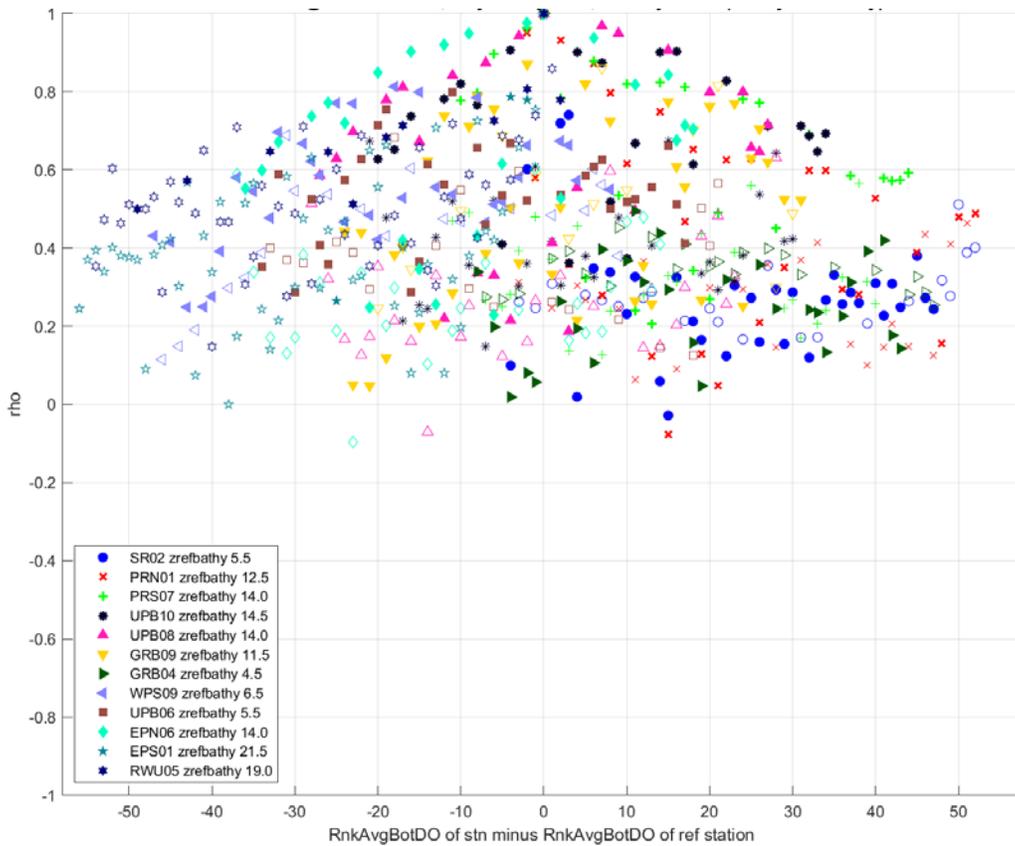


Figure 7-6. Decorrelation time scale for deep DO concentration.

From lagged autocorrelations of time series measurements at individual fixed site locations.

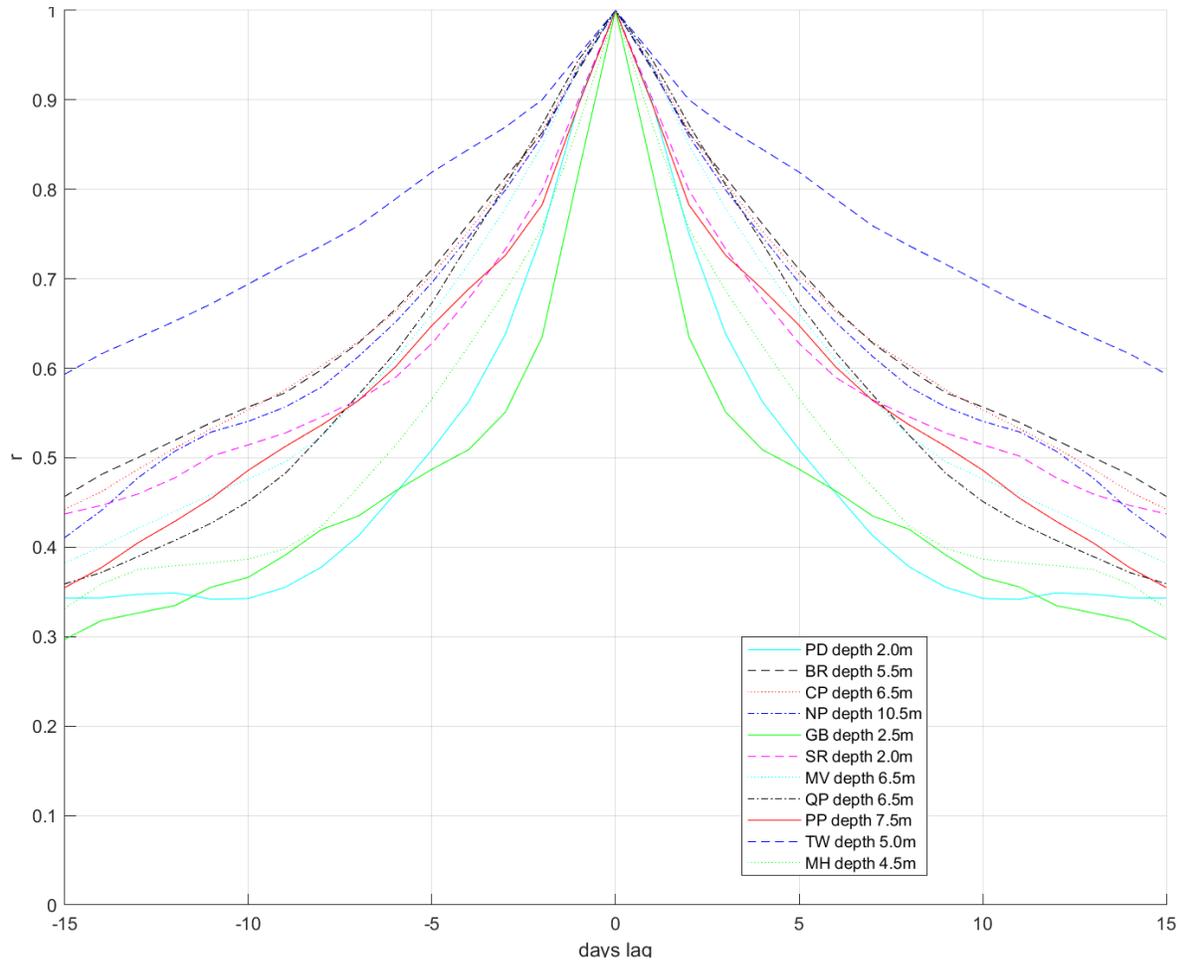


Figure 7-7. Correlations of spatial survey near-bottom DO concentration and daily mean of 15-minute temporal resolution time series on date of spatial survey.

(a) Fixed site stations PD, BR, CP, NP.

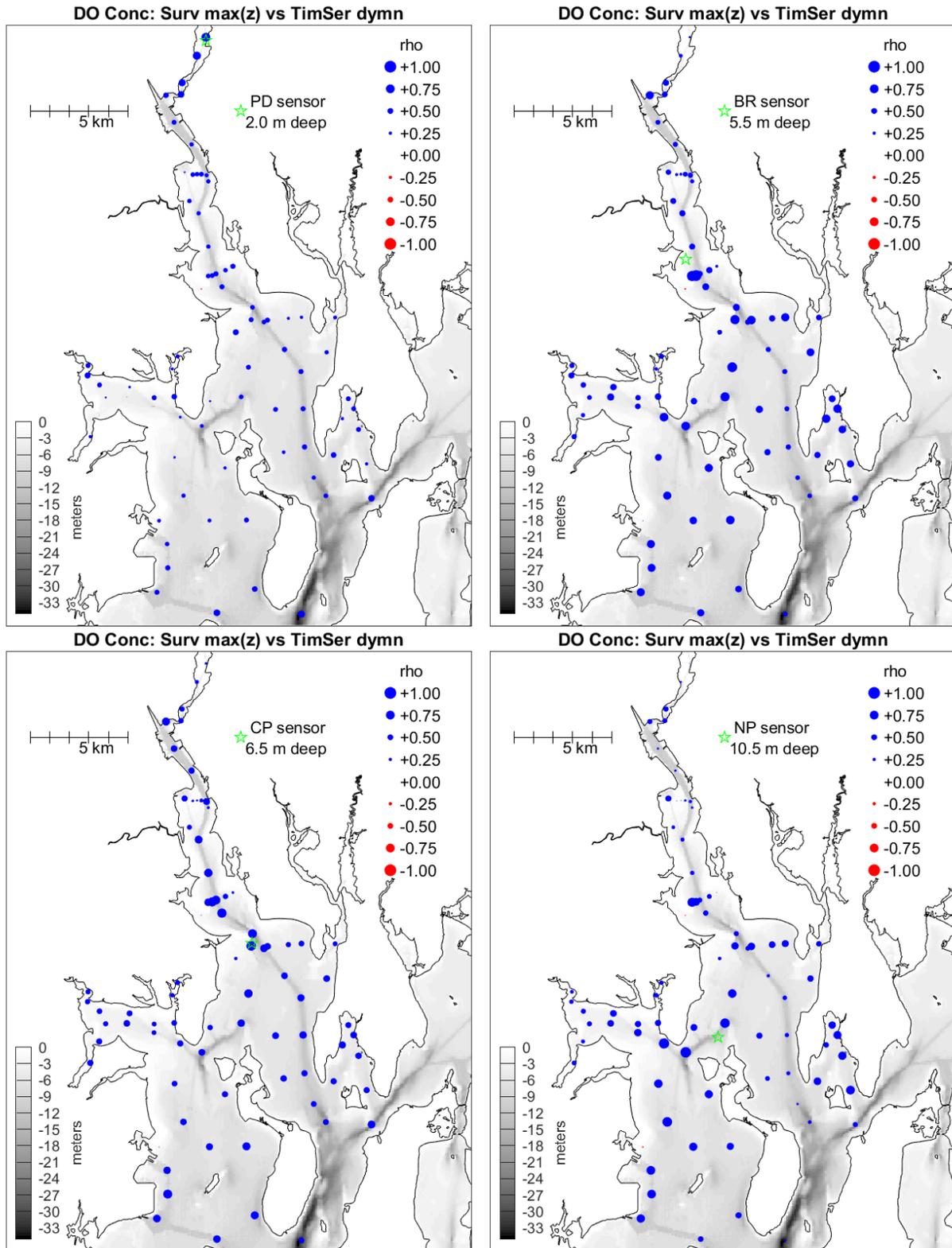


Figure 7-7. (b) Fixed site stations GB, SR, MV, QP.

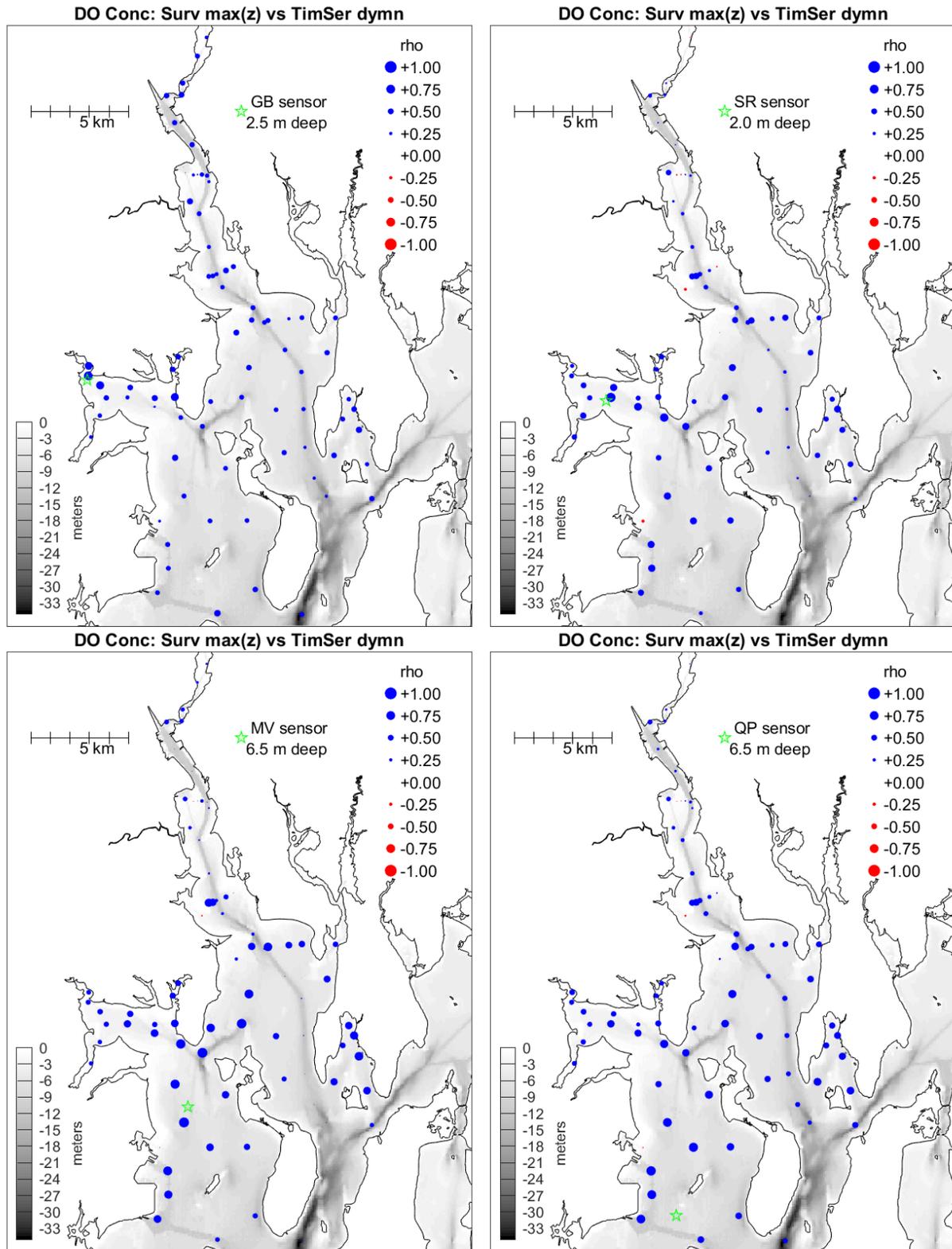


Figure 7-7. (c) Fixed site stations PP, TW, MH.

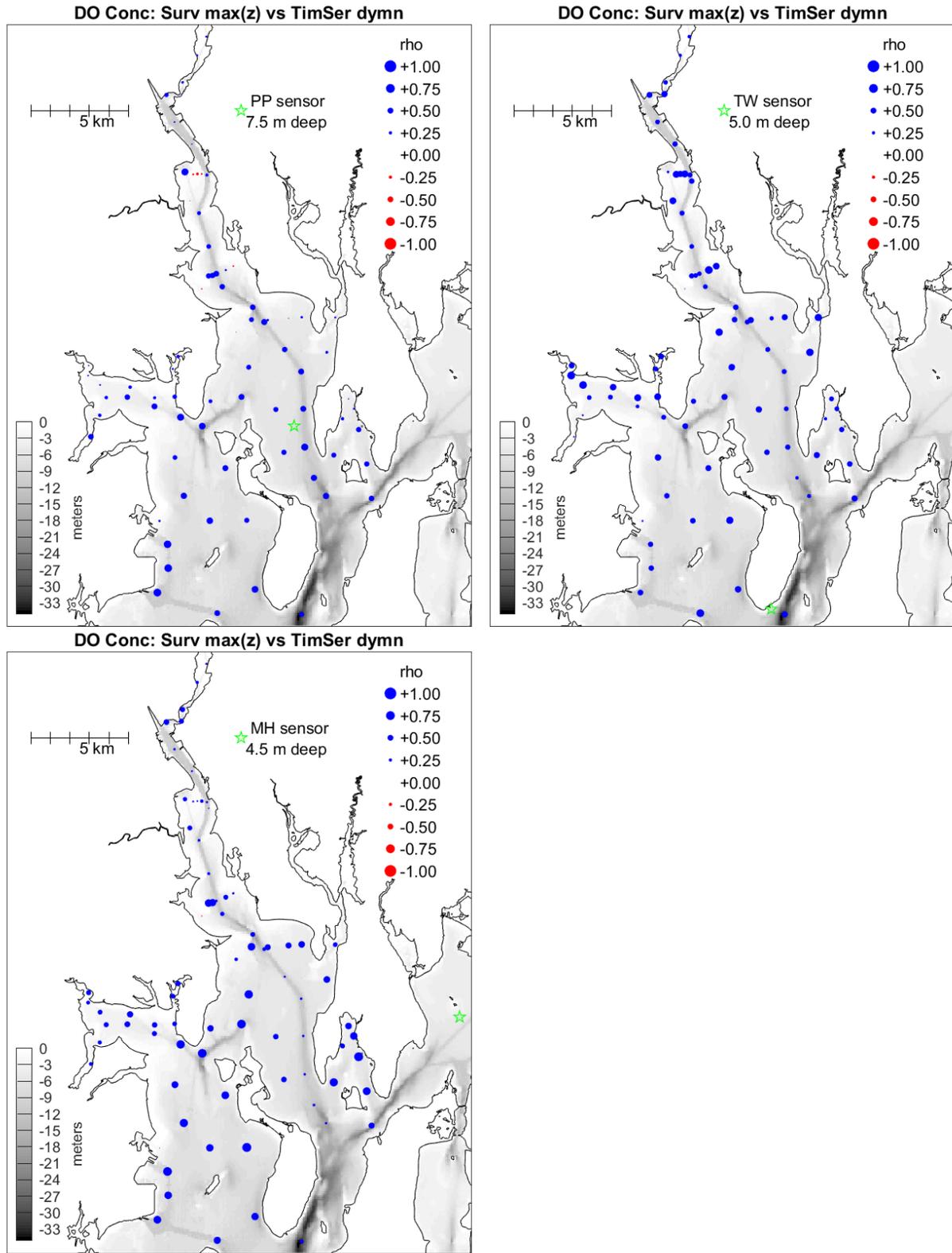


Figure 7-8. Correlation matrix between spatial surveys and time series oxygen.

Left: Matrix of rho values from results (shown in Figure 7-7 for correlation between spatial survey near-bottom DO concentration and daily-mean of time series at fixed site station on the date of each spatial survey). Vertical axis is spatial survey station rank (Figure 7-2), by increasing (1 = lowest) all-survey mean near-bottom DO concentration, with deeper stations from 1 to 57 and shallower stations from 58 to 77 (Table 7-2). Horizontal axis is fixed site station, in order shown in Figure 7-7 (PD, BR, CP, NP, GB, SR, MV, QP, PP, TW, MH). Right: Difference from matrix at left, when hourly mean of time series, for hour centered on time of each spatial survey profile, is used instead of daily mean.

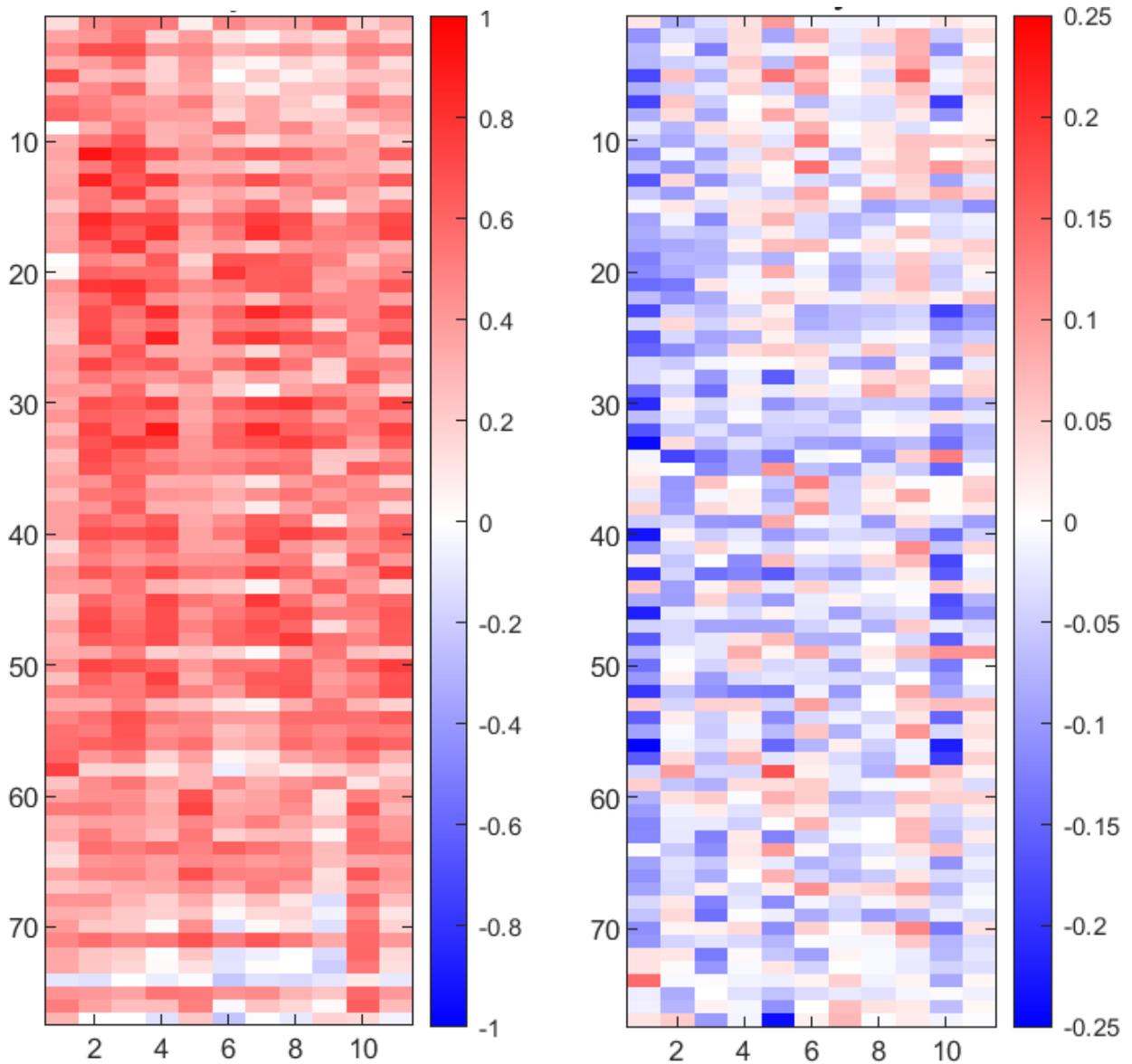


Figure 7-9. Lagged correlations between spatial survey and time series oxygen.

Bullocks Reach (BR) site (5.5 m sensor depth) daily-mean time series measurements and spatial survey measurements, near-bottom DO concentration. Correlation at each lag (positive or negative) is computed independently. Twelve representative spatial survey stations are included, as listed in the legend with their median near-bottom sample depth and distance in km from BR.

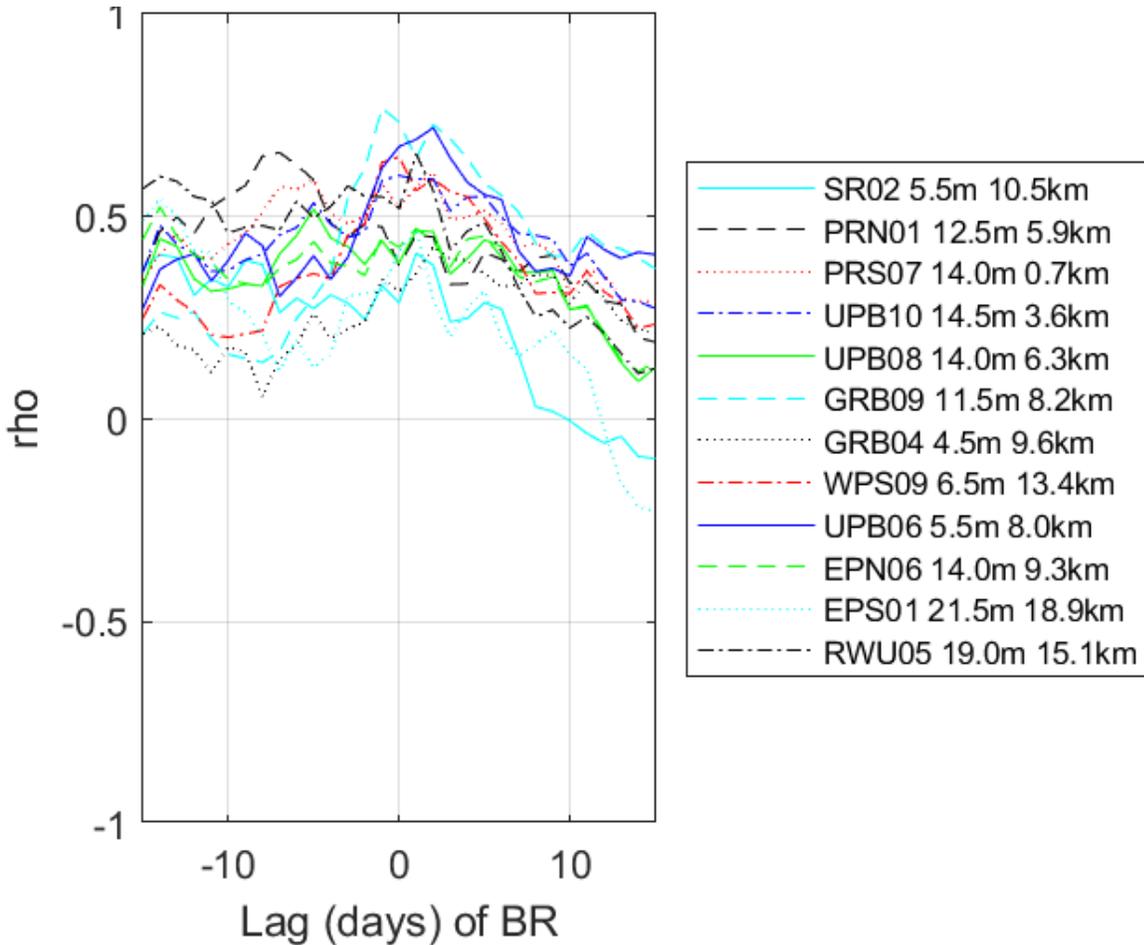


Figure 7-10. Percent hypoxic area from individual spatial survey dates, DO thresholds 4.8 mg L⁻¹ (green), 2.9 mg L⁻¹ (blue), and 1.4 mg L⁻¹ (red).

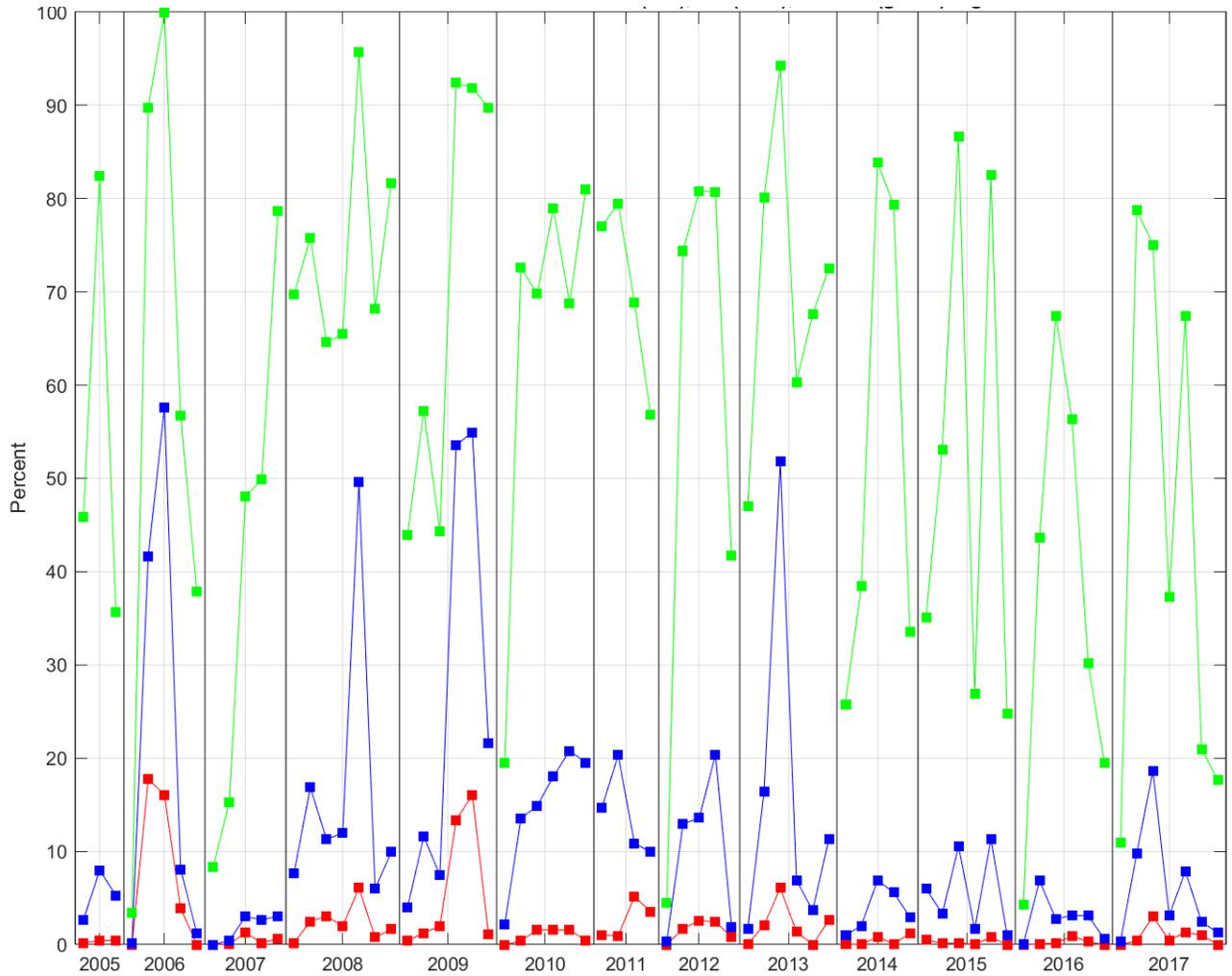


Figure 7-11. Percent hypoxic area as function of time series oxygen.

Percent hypoxic area on individual survey dates relative to the 1.4 (red), 2.9 (blue), and 4.8 (green) mg L⁻¹ thresholds, plotted as a function of daily-mean deep DO time series concentration at each of the 11 fixed site stations on the dates of the surveys. Correlation coefficient results are in Table 7-3.

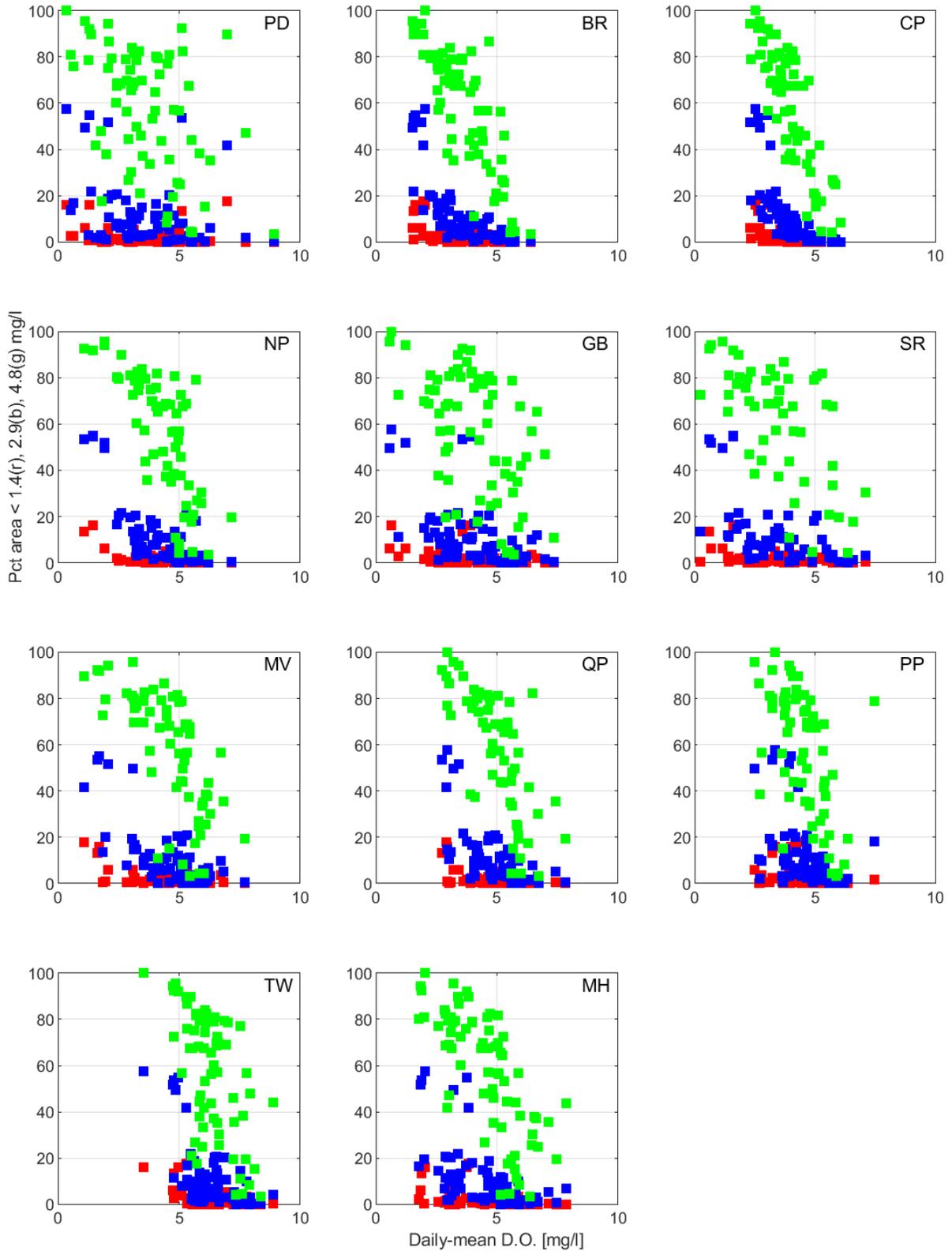


Figure 7-12. Same as Figure 7-11 but horizontal axis is the daily deficit-duration relative to 1.4 (red), 2.9 (blue), and 4.8 (green) mg L⁻¹ threshold.

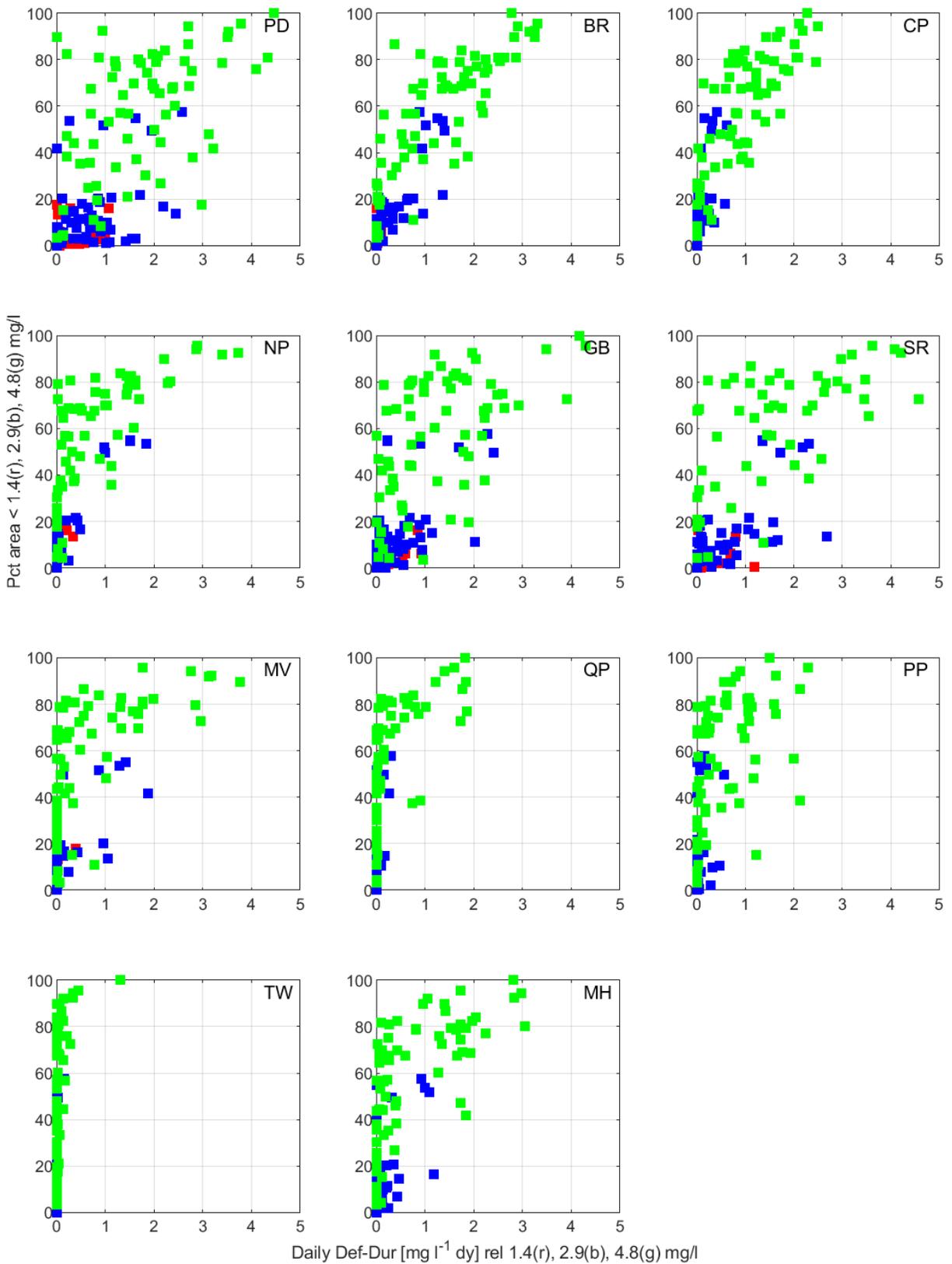


Figure 7-13. Seasonal Hypoxia Index from time series vs seasonally-averaged percent hypoxia area from spatial surveys.

Red symbols show Hypoxia Index relative to 1.4 mg L⁻¹ threshold, in mg l⁻¹ day normalized by 12 mg l⁻¹ day, vs seasonal-mean percent hypoxic area relative to 1.4 mg L⁻¹ threshold, normalized by 10%. Blue symbols show same, but relative to 2.9 mg L⁻¹ threshold, and normalized by 70 mg l⁻¹ day and 30% respectively. Green symbols show same, but relative to 4.8 mg L⁻¹ threshold, and normalized by 210 mg l⁻¹ day and 80% respectively. Correlation coefficient results are in Table 7-4.

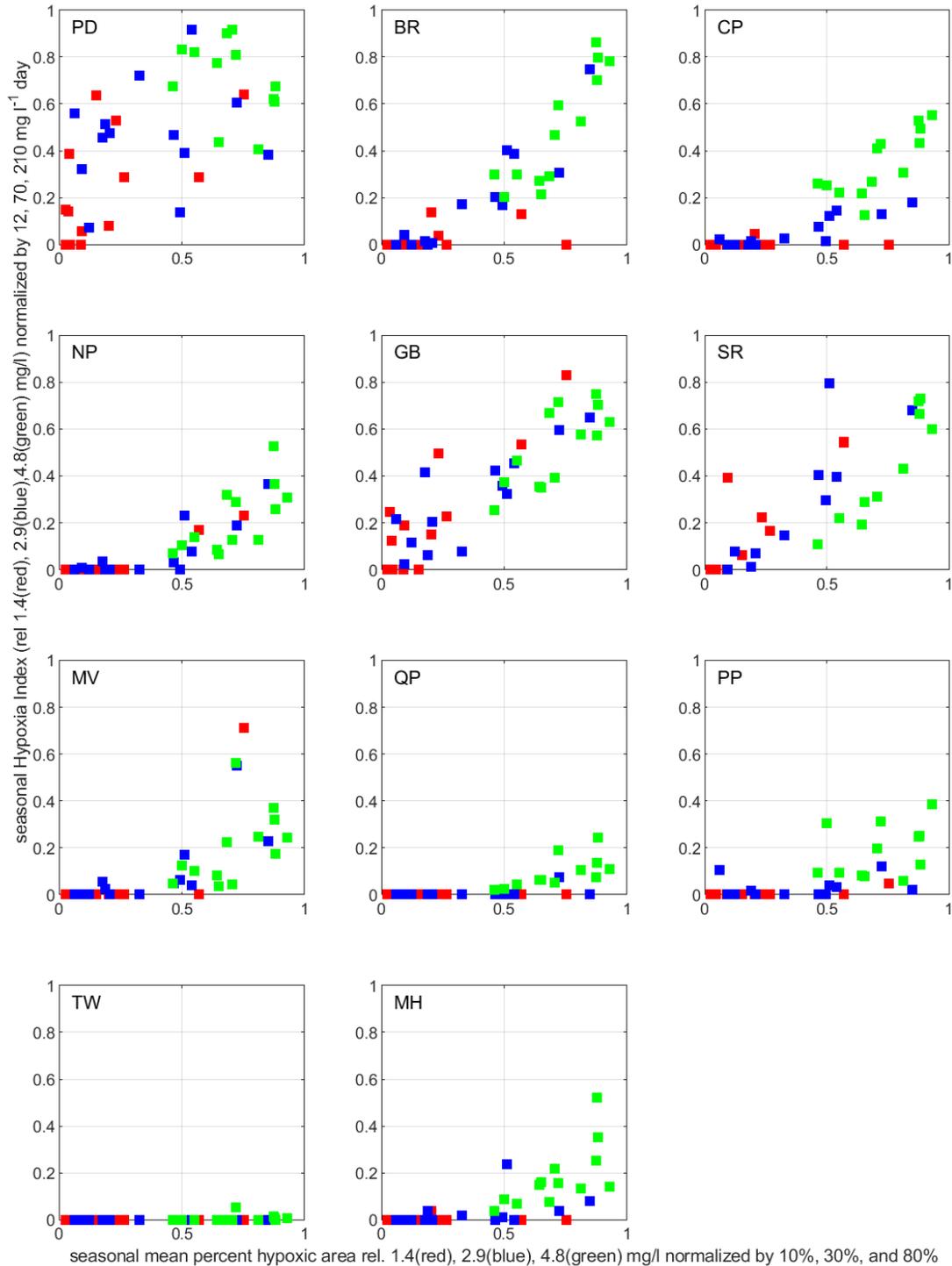


Figure 7-14. Histograms to gauge representativeness of spatial survey dates.

Daily-mean near-bottom DO concentration (horizontal axis) from fixed site time series using all years, date range from the earliest spatial survey to the latest spatial survey: all dates sampled in time series (blue, left vertical axis) and only dates sampled by spatial surveys (orange, right vertical axis). Vertical axes are number of daily-mean values in each bin of DO ranges. Results of Kolmogorov-Smirnov test shown in upper right below station code.

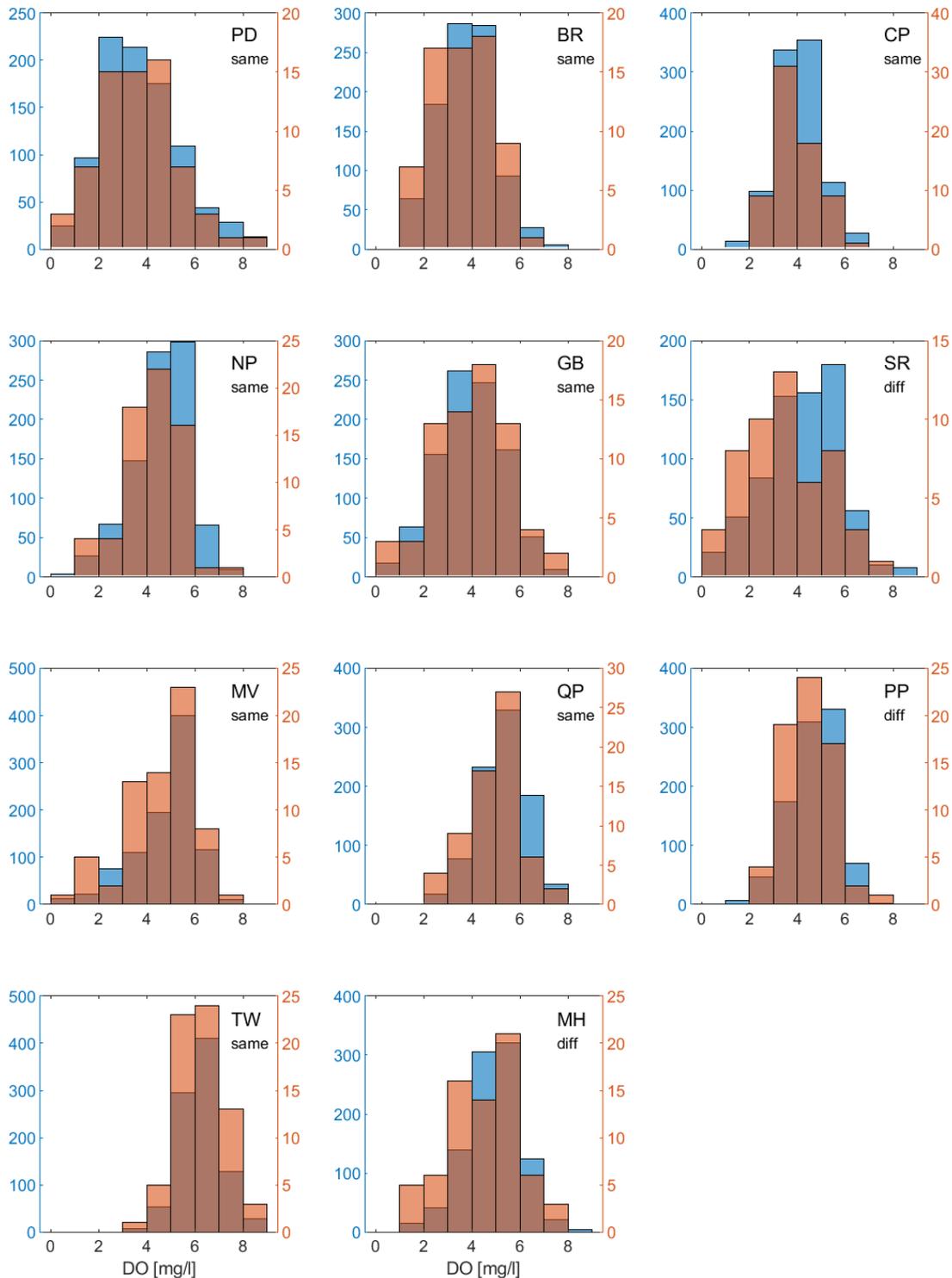
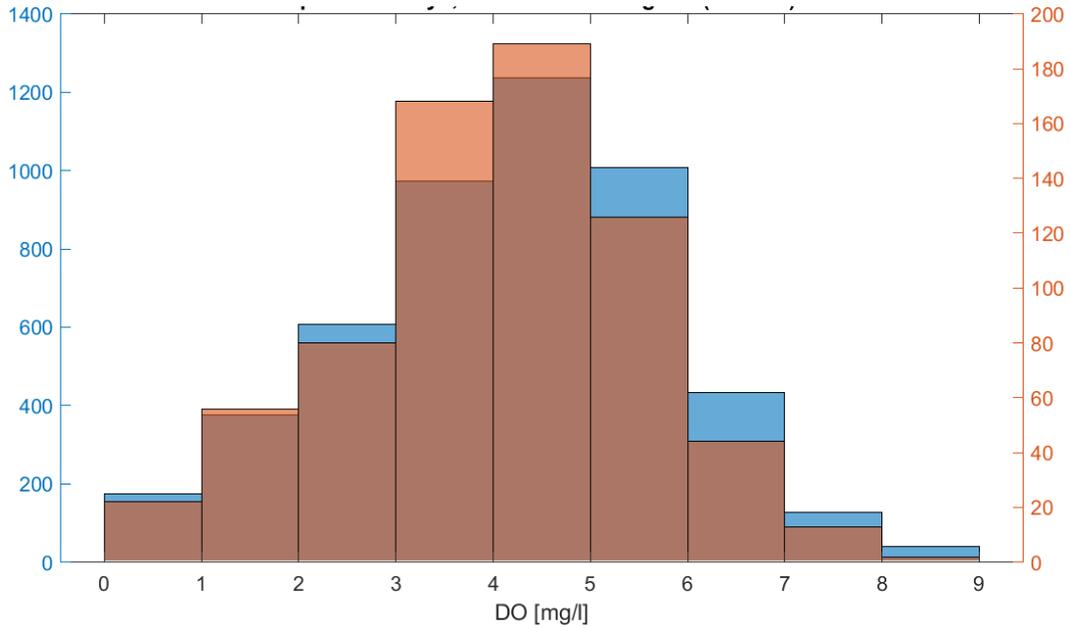


Figure 7-15. Histograms to gauge representativeness of time series locations.

Near-bottom DO concentration (horizontal axis) from spatial surveys using all years: all stations sampled (blue, left vertical axis) and only the 11 stations nearest to the fixed site station locations (orange, right vertical axis). Vertical axes are number of in each bin of DO ranges.



Task 8 Toward Improved Chlorophyll Time Series Using Grab Samples

8.1 Scope

Objective. Help facilitate improved quality of NBFSMN chlorophyll time series. **Methodology.** NBFSMN data undergo quality control and pre-deployment calibrations as documented in the QAPP. Extensive grab sample datasets have also been collected by NBFSMN operators and facilitate potential post-deployment recalibration. Comparisons of the grab samples and time series will be used to assess potential improvements to the time series and make it possible for them to be applied if justified.

Expected outputs and outcomes. Colleagues from URI/GSO and NBC who collected the grab samples had begun this process already. In this task, the goal was to help advance their work.

Note about task scope adjustment. As originally conceived this task was to explore potential for post-deployment recalibration of NBFSMN time series using the spatial surveys dataset. However, the grab sample dataset is better-suited for this purpose than the spatial surveys. The key people at labs that collect and analyze the grab samples were consulted and were willing to provide their expertise and datasets, so with agreement from NBEP and NEIWPC the scope was adjusted.

8.2 Methods

Grab samples. “Grab sample” refers to *chlorophyll a* concentration, from analysis by the acetone-extraction method, on a physical water sample collected in the field and brought to the laboratory (e.g., SNBIW Chapter 16). For simplicity the term “chlorophyll” is used here when referring to *chlorophyll a*.

Spreadsheets with grab sample results from the NBFSMN site locations through 2017 were obtained for PD and BR from NBC (Eliza Moore, Molly Welsh) and for all other stations from URI/GSO (Heather Stoffel), including TW for which the data are from NBNERR. The grab sample data span all sampled years up through 2017, with few exceptions, at all stations except PD and BR, where samples are all from 2011-2017 except a small number from 2005 at PD. Years 2005 to 2010 at PD and BR were sampled and the results are still being compiled.

Only grab samples for the near-surface sondes were included in the analysis. NBC included only good quality (unqualified) results in their spreadsheet. Screening was applied to omit results that had remarks in the comments and/or notes columns (URI/GSO stations), and that were suspect or had errors in the QC codes (TW station; qualifier codes 0, 4, and 5 kept and all others omitted, see for example Durant and Raposa, 2010).

Up to three individual *chlorophyll a* concentration results (field duplicates or lab replicates) were available for each date/location of grab sampling. At TW multiple values are different field duplicates, at PD and BR multiple values are either field duplicates or lab replicates, and at other stations multiple values are lab replicates.

The grab results were loaded in to Matlab and manipulated to a single table for all 11 stations containing the individual values, and for cases with more than one value some basic statistics were tabulated including their mean, standard deviation, and difference between the maximum and minimum.

Time series. The time series measurements are *chlorophyll fluorescence* from sonde sensors (YSI, 2012). Sondes are typically deployed for between 2 and 4 weeks at a time unattended and record a value each 15 minutes. One-point (for zero concentration) pre-deployment calibrations are done using deionized water, following the QAPP (RIDEM, 2014), and post-deployment recalibrations have not yet been implemented. Values are recorded with units of $\mu\text{g L}^{-1}$ and the sonde manufacturer indicates the sensors measure *total chlorophyll*. However a common practice for *chlorophyll fluorescence* is to do post-deployment recalibrations using grab samples, as is being pursued here. In this approach, because the grab samples are *chlorophyll a* (not *total chlorophyll*), after the recalibrations are applied the sonde results are taken to be *chlorophyll a*. This is typically acceptable because *chlorophyll fluorescence* is known to be an approximate measure of actual chlorophyll concentrations, which varies widely depending on a number of factors including phytoplankton physiology. The assumption inherent in this approach to recalibration is that this measurement variability is of comparable magnitude to typical differences between *chlorophyll a* and *total chlorophyll*. Confirming that this is the case, for bay waters, would help justify this approach.

It is important to bear in mind that a sonde samples a much smaller volume of water than is treated by a grab sample. This sets an important limitation on agreement between sonde and grab sample results, even when the location, depth, and time of the of the sonde measurement perfectly matches that of the grab sample.

The date/time of each grab sample was used to obtain, and add to the table, the time series results from the corresponding sonde. The nearest individual 15-minute resolution value from the sonde, as well as the minimum, mean, and maximum of the values within 1 hour, 3 hour, and 8 hour periods centered on the time of the grab sample were tabulated.

The sonde chlorophyll sensor specification from the manufacturer indicates it has range up to “approximately $400 \mu\text{g L}^{-1}$ ” (YSI, 2012). There are two sonde values in the table higher than $400 \mu\text{g L}^{-1}$ (557.6 and $541.9 \mu\text{g L}^{-1}$). They were not excluded from the analysis.

See page 155 for information about supporting code and data files, and page 150 for how to use them.

8.3 Results

Grab results alone. The number of grab sample results ranges from 87 to 130 at all stations except PD and BR, for which there are 175 and 162 respectively (Table 8-1). For PD and BR on most grab sample dates there is a single grab sample value, with a smaller number of cases having two values and none with three values; about half of the cases with two values are field duplicates and about half are lab replicates. For all other stations except TW, on most grab sample dates there are 3 values, with a small number having 2 values and fewer with a single value; these are all analytical replicates. For TW all cases have two values, which are field duplicates.

The minimum, average, and maximum values of the mean of the (up to 3) results for each grab span most of the range of concentration values that occur at each station in the time series (shown below). Comparisons to time series are therefore expected to be representative of their range.

For standard deviation among the 2 or 3 grab sample values, the means across all grabs are between 1 and $2 \mu\text{g L}^{-1}$ for most stations; they are less than $1 \mu\text{g L}^{-1}$ at a stations QP and TW where chlorophyll

levels are generally lower, and more than $2 \mu\text{g L}^{-1}$ where chlorophyll levels are generally higher (northern stations NP and GB). The PD and BR stations results may not be representative because they are based on a small fraction of cases for which there are multiple values. This likely explains why PD, the station with the highest chlorophyll levels, has the lowest mean standard deviation in contrast to overall pattern just described. As expected, results for the minimum-maximum difference, among the 2 or 3 grab sample values, are similar to, but higher than, results for the standard deviation.

Grab results with time series. Plots showing mid-May to mid-Oct for all years of time series measurements, with the individual grab results superposed, enable coarse visual inspection (Figure 8-1 to Figure 8-11). The presentation with all years on a single page does not facilitate a detailed quantitative comparison but serves the purpose of generally confirming reasonably good agreement between the time series and the grabs. There is no obvious or pronounced bias between the two datasets and the number of cases where there is notable mismatch between grab results and the time series is generally small. Some high chlorophyll bloom events are captured by grab samples that agree with the time series (for example, mid-September 2016 at NP). At GB, and to some extent PD, there are at least a few high chlorophyll grab samples for which the time series data are not elevated.

Scatter plots showing sonde time series results on the vertical axis and grab sample results on the horizontal axis, with a diagonal line superposed to indicate where data with perfect agreement between sonde and grab would fall, are useful to assess potential for post-deployment calibrations. Results for PD are shown in are shown Figure 8-12. The left column of frames uses linear axes and the right column of frames shows the same results on log-log axes. The top row of frames shows sonde results from the nearest individual 15-min timestamp to when the grab was collected, with each of the up to 3 individual grab sample results shown. The horizontal errorbars show the range of variability in grab samples (some field duplicates and some lab replicates). The lower three rows of frames show sonde results from the 1-hour, 3-hour, and 8-hour intervals centered on the time of the grab sample. The horizontal lines are located vertically at the mean of the sonde results during each period. The vertical errorbars show the range (minimum to maximum) of temporal variability in the sonde results (zero for the individual 15-min values, nonzero for the 1-hour, 3-hour, and 8-hour interval results).

The differences in frames from row to row in the figure quantify the extent to which temporal variability in sonde data increases for longer periods. The temporal variability during a one hour period (four individual 15-min values) is non-trivial, roughly comparable to the variability in grab samples. This underscores the importance of accurately recording the times when grab samples are collected, so they can be matched with the proper sonde measurement. Temporal variability of sonde measurements over 3-hour and 8-hour periods increases substantially relative to the 1-hour results, suggesting that when comparing time series chlorophyll to grab sample it is most appropriate to use individual 15-min. Hourly-mean sonde data is less suited but perhaps comparably appropriate as 15-min values, and sonde values averaged over periods longer than an hour are much less appropriate.

The same scatterplots for the other 10 stations are shown in Figure 8-13 and Figure 8-14, for the 15-minute sonde values only. The horizontal error bars for TW, due to variations between field duplicates, are comparable to error bars for stations other than PD and BR, due to variations between lab replicates. This indicates that field collection of grab samples contributes to variability in their results at a comparable level as laboratory analysis methods. The PD and BR horizontal errorbars, due to variability in field replicates in some cases and lab replicates in some cases, are consistent with this

conclusion. The total variability in grab samples, due to combined influences of field sample collection and laboratory analysis methods, is therefore modestly larger than the horizontal errorbars on the figures.

Overall, at all stations there is a relationship with higher sonde values corresponding to higher grab samples, as expected. By qualitative visual inspection the relationship is strongest at PD, GB, and BR stations, where the values span the largest ranges. The relationship is notably weaker at TW, where the values span the smallest range. At most stations there is a perceptible pattern in which at the upper end of the range the sonde results have lower values than grab results, and at the lower end of the range they are higher than grab results. This pattern seems strongest at stations with a high range of values, such as PD, GB, and to some extent BR.

There are multiple methods by which the sonde data could be recalibrated. If it was possible, perhaps one of the best approaches would be to determine the needed recalibration using only the results from grab samples during each individual 2-4 week deployment period, from the site of the deployment, and apply it only to that particular deployment at that site. Despite that this dataset has an extensive number of grab samples, they have likely not been collected frequently enough to make this approach successful. For some deployments there may be no grab results available and for others there are grab results from typically at most up to 2 times per deployment. The inherent variability of individual sonde and grab results means that a recalibration based on such a small number of sonde-grab pairs would likely introduce a high degree of inaccuracy in the recalibrated records.

A more common method to determine and apply needed recalibrations is to group sonde-grab pairs together across multiple deployments, assume a linear regression relationship applies, compute the regression and apply it to sonde results from all deployments. This approach can be taken using all deployments from all sites, or using deployments only from a given site. Using regressions based on deployments only from a given site is more appropriate, if there is a large difference in the range of values from site to site as is the case with NBFSMN stations. (In addition, the chlorophyll sensor on a sonde is removable; deployments having the same individual chlorophyll sensor could be identified and used. This would make sense if a sufficient number of grabs was available, and if there was reason to believe that the regression was expected to vary from sensor to sensor.) A regression against the grab results could be computed using all 15-min sonde values from each individual station (as shown in upper row of Figure 8-12, and in Figure 8-13 and Figure 8-14). If the regression result is statistically meaningful and judged to justify recalibration, it could be used to recalibrate all sonde data from that station.

It may be that the statistics of the regression are not sufficiently strong, and/or the recalibrations (once computed and examined) have magnitude or other characteristics at odds with the judgement of experts familiar with these datasets. In this case, the decision not to recalibrate is justified, and the comparisons presented here can serve as quantitative demonstration of the sonde-grab relationship deemed to be acceptable without recalibration.

Table 8-1. Basic statistics of grab sample datasets in the provided spreadsheets.

	PD	BR	CP	NP	GB	SR	MV	QP	PP	MH	TW
DateSt	04-Aug-2005	16-Mar-2011	01-Jun-2006	01-Jun-2006	23-Jun-2006	08-Jul-2008	01-Jun-2006	01-Jun-2006	01-Jun-2006	01-Jun-2006	31-Jan-2003
DateFn	20-Dec-2017	20-Dec-2017	18-Oct-2017	18-Oct-2017	27-Oct-2017	27-Oct-2017	18-Oct-2017	26-Sep-2017	18-Oct-2017	18-Oct-2017	13-Dec-2016
nTot	175	162	100	96	130	87	95	95	106	99	124
n3val	0	0	89	89	111	78	87	87	92	91	0
n2val	16	45	9	5	18	6	5	5	13	6	124
n1val	158	117	2	2	1	3	3	3	1	2	0
min(avg)	0.1500	0.7124	1.2283	0.1156	0.7890	0.8800	0.7800	0.3500	1.2467	1.1791	0.1720
avg(avg)	16.1579	13.0219	9.5700	7.9109	19.0891	8.6989	6.3358	4.2821	7.0218	7.7700	2.3521
max(avg)	262.1000	67.2880	49.8621	139.9233	190.1523	42.1667	28.2786	16.8767	41.9420	33.7467	9.8250
min(std)	0.0071	0	0	0	0	0	0	0	0	0	0
avg(std)	0.3114	2.1193	1.3461	2.2098	2.6304	1.4305	1.0698	0.5418	1.0447	1.2692	0.4335
max(std)	1.9071	22.2428	14.3621	76.3625	34.6368	13.3232	6.5965	8.1508	17.5139	9.3765	8.1388
min(dif)	0.0100	0	0.1000	0.0555	0.0925	0.0400	0.0700	0.0300	0.0280	0.0720	0
avg(dif)	0.4404	2.9972	2.5150	4.2733	4.9186	2.8045	2.0970	1.0693	1.9765	2.3988	0.6130
max(dif)	2.6970	31.4560	25.9520	149.7900	68.6670	25.1800	13.1761	16.1910	34.7000	16.9300	11.5100

Rows in the table are as follows:

- DateSt: Earliest date of grab sample for the station, in the spreadsheet provided.
- DateFn: Latest date of grab sample for the station, in the spreadsheet provided.
- nTot: Total number of grab sample analyses (i.e. unique sample collection date/times).
- n3val: Number of unique sample collection date/times for which there are 3 values.
- n2val: “ ” for which there are 2 values.
- n1val: “ ” for which there is 1 value.
- min(avg): Minimum of all results for the average of the (up to 3) values.
- avg(avg): Mean of “ ”.
- max(avg): Maximum “ ”.
- min(std): Minimum of all results for the standard deviation of the (2 or 3) values.
- avg(std): Mean of “ ”.
- max(std): Maximum of “ ”.
- min(dif): Minimum of all results for the difference between the max and min of the (2 or 3) values.
- avg(dif): Mean of “ ”.
- max(dif): Maximum of “ ”.

Figure 8-1. Phillipsdale. Sonde timeseries with grab sample results superposed.

Sonde timeseries (blue; green/cyan if above/below vertical axis limits). Individual grab sample results (red). Ratio of number of 15-minute intervals with good values to total number, and resulting percent of deployment duration with good value coverage, is shown at upper right in each frame.

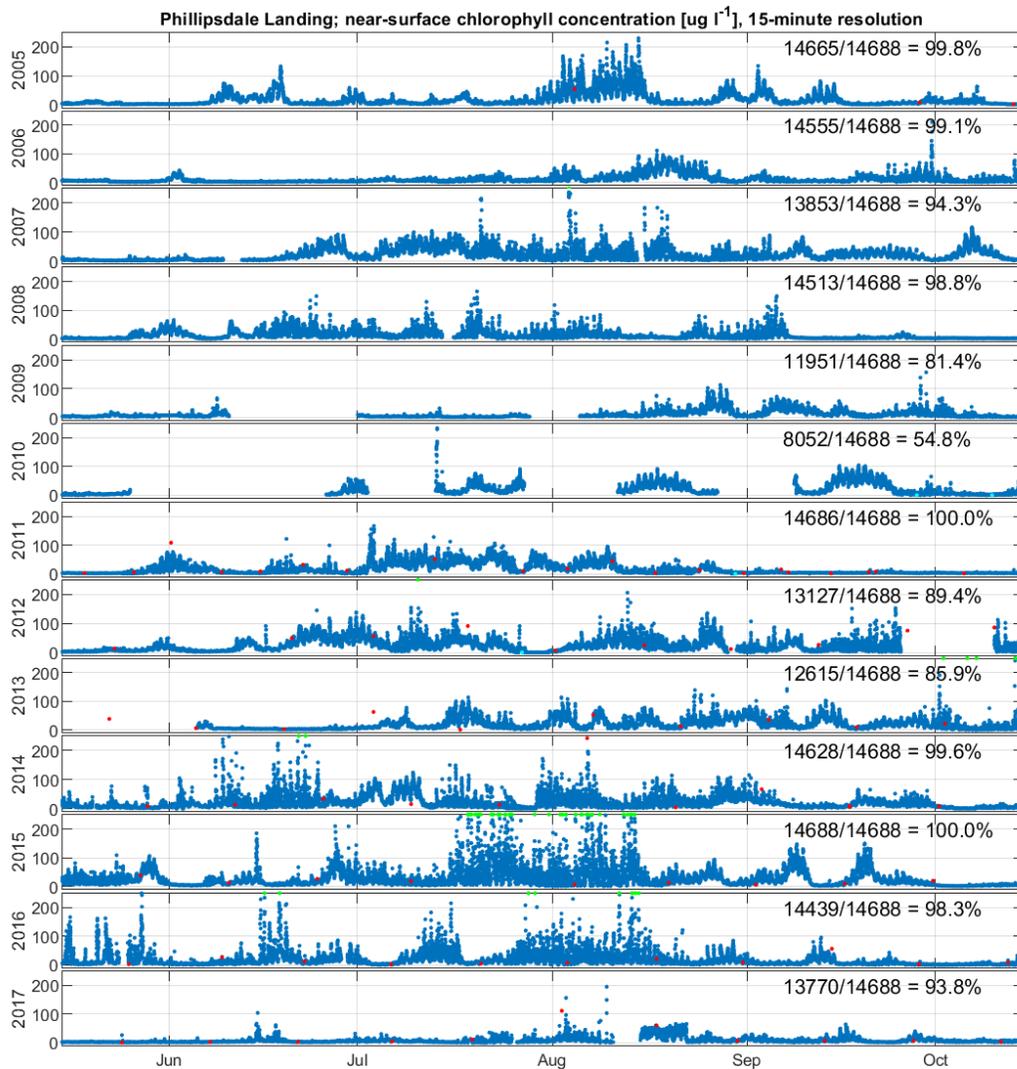


Figure 8-2. As Figure 8-1, Bullock Reach.

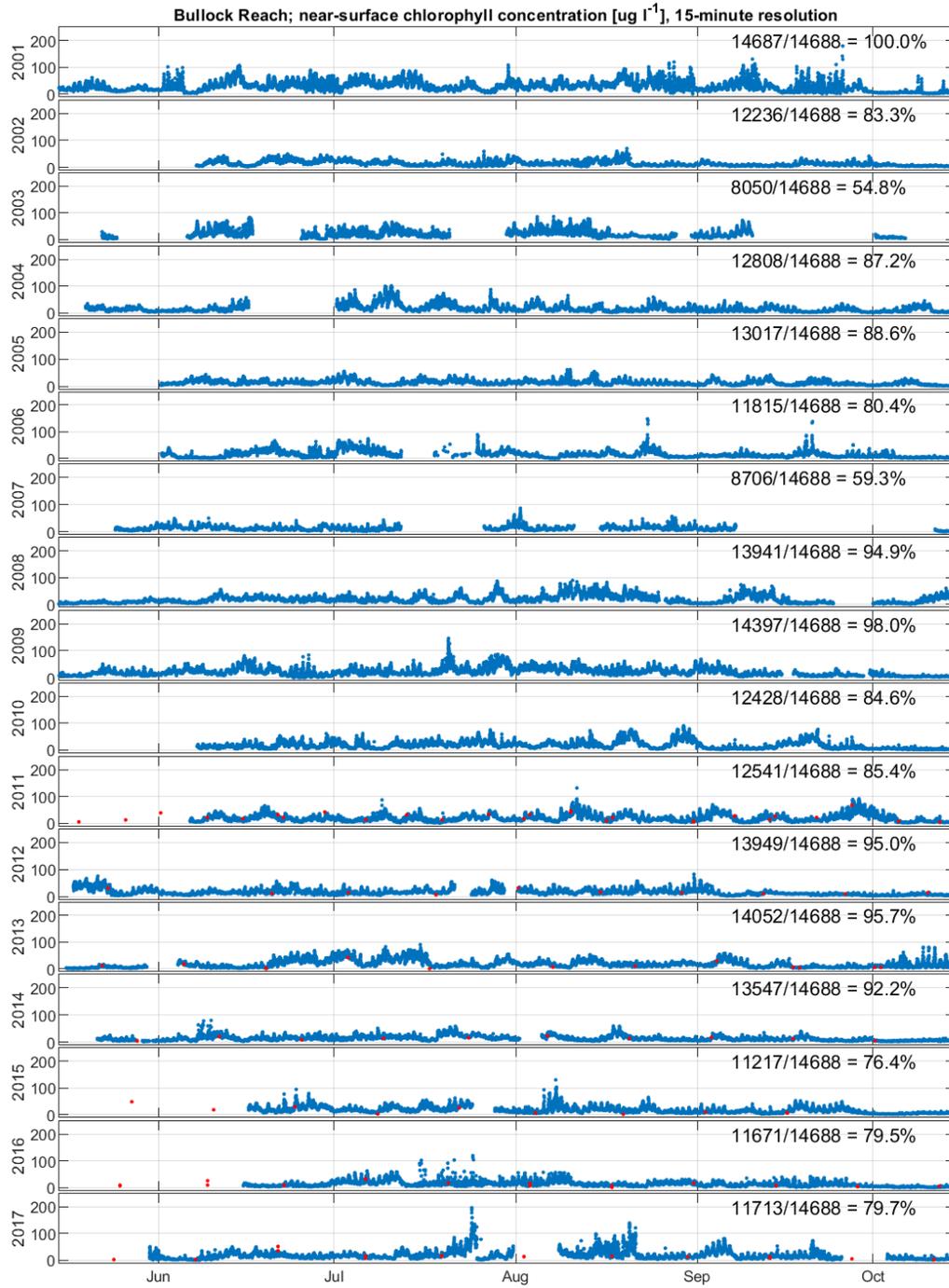


Figure 8-3. As Figure 8-1, Conimicut Point.

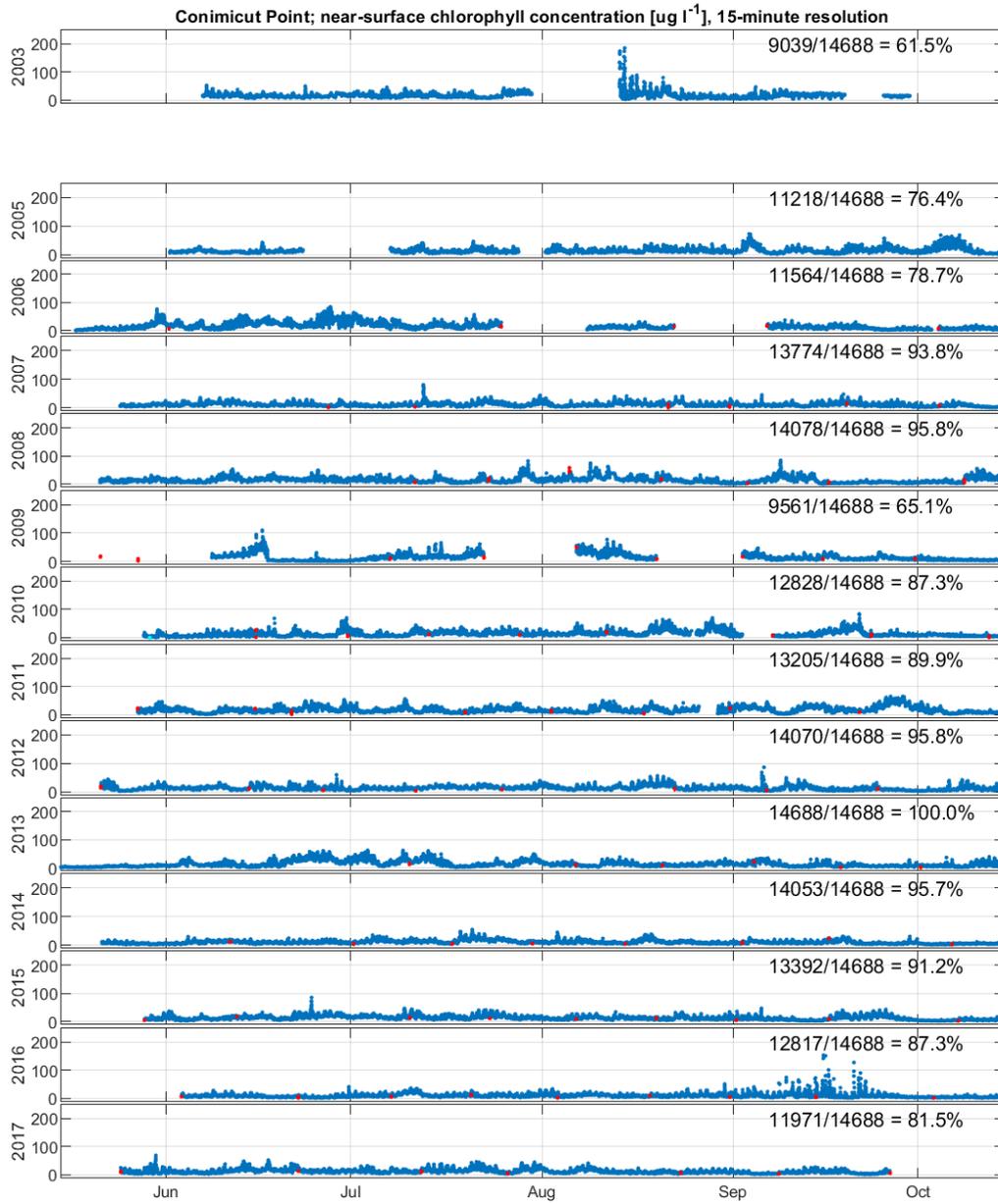


Figure 8-4. As Figure 8-1, North Prudence.

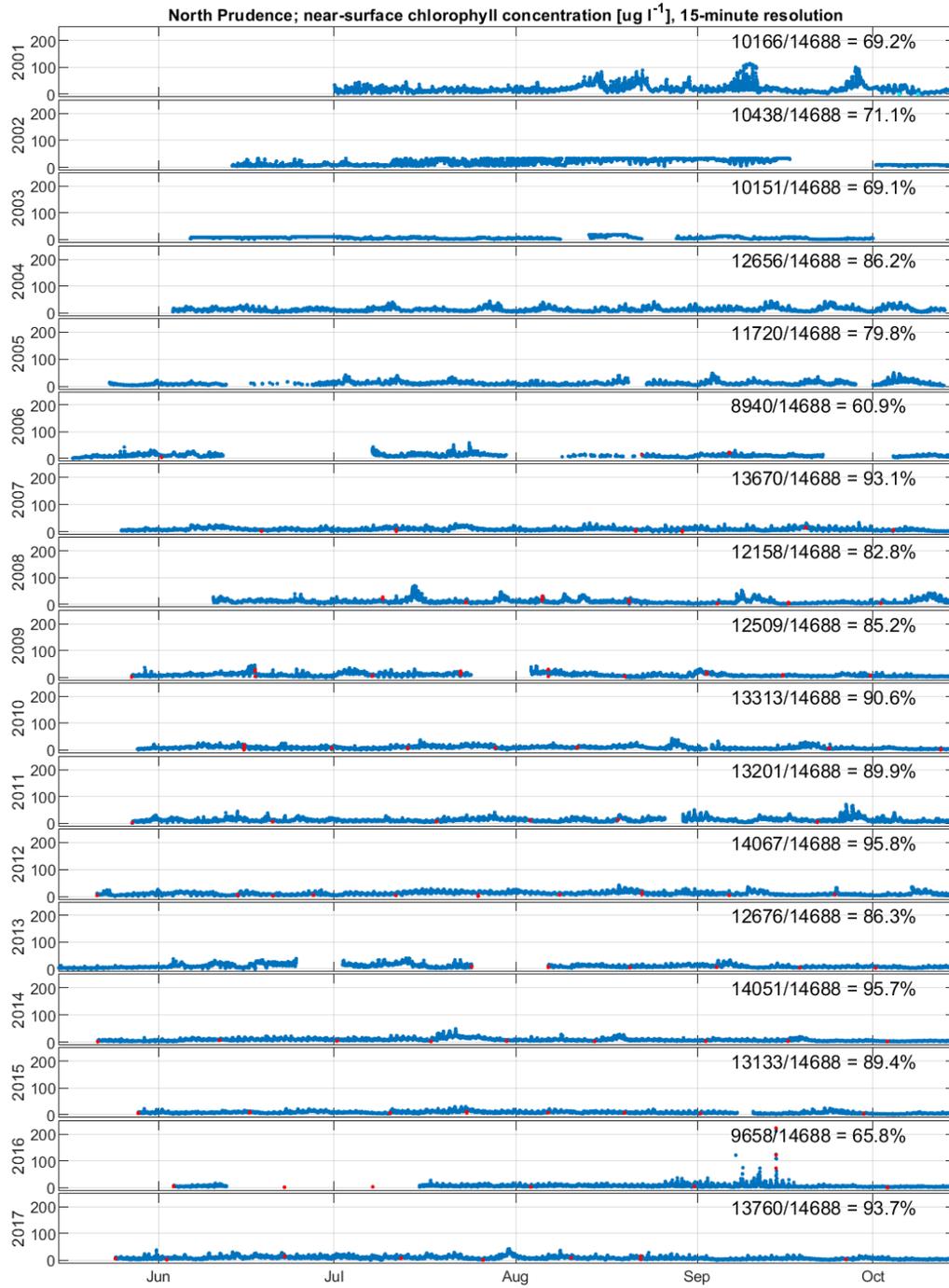


Figure 8-5. As Figure 8-1, Greenwich Bay.

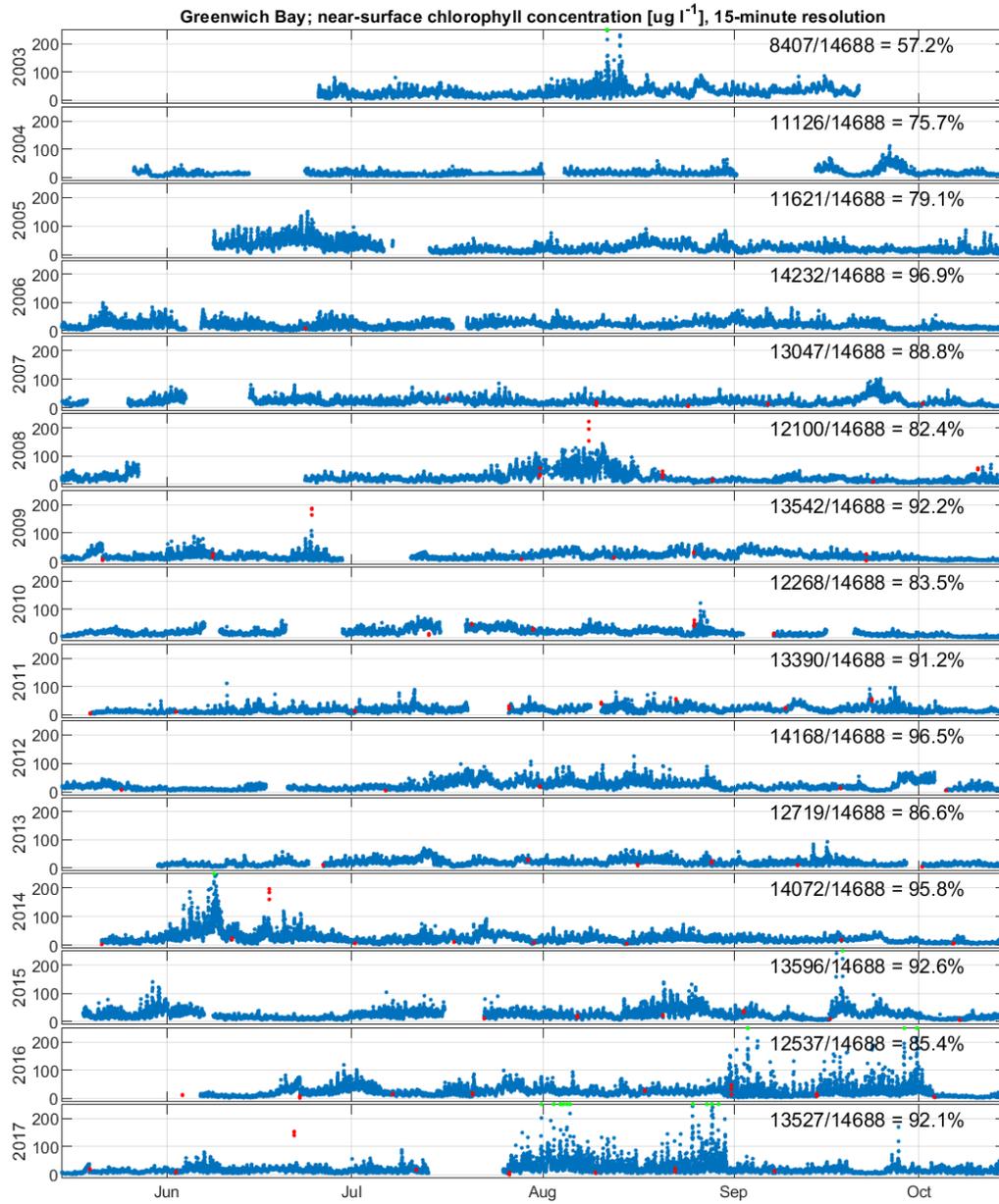


Figure 8-6. As Figure 8-1, Sally Rock.

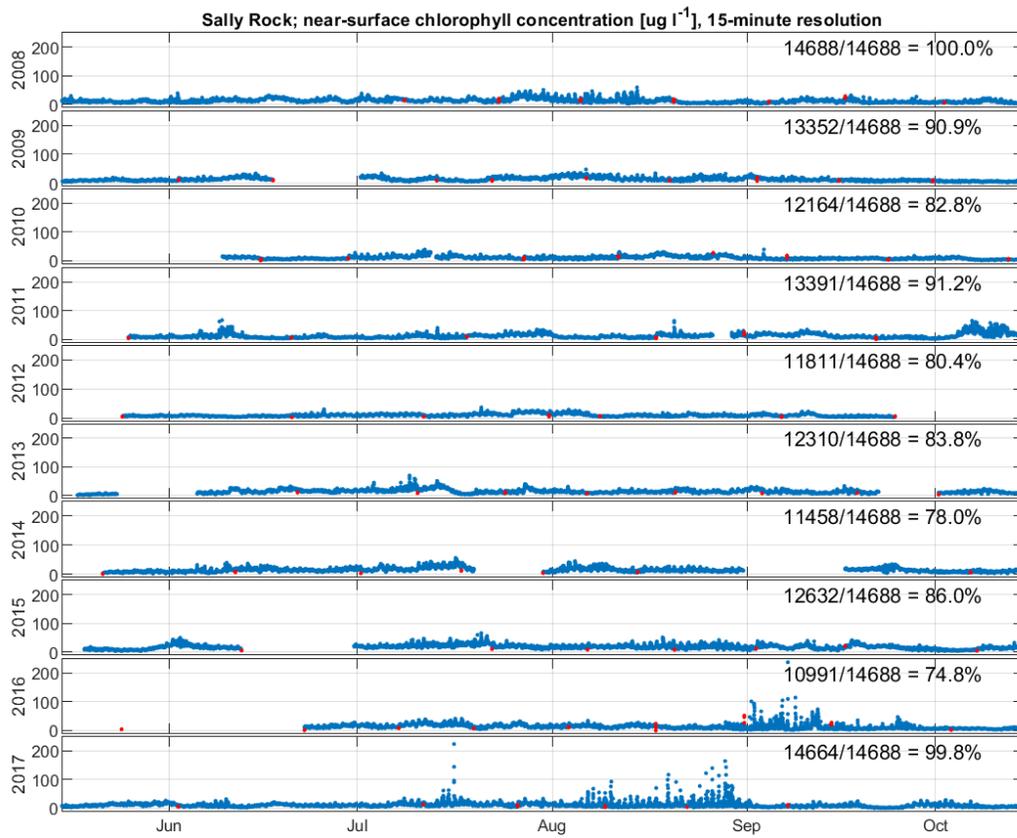


Figure 8-7. As Figure 8-1, Mount View.

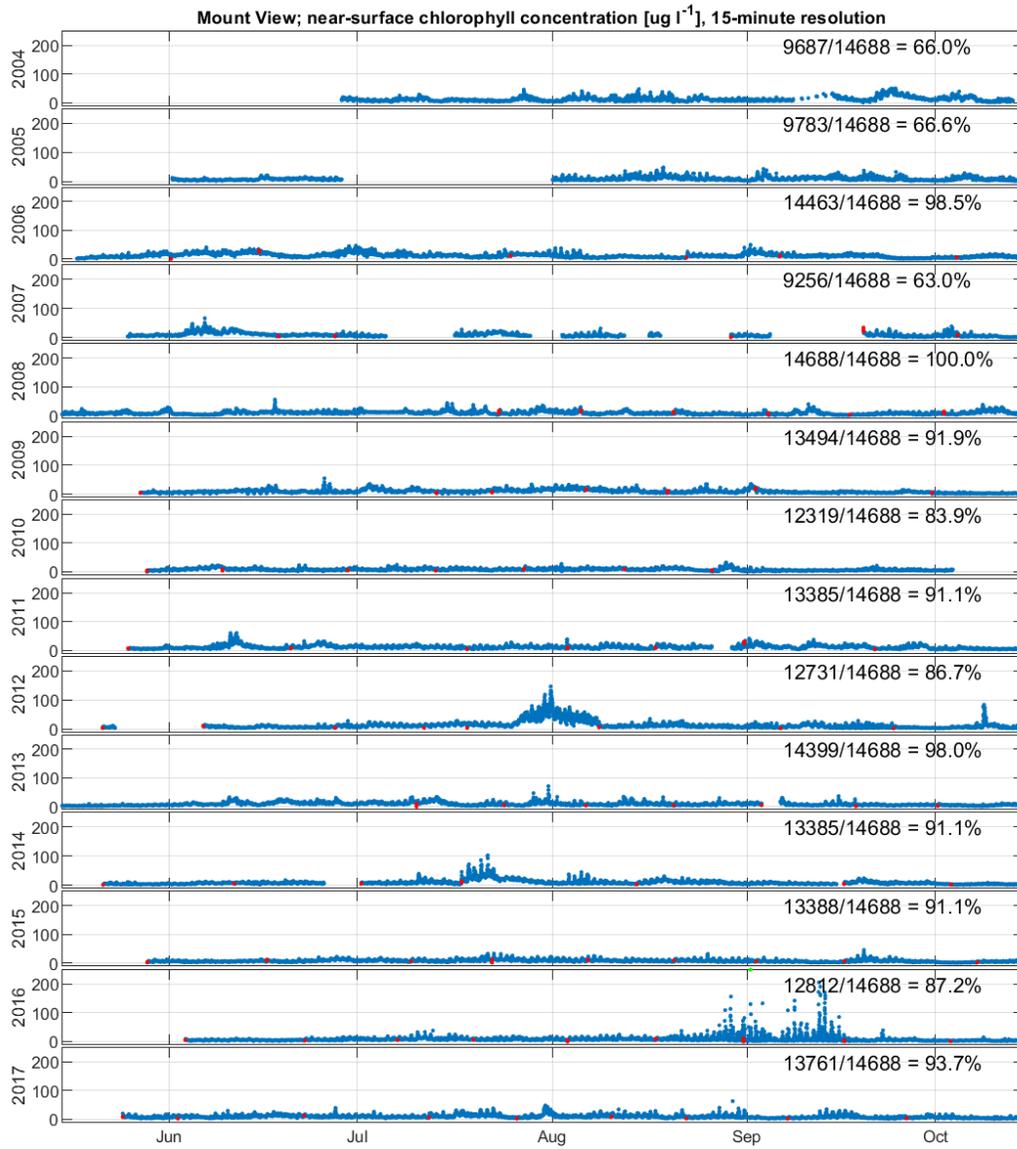


Figure 8-8. As Figure 8-1, Quonset Point.

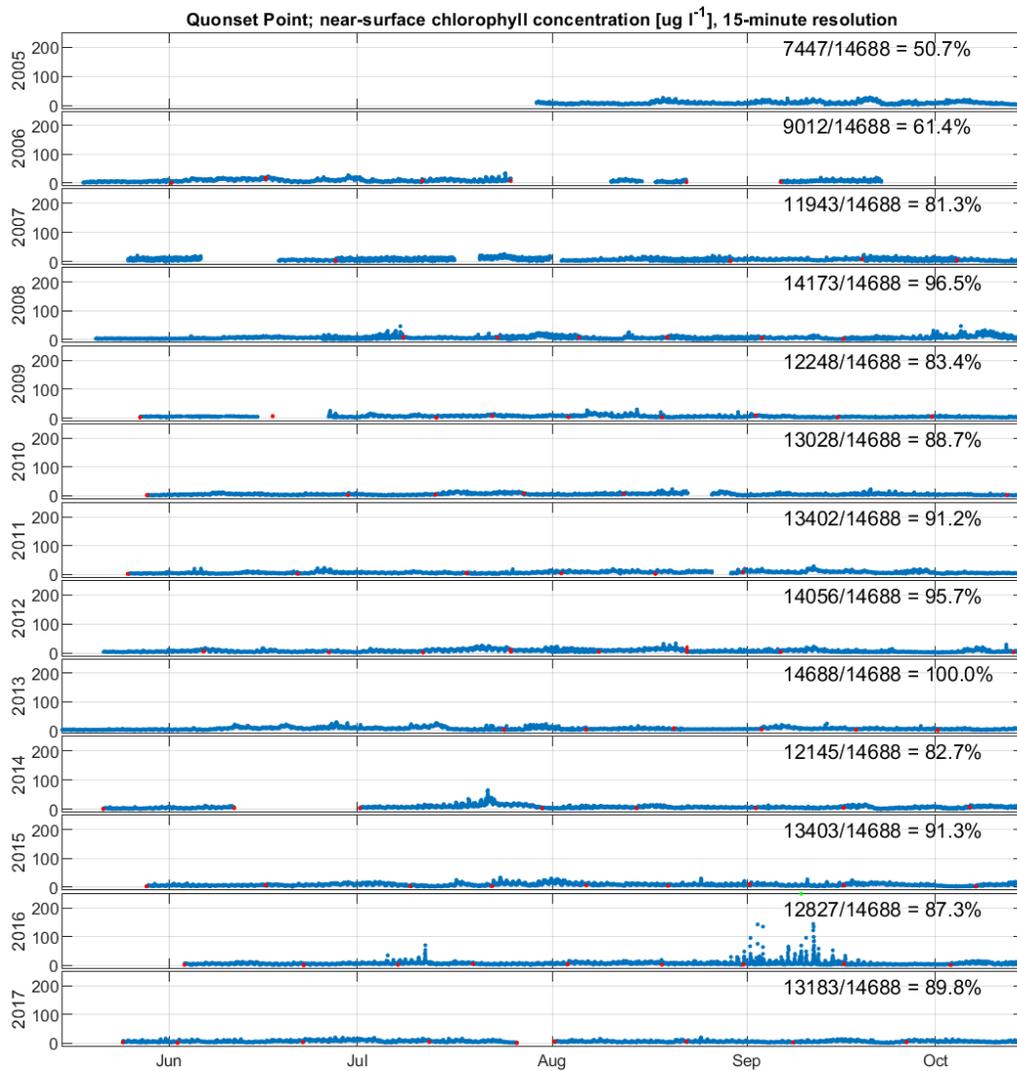


Figure 8-9. As Figure 8-1, Poppasquash Point.

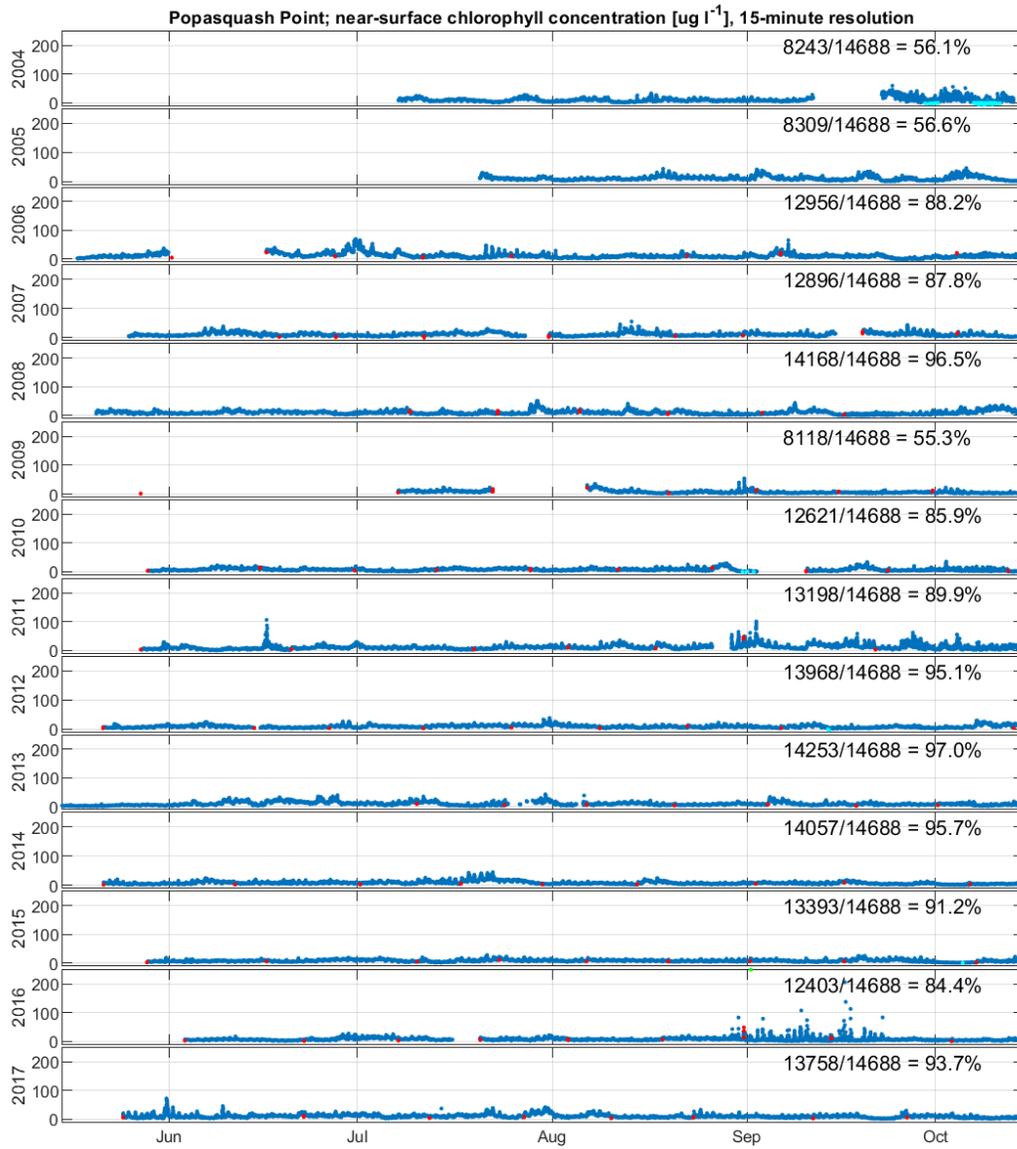


Figure 8-10. As Figure 8-1, T-Wharf.

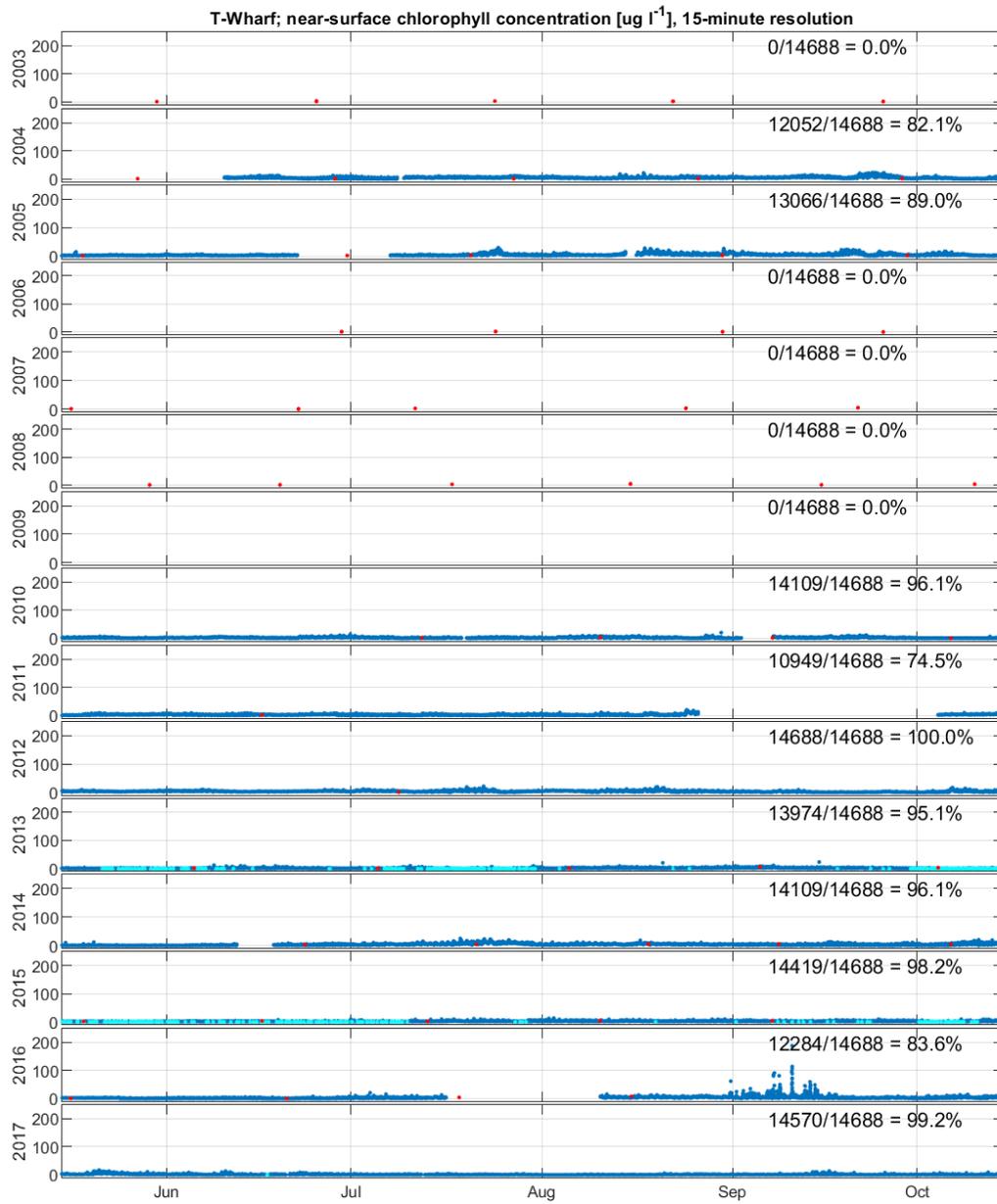


Figure 8-11. As Figure 8-1, Mount Hope.

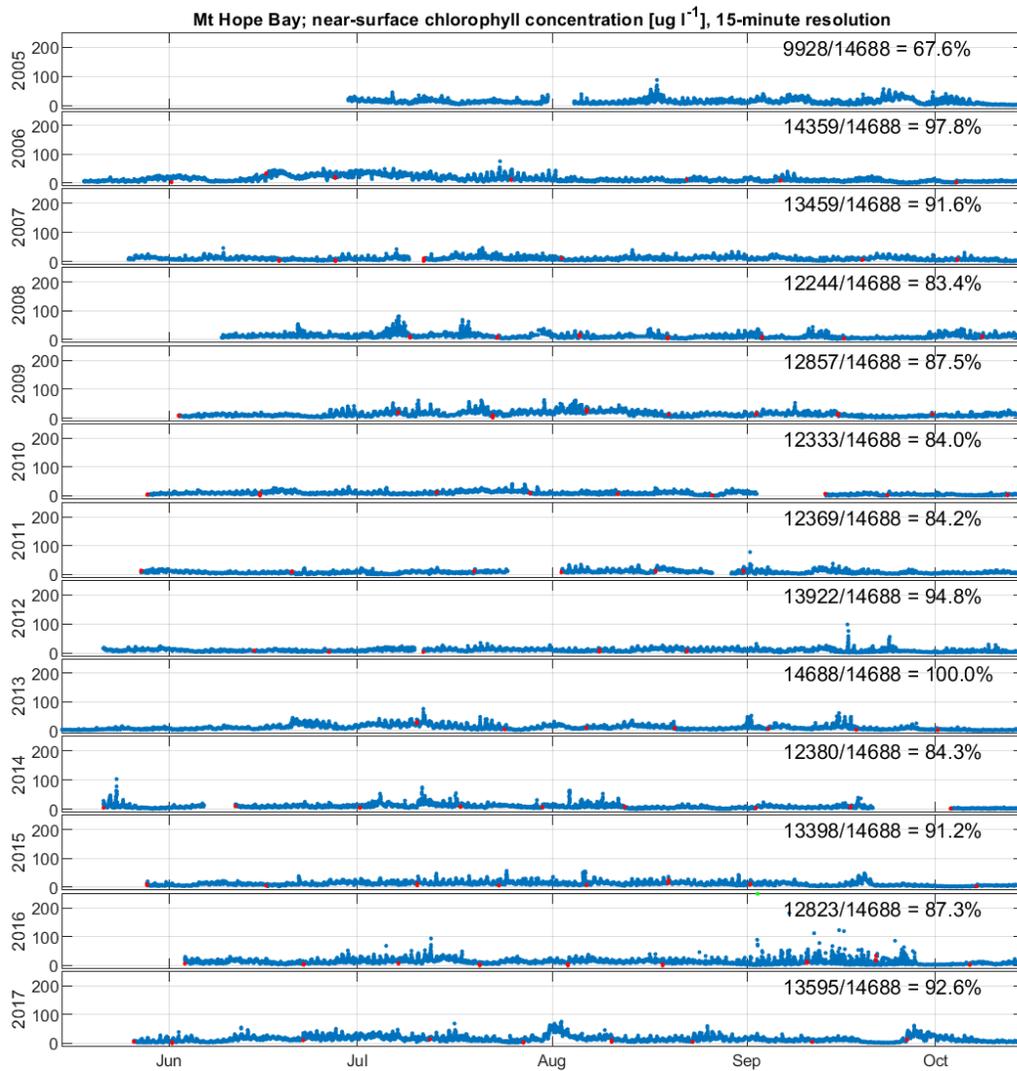


Figure 8-12. Phillipsdale. Scatterplots. See text for discussion.

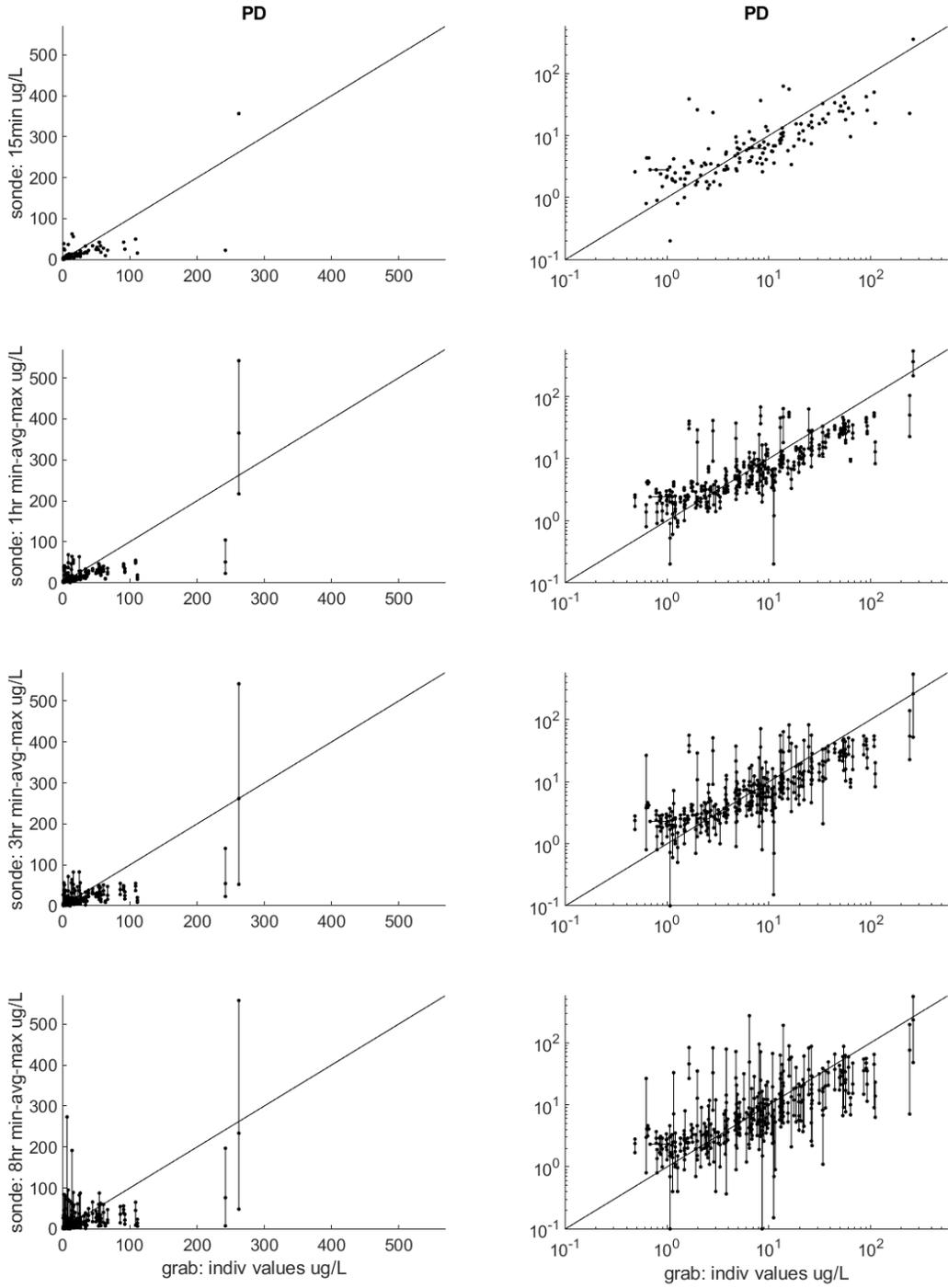


Figure 8-13. Scatterplots using 15-min sonde values. Stations BR, CP, NP, GB, SR.

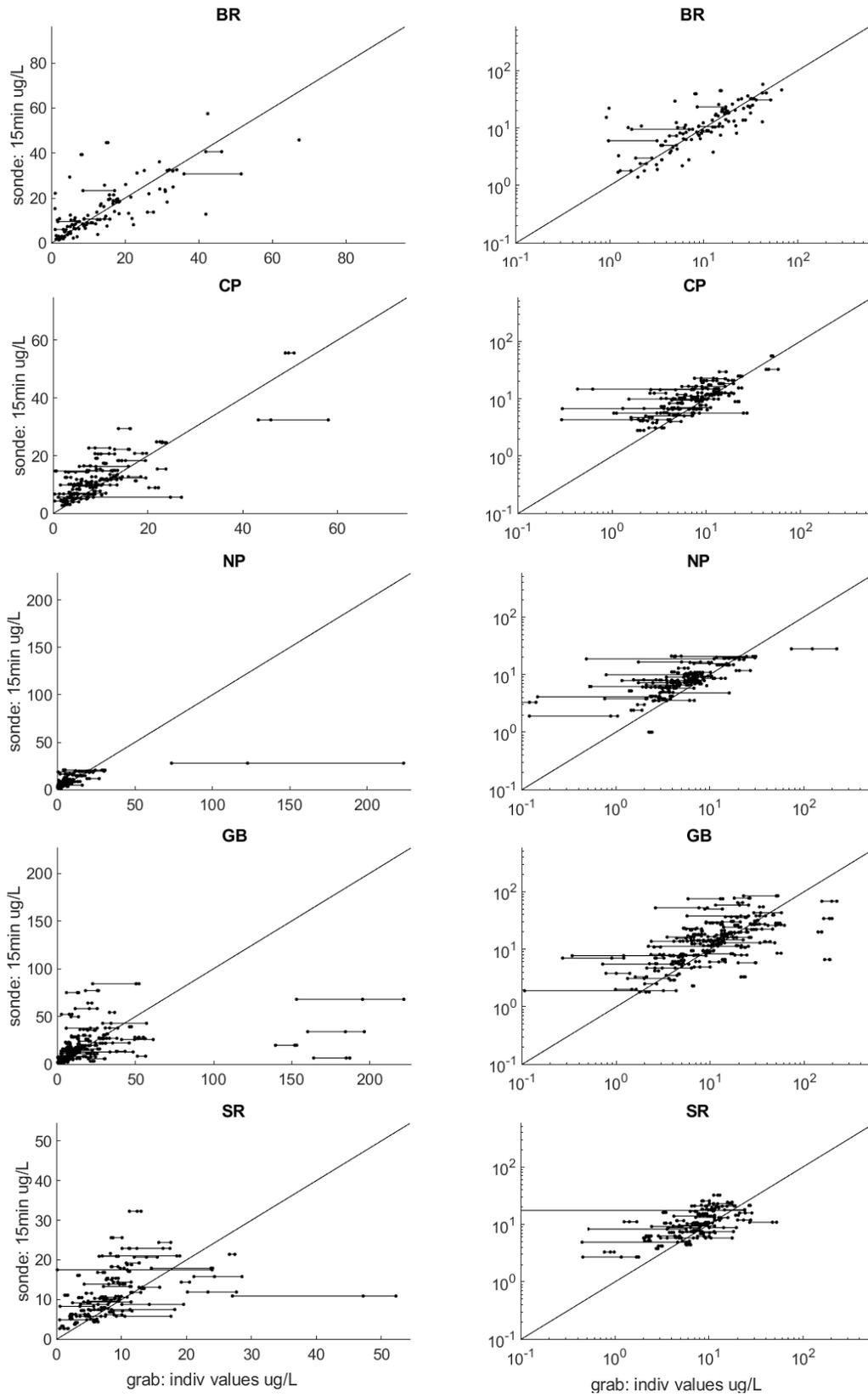
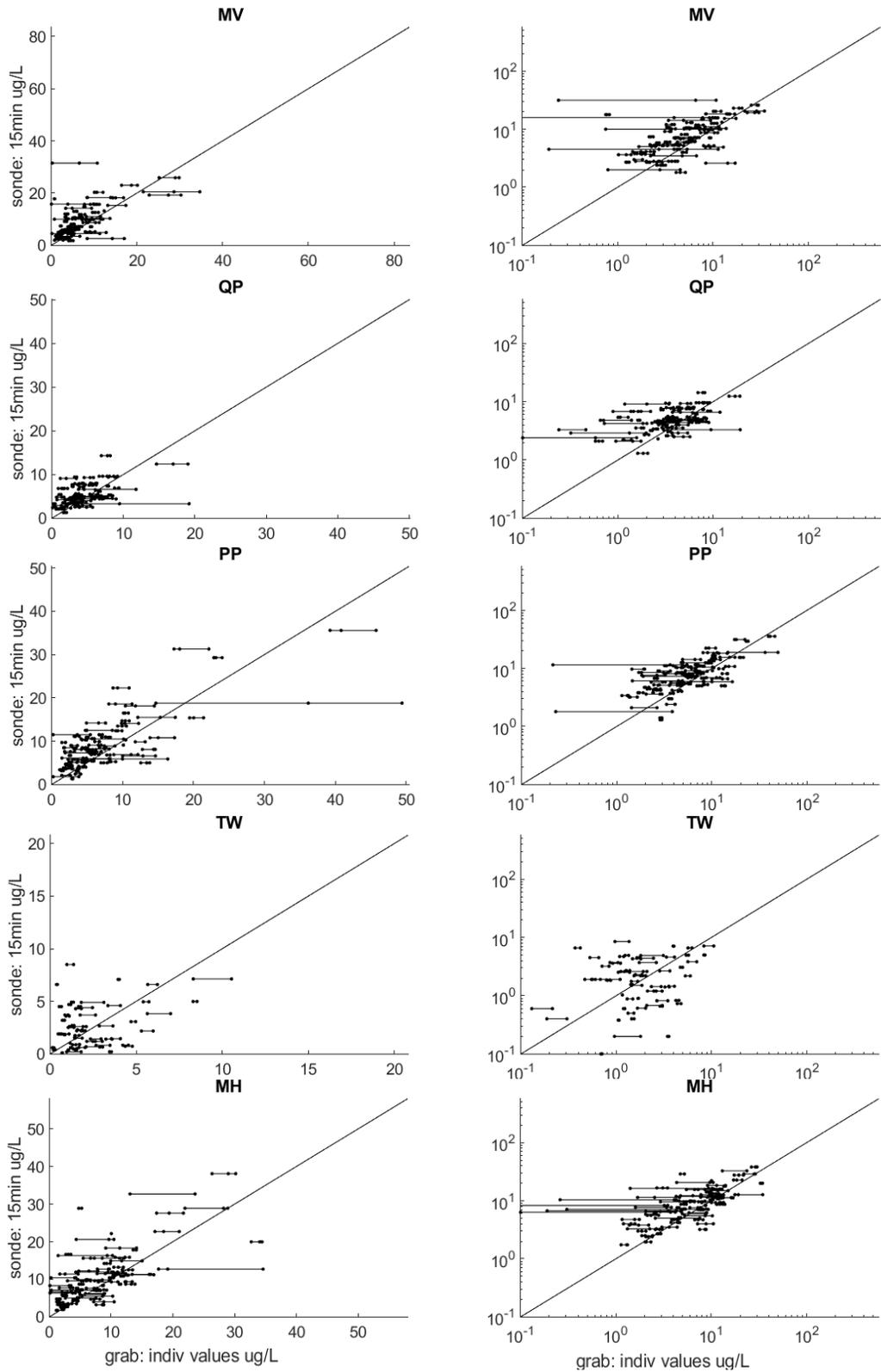


Figure 8-14. Scatterplots using 15-min sonde values. Stations MV, QP, PP, TW, MH.



Task 9 Update SNBIW Chlorophyll Results with Phillipsdale

9.1 Scope

Methodology. Same as in SBNIW. **Expected outputs and outcomes.** Updated versions of figures and tables presenting fixed-site results in the SNBIW chlorophyll chapter, now including Phillipsdale; and an explanation of Phillipsdale results in the context of other stations, paralleling the SNBIW narrative.

9.2 Methods

Methods are the same as in the SNBIW.

The potential recalibration (described in Task 8) of time series chlorophyll (used in this task) has not been implemented (nor had it implemented on the data used in the SNBIW). If the recalibration is judged appropriate and carried out, results in this section may need to be revised.

See page 156 for information about supporting code and data files, and page 150 for how to use them.

9.3 Results

Table 9-1 shows the percent time for each record with chlorophyll data (an update to Table 2 in SNBIW Chapter 16 on chlorophyll).

The index is the season-cumulative (mid-May to mid-Oct) surplus-duration, or sum of the areas swept out above the threshold by the chlorophyll concentration curve during all above-threshold events that season. Its units are $\mu\text{g l}^{-1} \text{ day}$. Its meaning and interpretation are explained in SNBIW Chapter 16, in particular its pages 300-301 and 302-304, and its Figure 3.

The SBNIW included two figures (Figs 11 and 12 in Chapter 16) presenting results of the initial index calculations, for the site-specific 80th percentile threshold values. Fig 11 showed results for individual stations; stations located near to each other in sub-regions of the bay were shown on the same frames of Fig 11 (sub-regions acronyms: PRUB for Providence River and Upper Bay, UWP for Upper West Passage, UEP for Upper East Passage, and GRBY for Greenwich Bay). Fig 12 showed results for the means of stations within the sub-region station groups, all superposed on a single frame. In these two figures gray vertical background bands were included that indicated the years determined to have relatively higher river flow and designated “wet” (see Table 3 of SNBIW Chapter 16).

Here, new versions of the two figures (Figure 9-1 and Figure 9-2) are presented, as updated in the following ways:

- The Phillipsdale station is now included.
- The results for the years 2016 and 2017 have been included.
- The river flow conditions are presented using the wet/typical/dry designations (dark, medium, and light gray vertical background bands) developed in Task 6 above.

Because the years 2016 and 2017 were included in the new calculations, the site-specific thresholds were recomputed. The results are shown in Table 9-2. They are not meaningfully different than the

values computed for 2001-2015 (for stations other than Phillipsdale) in the SNBIW (Table 1 of Chapter 16). The Phillipsdale threshold ($28.2 \mu\text{g L}^{-1}$) is second-highest and comparable to that for Greenwich Bay ($28.8 \mu\text{g L}^{-1}$), as expected because Phillipsdale is known to be a shallow site with weak flushing that is more susceptible to chlorophyll blooms, like Greenwich Bay.

The new results at individual stations (Figure 9-1) include Phillipsdale (top frame) and put it in the context of the other stations. Phillipsdale has the highest values of the index in nearly all years. In many years Phillipsdale is higher than the next highest station, Greenwich Bay, by a substantial amount. Inter-annual variations at Phillipsdale differ from those at other sites, for example lower values in 2008 and 2009 were followed by a 5-year period of higher values from 2010 to 2014. These patterns were relatively independent from the wet, typical or dry year river flow conditions (Figure 6-2). This is consistent with our understanding that other factors, in addition to river flow, have a strong influence on variability at the Phillipsdale site (see above tasks).

The results for 2016 and 2017 (Figure 9-1) at most stations were not markedly different from the prior several years. However they do show steady or declining index values at the stations where the highest values have typically occurred (Phillipsdale, Greenwich Bay, and Sally Rock). This is consistent with easing of eutrophication associated with completion of treatment facility nutrient load reductions in 2014, as discussed in earlier tasks above.

In contrast to this pattern, the 2017 index results for Mount Hope, a deep station where values are typically low, were notably higher and the reason for this is not known. Long-term trends at that station differ from other stations, suggesting different processes may be at play there (see Task 12).

Presenting group-averaged index results in the same frame (Figure 9-2) reinforces that PD has the highest values in many years. It also helps highlight the declines in 2016-2017 at PD and the GRBY group (GB and SR stations), and the increase at MH in 2017.

It bears repeating that these results are relative to site-specific thresholds. A different approach is to apply constant thresholds to all stations bay-wide, as is done in analysis of oxygen time series to compute the Hypoxia Index. The bay-wide thresholds approach is carried out in Task 10.

Table 9-1. Percent of time during mid-May to mid-October analysis period with valid near-surface chlorophyll values at each of the eleven sites in each of the seventeen years. Asterisks (*) indicate records with less than 55 percent valid values, which were not included in the analysis. Blanks indicate sites for which data records were not obtained.

	PD	BR	CP	NP	MV	QP	PP	TW	GB	SR	MH
2001		100		69							
2002		83		71							
2003		55	62	69				*	57		
2004		87		86	66		56	82	76		
2005	100	89	76	80	67	*	57	89	79		68
2006	99	80	79	61	99	61	88	*	97		98
2007	94	59	94	93	63	81	88	*	89		92
2008	99	95	96	83	100	97	97	*	82	100	83
2009	81	98	65	85	92	83	55	*	92	91	88
2010	55	85	87	91	84	89	86	96	84	83	84
2011	100	85	90	90	91	91	90	75	91	91	84
2012	89	95	96	96	87	96	95	100	97	80	95
2013	87	96	100	86	98	100	97	95	87	84	100
2014	100	92	96	96	91	83	96	96	96	78	84
2015	100	76	91	89	91	91	91	98	93	86	91
2016	98	80	87	66	87	87	84	84	85	75	87
2017	94	80	82	94	94	90	94	99	92	100	93

Table 9-2. Chlorophyll thresholds for each site, computed as the 80th percentile of all years' data between May 15 and October 14 from that site.

Site group	Site	Acronym	Threshold [$\mu\text{g L}^{-1}$]
(PD)	Phillipsdale	PD	28.2
PRUB	Bullock's Reach	BR	25.7
PRUB	Conimicut Point	CP	18.9
PRUB	North Prudence	NP	14.5
UWP	Mount View	MV	12.8
UWP	Quonset Point	QP	9.2
UEP	Poppasquash Point	PP	12.7
UEP	T-Wharf	TW	4.8
GRBY	Greenwich Bay	GB	28.8
GRBY	Sally Rock	SR	17.5
(MH)	Mount Hope	MH	15.3

Figure 9-1. Chlorophyll Index, individual sites, site-specific 80th percentile thresholds (Table 9-2).

Gray shaded bars: wet, typical, and dry years (see Task 6). To improve clarity by reducing overlap of lines and symbols, symbols from each site are systematically offset a small distance horizontally relative to those of other sites in the frame. Uncertainties are due to gaps in the time series due to sensor malfunction or data not meeting quality standards. In most cases, the uncertainties are smaller than the symbol size.

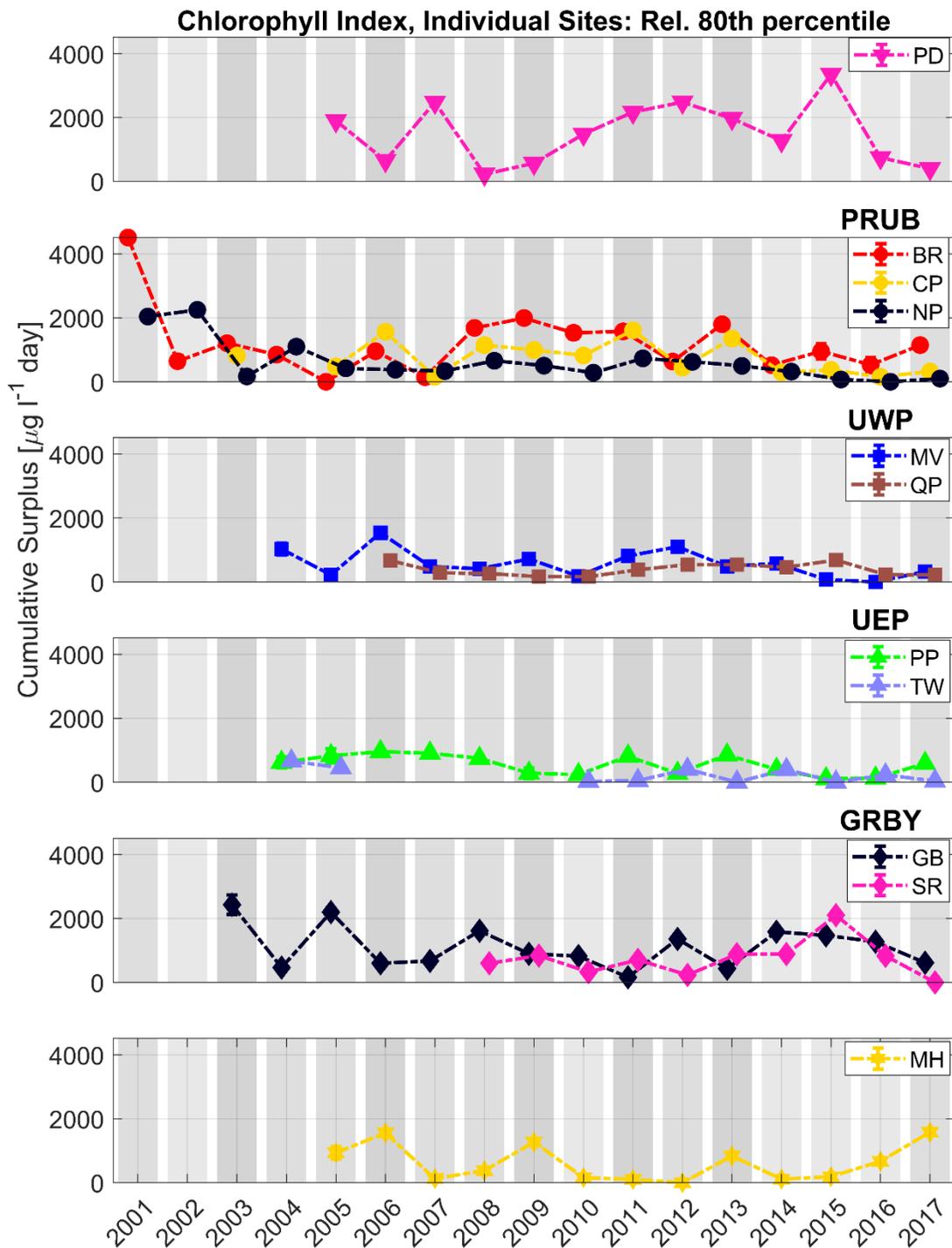
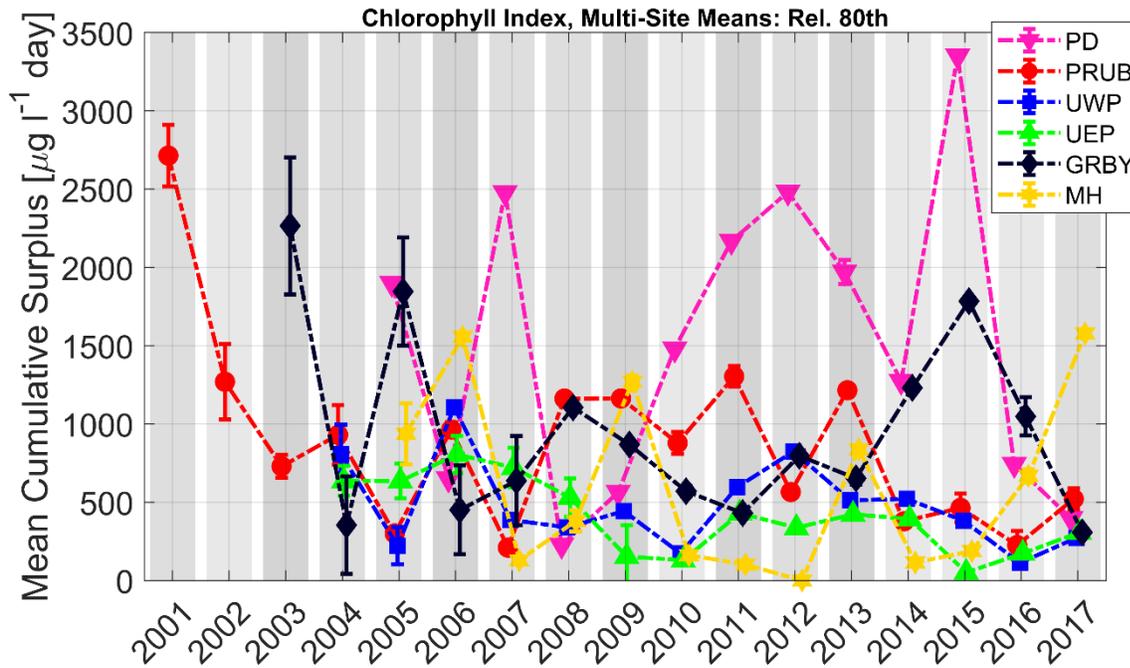


Figure 9-2. Chlorophyll Index, multi-site averages, site-specific 80th percentile thresholds (Table 9-2).
 Shown as in Figure 9-1. The Phillipsdale (PD) and Mount Hope (MH) curves are single sites, repeated from Figure 9-1.



Task 10 Refine Chlorophyll Index

10.1 Scope

Objective. Revisit initial definitions of chlorophyll index used in SNBIW. **Methodology.** In consultation with the NBEP-formed subcommittee of phytoplankton experts (including Oviatt and Borkman, URI/GSO) that guided the 80th-percentile SNBIW index definition, other definitions will be explored; for example, concentration-based thresholds (instead of percentile-based). **Expected outputs and outcomes.** Revised index results, based on alternative definitions, with more robust meanings.

10.2 Methods

Methods are the same as in SNBIW except that different values of the thresholds are used in the event calculations, and the thresholds used are held fixed across all stations bay-wide (in the SNBIW, and Task 9 above, the thresholds were site-specific).

The potential recalibration (described in Task 8) of time series chlorophyll (used in this task) has not been implemented (nor had it been implemented on the data used in the SNBIW). If the recalibration is judged appropriate and carried out, results in this section may need to be revised.

A group videoconference was held on 3/31/2020 with a number of phytoplankton experts. Courtney Schmidt (NBEP) e-hosted the meeting and the following participated: from URI/GSO, Candace Oviatt, Heather Stoffel, Kristin Huizenga; from RI DEM, David Borkman; from Brown U, Warren Prell, David Murray; from NBC, Eliza Moore, Molly Welsh, Luis Cruz. The following meeting minutes were distributed afterwards and no requests for modifications were made:

Meeting Minutes, March 31, 2020

The calculation of the Index for the 2017 State of Narragansett Bay and Its Watershed (SNBIW) was briefly reviewed. The index is a season-cumulative measure, based on treating time series measurements from mid-May through mid-Oct. The threshold value must be specified. For SNBIW, the threshold used was site-specific (different at each fixed site monitoring station), and computed as the 80th percentile of all measurements made at that site in all years of sampling (threshold values shown in Table 1 of SNBIW Chapter 16).

A way to refine the index was proposed, by using a group of three fixed thresholds at all sites (rather than site-specific values). This is parallel to how oxygen is analyzed (thresholds 1.4, 2.9, and 4.8 mg L⁻¹—each applied bay-wide).

For threshold values, an initial suggestion was made to use the 5 µg L⁻¹ and 20 µg L⁻¹ thresholds from the NCCA report, plus some third threshold that is in the 30-35 µg L⁻¹ range in order to help capture high-concentration events.

Discussion led to the general conclusion that doing multiple sets of calculations of the index--using both site-specific thresholds, and constant thresholds applied bay-wide, will both be valuable because they provide complementary information. Site-specific thresholds (at 80th percentile for example) can help characterize the nature of events at sites no matter how high or

low their long-term mean concentration; using a set of three constant thresholds applied bay-wide can help identify regional patterns and trends.

Multiple participants mentioned they think it will be important that, in association with the event-based index calculations, summary statistics of chlorophyll should also be reported based on more traditional methods.

It was suggested that in addition to the season-cumulative index, the more granular underlying statistics of the multiple individual events the index groups together, should be documented and used to characterize chlorophyll conditions.

More than one participant considered the $5 \mu\text{g L}^{-1}$ value to be too low to be useful. It also was pointed out that the "good/fair/poor" language used in the NCCA report should be avoided if at all possible and other terminology used instead (low/intermediate/high, or oligotrophic/mesotrophic/eutrophic, as applicable).

A request was made, to circulate a subset of the initial results of the calculations to the group for their review, so they have the opportunity to suggest adjustments to the approach taken. This initial set of results could be from three representative stations (e.g. Phillipsdale, Quonset Point, T Wharf).

There was support for the suggestion to define the three constant thresholds by simply grouping all data from all sites and all years, and computing the 20th, 50th, and 80th percentiles from them. This is parallel to the analysis approach used on the spatial surveys dataset.

One participant suggested defining the thresholds only using data from about 2014 (or perhaps 2012) onward, i.e. after nutrient load reductions were mostly or fully completed.

There also was discussion of potentially defining thresholds using (i) subsets of years (examples being "pre-reduction" years up through 2012, "reduction transition" 2012 through 2014", and/or "post-reduction" 2015 and beyond), and/or (ii) using subsets of fixed-site stations (groups in or near each other, or certain embayments, for example). No particular combination was settled on during the discussion.

Measurement methods were discussed. Lab extractions of chlorophyll have been regressed against time series measurements at the GSO dock and show a strong correlation, with the time series values lower. The lab extractions yield chlorophyll a, and are higher than the time series in the regression relationship, which led to some discussion of how this could be consistent with the time series values being total chlorophyll. Separately it was explained that the spatial surveys have been calibrated to lab extraction values.

There was some discussion of improving the fidelity of the time series dataset (susceptible to drift and calibration difficulties, for example due to fouling) by adjustment toward independent measurements treated as "truth". The Oviatt lab at URI/GSO is moving forward on this, using their grab samples dataset. NBC staff are intending to follow a similar approach. The spatial surveys dataset, having been calibrated to extracted chl results, should also be useful.

Following the meeting there was further discussion of a number of the topics by email. Some of the requested calculations, such as tabulation of event statistics at a more granular level than the seasonal

index and use of more traditional summary statistics were considered outside the scope of this project. It was decided that using subsets of years and/or subsets of stations to define thresholds should not be pursued at this time, as they are not yet sufficiently well motivated and will lead to interpretation challenges. Initial results of the calculations were circulated to the group as requested.

See page 156 for information about supporting code and data files, and page 150 for how to use them.

10.3 Results

Based on consensus support from the experts, during the meeting and in follow-up email discussions:

- The index calculations should be refined by doing them using a set of fixed thresholds applied bay-wide (in addition to the site-specific thresholds in SNBIW Chapter 16 and Task 9).
- Three bay-wide chlorophyll thresholds should be used, parallel to the way three oxygen thresholds are used bay-wide for the Hypoxia Index.
- The three thresholds should be defined as the 20th, 50th, and 80th percentiles of the chlorophyll data from all sites and all years grouped together.
 - The thresholds should be computed as percentiles of the observations because water quality standard thresholds from investigations of other waterbodies, or as developed for national programs, may not be appropriate for Narragansett Bay. For chlorophyll there is no objective standard by which to select the threshold values, in contrast to the threshold values for oxygen, which are based on biological responses of affected organisms; the same chlorophyll concentration can mean poor or good water quality depending on the time of year, the location, the circulation/mixing conditions, etc.
 - The 20th, 50th, and 80th percentiles are chosen to span a representative range, similarly to what was done for analysis of the spatial surveys dataset in the SNBIW.
- The index should be denoted “Chlorophyll Index” from here forward (not “Chlorophyll Bloom Index”), to avoid the term “bloom” which causes confusion during interpretation of the results.

The results from using site-specific thresholds (Task 9) help identify patterns of inter-annual variability and long-term trends in conditions at each individual site, regardless of its long-term mean. This is useful because there is a wide range of long-term means from station to station (see Table 9-2).

The Chlorophyll Index is the season-cumulative (mid-May to mid-Oct) surplus-duration, or sum of the areas swept out above the threshold by the chlorophyll concentration curve during all above-threshold events that season. Its units are $\mu\text{g l}^{-1} \text{ day}$. Its meaning and interpretation are explained in SNBIW Chapter 16, in particular its pages 300-301 and 302-304, and its Figure 3.

The results presented here, from using three fixed thresholds (not site-specific) applied bay-wide, are useful to help gauge regional patterns and trends. Compared to the site-specific threshold results, they are in some ways easier to interpret and more robust.

The chlorophyll measurements from all 11 stations and all sampled years, between mid-May and mid-October, were grouped together and used to compute 20th, 50th, and 80th percentiles. The results are $4.9 \mu\text{g L}^{-1}$, $9.4 \mu\text{g L}^{-1}$ and $17.6 \mu\text{g L}^{-1}$. These are comparable to concentration levels used in other studies. For example, in recommendations intended to be general across all waterbodies, USEPA (2015)

designates $5 \mu\text{g L}^{-1}$ (very close to the 20th percentile computed here for Narragansett Bay) as the boundary between what they describe as “good” and “fair” water quality and $20 \mu\text{g L}^{-1}$ (reasonably close to the 80th percentile computed here for Narragansett Bay) as the boundary between water quality conditions they denote as “fair” and “poor”.

The new Chlorophyll Index results, from the event-based calculations using bay-wide thresholds $4.9 \mu\text{g L}^{-1}$, $9.4 \mu\text{g L}^{-1}$ and $17.6 \mu\text{g L}^{-1}$ at individual stations, are presented in Figure 10-1, Figure 10-2, and Figure 10-3 respectively.

For the lowest $4.9 \mu\text{g L}^{-1}$ threshold (Figure 10-1), the site with the highest values sustained over many years of the record is GB, while PD had the highest individual index value in 2015 and values comparable to GB during 2012-2014. These results reflect that the GB and PD stations are in shallow embayments with persistently high chlorophyll concentrations, and prone to more chlorophyll blooms. The next-highest index values are generally at SR and BR, the two stations most proximal to GB and PD respectively, but in deeper water somewhat less prone to blooms. The CP and NP stations, each located increasingly south of BR in the Providence River, have index results progressively lower than BR and illustrate the southward-decreasing north-south gradient in chlorophyll levels. Index results for stations MV, PP, and MH, all more distant from the Providence River, are comparable to the NP or CP results, and the results are slightly lower at farther distant QP in the southern part of the Upper West Passage. The only station for which the index is not substantially above zero in all years is TW, the site closest to the open ocean, with the lowest long-term mean chlorophyll levels.

Inter-annual variability in the Chlorophyll Index for the lowest $4.9 \mu\text{g L}^{-1}$ threshold (Figure 10-1) is not uniform bay-wide. At PD it is most pronounced, with intermediate values before 2008, then relatively low values 2008-2011, followed by relatively high values 2012-2015 and the lowest value in 2017. At other stations the inter-annual variations are more modest, and generally independent from location to location. The inter-annual variability of the index is not strongly associated with year to year changes in river runoff (shown as gray bars in Figure 10-1). In addition, it does not show widespread or marked declines in the years since 2014, when several years of treatment plant upgrades were fully completed leading to substantial decreases of nutrient loading to the bay; however, a decline in 2017 is evident at multiple stations (PD, GB, and SR).

The Chlorophyll Index results for intermediate $9.4 \mu\text{g L}^{-1}$ threshold (Figure 10-2) have many similarities to those just described for the lower threshold. As expected for a higher threshold, the index values are generally lower. This decrease is most strongly apparent for the sites at deeper and more southern locations, where the long-term mean chlorophyll levels (Table 9-2) are not markedly different from the $9.4 \mu\text{g L}^{-1}$ threshold. In contrast, for the shallower sites with higher long-term mean chlorophyll levels, the results for the $9.4 \mu\text{g L}^{-1}$ threshold (Figure 10-2) are very weakly reduced from the results for the $4.9 \mu\text{g L}^{-1}$ threshold (Figure 10-1), as expected given that a high proportion of their values are substantially higher than both thresholds. With respect to inter-annual variability the results are quite similar to those of the lower threshold, with the 2017 declines at GB and SR somewhat more pronounced.

The Chlorophyll Index results for the highest $17.4 \mu\text{g L}^{-1}$ threshold (Figure 10-3) continue the pattern of changes between the intermediate and low threshold results (Figure 10-2 relative to Figure 10-1). The index is yet lower for all sites, as expected for a higher threshold, with the result that all of the deeper down-bay stations (NP, MV, QP, PP, TW, MH) have index at or near zero in many of the years. For this

high threshold only sites nearer to or within embayments, where long-term mean chlorophyll levels are higher, have non-zero index in most years (PD, BR, CP, GB, SR). The inter-annual variability of the index at these stations is more pronounced than that of the index for lower thresholds, but is similar in that it is relatively independent from station to station, not closely linked to river flow, and not strongly different in the years since 2014 when plant upgrades were completed. The decline in 2017 at PD, GB, and SR is most pronounced for this high threshold.

When individual station index results are averaged across station groups from different geographic portions of the bay, the patterns and features just described are clear from a single 3-frame plot (Figure 10-4) showing results for all three thresholds.

It is notable that inter-annual variability in the Chlorophyll Index is not very similar to that of the Hypoxia Index (Task 4). The latter is tied closely to river flow, with higher index during wet years, and also showed a somewhat stronger decline across many sites since 2014 when plant upgrades were completed; these features are not characteristic of the Chlorophyll Index.

Figure 10-1. Chlorophyll Index relative to (lowest) 4.9 $\mu\text{g L}^{-1}$ threshold.

Individual stations, same threshold concentration applied bay-wide. Gray shaded bars: wet, typical, and dry years (see Task 6). To improve clarity by reducing overlap of lines and symbols, symbols from each site are systematically offset a small distance horizontally relative to those of other sites in the frame. Error bars indicate uncertainties based on gaps in the time series due to sensor malfunction or data not meeting quality standards. In most cases, the uncertainties are smaller than the symbol size.

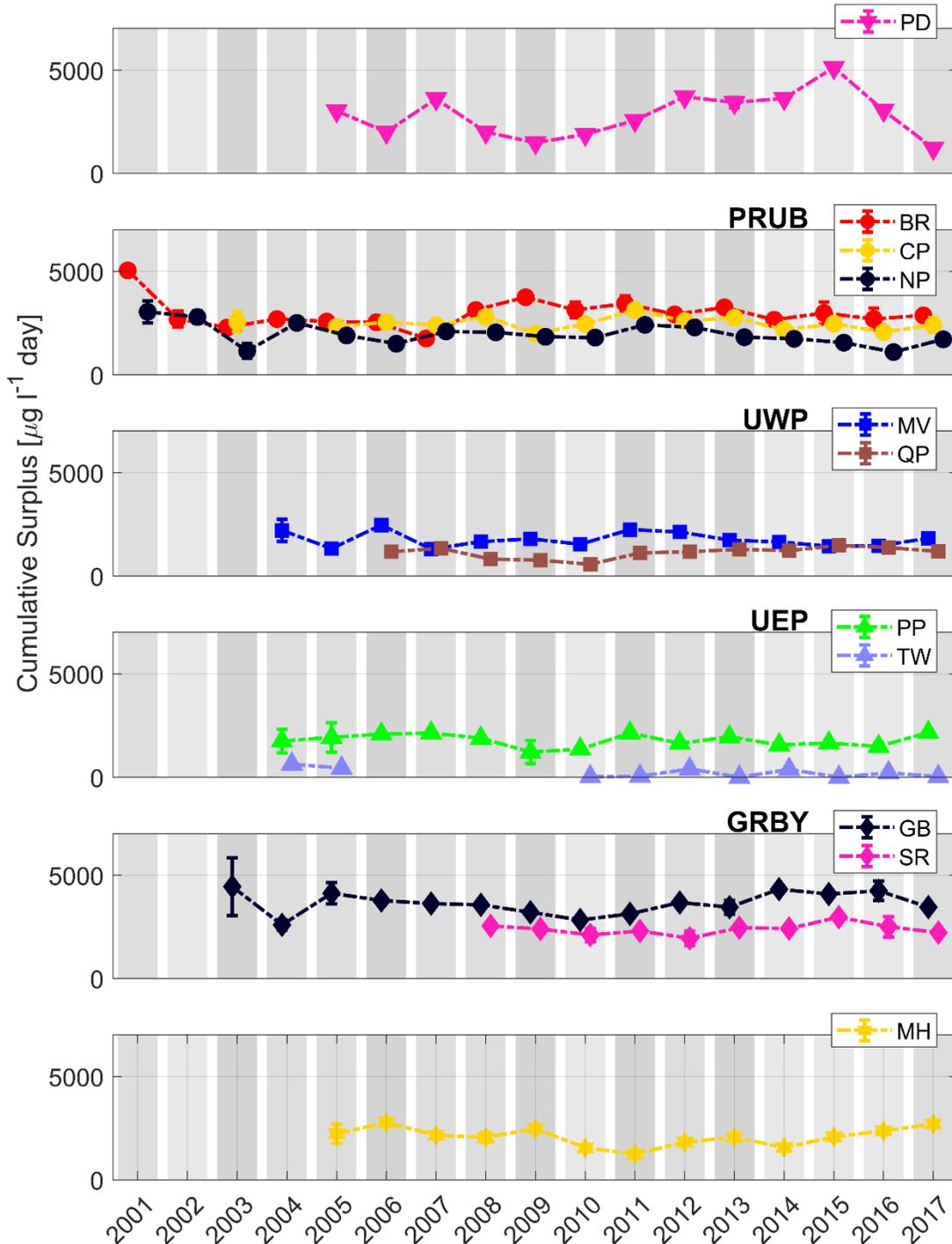


Figure 10-2. Same as Figure 10-1 except (intermediate) 9.4 $\mu\text{g L}^{-1}$ threshold.

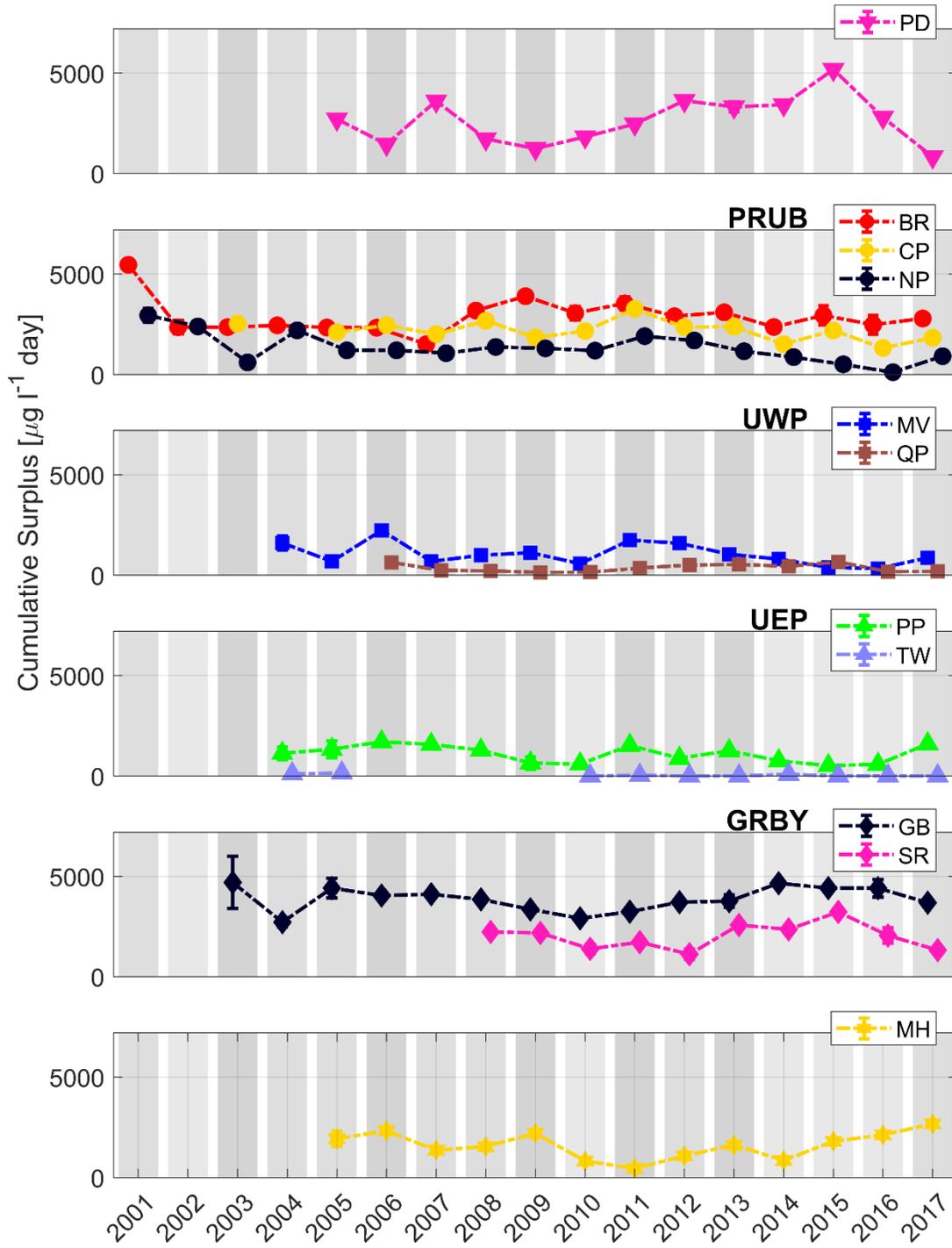


Figure 10-3. Same as Figure 10-1 except using (highest) 17.6 $\mu\text{g L}^{-1}$ threshold.

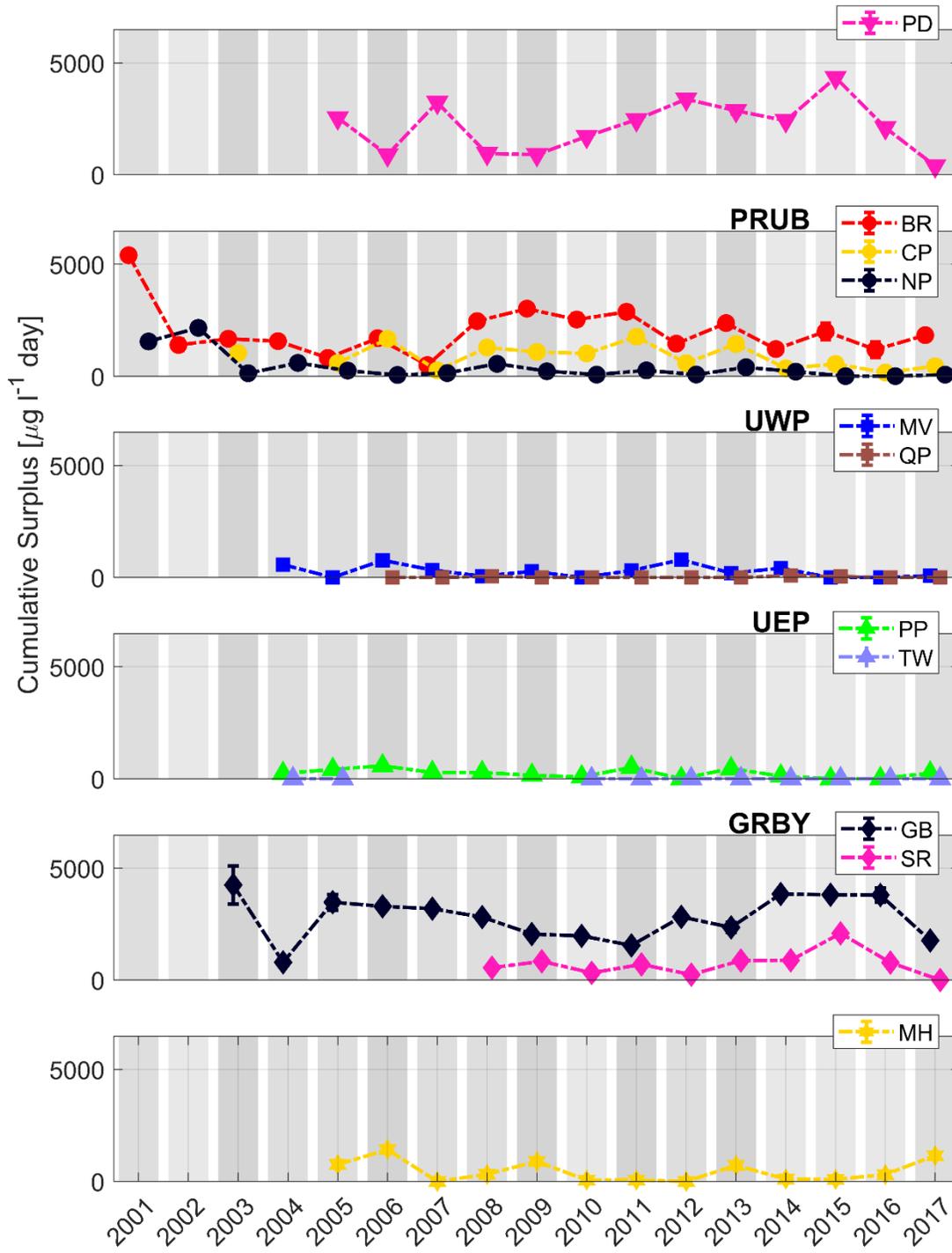
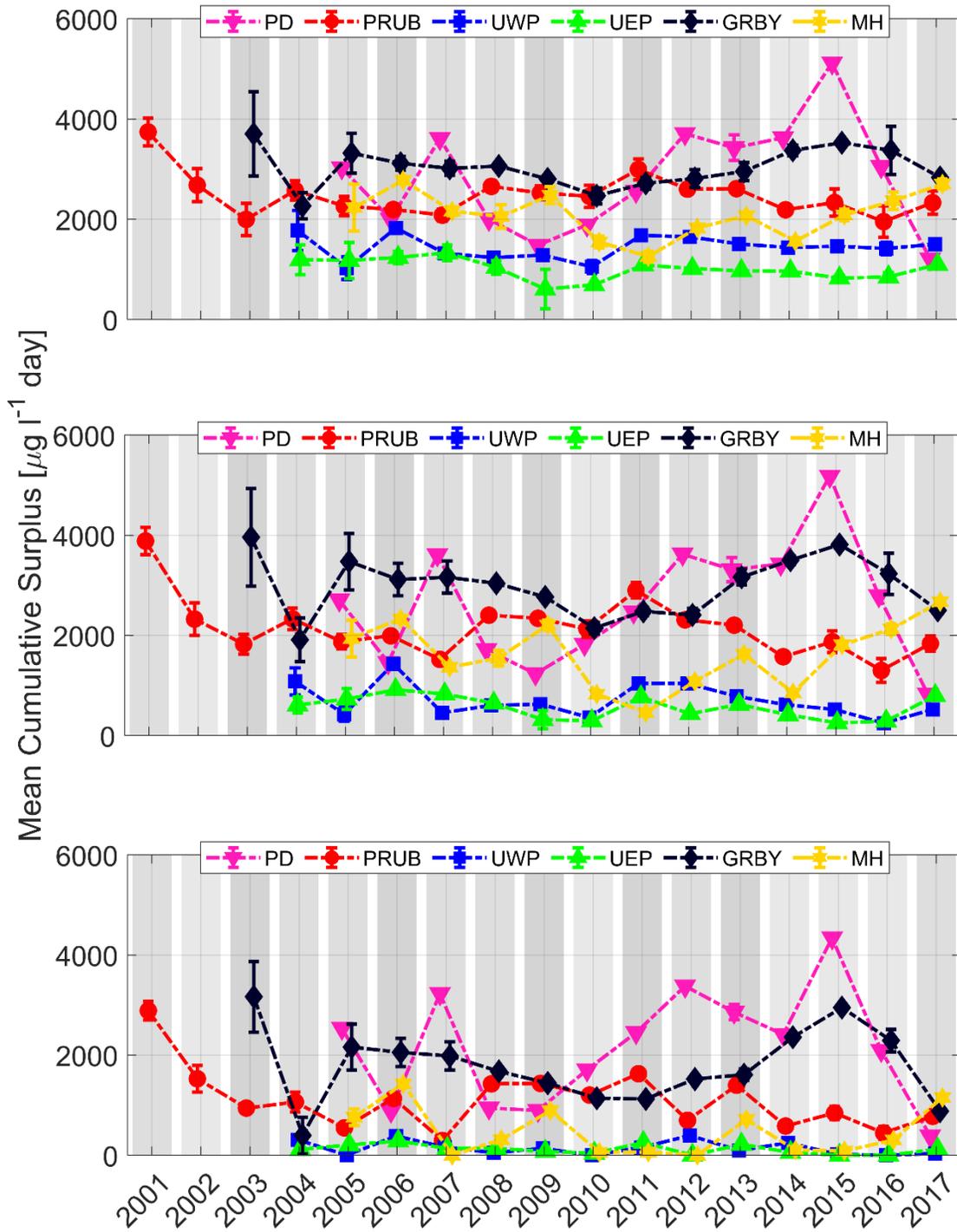


Figure 10-4. Chlorophyll Index, bay-wide thresholds, all three thresholds.

Top, middle, and bottom frames are results for averages of groups of stations for the 4.9 $\mu\text{g L}^{-1}$, 9.4 $\mu\text{g L}^{-1}$, and 17.6 $\mu\text{g L}^{-1}$ thresholds respectively (corresponding to individual station results in Figure 10-1, Figure 10-2, and Figure 10-3).



Task 11 Time Series and Spatial Survey Datasets Together: Chlorophyll

11.1 Scope

Objective (adapted from objective of Task 7). Improve knowledge of chlorophyll spatio-temporal characteristics. Shed light on the representativeness and uncertainty in the spatial information from the NBFSMN time series dataset, and in the temporal information from the spatial surveys dataset.

Methodology. The spatial surveys dataset (Prell et al, 2016) will be provided by Warren Prell, Brown University. Decorrelation timescales and lengthscales will be used to characterize the two datasets and determine relationships between them. **Expected outputs and outcomes.** The time series data provide dense temporal coverage at few sites; the spatial surveys achieve better spatial coverage and resolution, but occur infrequently. This analysis will better quantify their complementary strengths and weaknesses. This will be useful to improve understanding of the spatial and temporal characteristics of chlorophyll. It will also be helpful for cost-effective planning of future monitoring strategies.

11.2 Methods

Both the spatial survey and time series results are referred to as “chlorophyll”. The spatial survey results are concentrations of *chlorophyll a*. The time series results are *chlorophyll fluorescence* for *total chlorophyll*; as explained in Task 8 the approach for post-deployment recalibration using grab samples assumes differences between *chlorophyll a* and *total chlorophyll* are comparable to or less than measurement uncertainty.

The same methods were applied to chlorophyll concentration that were applied to oxygen, for both the time series and spatial survey datasets, in Task 7. The station locations in the spatial surveys are show in Figure 7-1. The long-term average chlorophyll concentrations at spatial survey stations, with corresponding ranks, are shown in Table 11-1 and Figure 11-1a. Surface chlorophyll does not vary spatially with station bathymetric depth as strongly as near-bottom oxygen does, as demonstrated in Task 7, which motivated using groups of stations with shallower and deeper depths there; however, for some analyses the shallower and deeper groups of stations are used here also. Table 11-2 and Figure 11-1b isolate the deeper group of stations.

Warren Prell (Brown U) provided results for Percent High-Chlorophyll Area that are parallel to the Percent Hypoxic Area results treated in Task 7. The Percent High-Chlorophyll Area results are relative to thresholds of $27.5 \mu\text{g L}^{-1}$, $14 \mu\text{g L}^{-1}$ and $8 \mu\text{g L}^{-1}$. These threshold values were computed as the 80th, 50th, and 20th percentiles of all spatial survey chlorophyll measurements from all surveys during the years 2005 through 2015.

New daily-resolution surplus-duration values (units $\mu\text{g L}^{-1} \text{ day}$) were computed for each fixed site time series, relative to the bay-wide thresholds developed in Task 10 (17.6 , 9.4 , and $4.9 \mu\text{g L}^{-1}$). Other than the day-long duration limit, these surplus-duration results are the same as used in prior calculations at seasonal timescales (Task 10). If no values that day are above the threshold the daily surplus-duration is zero; if some are above the threshold, there is a non-zero surplus-duration, proportional to how long the values are above the threshold and how far above the threshold they extend, because the surplus-duration is a measure of the area swept out above the threshold.

The potential recalibration (described in Task 8) of time series chlorophyll (used in this task) has not been implemented (nor had it been implemented on the data used in the SNBIW). If the recalibration is judged appropriate and carried out, results in this section may need to be revised.

See page 156 for information about supporting code and data files, and page 150 for how to use them.

11.3 Results

Background context. To improve understanding of the characteristics of spatial and temporal variations of chlorophyll in Narragansett Bay, two datasets are analyzed: vessel-based spatial surveys and fixed-site mooring time series. The datasets are known to be complementary, and have been used together only in limited ways. This analysis uses them together and further quantifies the relationships between them.

For the spatial surveys the main strengths are (a) sampling coverage that spans a larger geographic area of the bay, including both deeper and shallower areas, with finer spatial resolution (77 stations), and (b) sampling throughout the water column, with fine resolution (0.5 m) of vertical structure. The main weakness is how infrequently the surveys occur (5-7 times per season), which limits their ability to capture temporal variability.

For the fixed site time series the main strengths are the span of temporal coverage throughout the late spring to early fall season (mid-May to mid-October), and the high frequency 15-minute resolution sampling. The main weakness is that there are relatively fewer stations (11), located mostly in the deeper channelized areas of the bay, and the sensors are at near-surface and near-bottom depths which do not fully resolve the vertical structure of the water column.

In preparation to use the spatial survey and fixed site measurements together, it is useful first to investigate (a) spatial structure using solely the surveys, and (b) temporal structure using solely the fixed site time series.

Spatial surveys only. Using the spatial surveys, correlations (ρ , Pearson's coefficient; n value between 62 and 71, the number of surveys across all sampled years) were computed between individual values of near-surface chlorophyll from all surveys and all years at each of 12 reference stations, and corresponding individual values from the other reference and non-reference stations with median bathymetric depth greater than 4m. The 12 reference stations were chosen to include a range of depths and locations within the deeper channel areas and the shallower areas outside of channels.

Maps of ρ values for correlation between one of the 12 reference stations and other stations (Figure 11-2) present distinct spatial structure characterizing surface chlorophyll variability. For example, ρ values for northern stations within the Providence River Estuary (SR02, PRN01; station locations in Figure 7-1) drop to less than 0.7 at distances of a few km, and are weakly negative at most of the stations farther south. For stations farther south near and in the central upper bay (PRS07, UPB10) ρ values are positive everywhere and fall off over a somewhat larger distance. Stations yet farther south (UPB08, WPS09, EPN06, others) show some of the longest distances over which ρ values decrease, and some also have weakly negative ρ values in northern areas most distant from them. Differences in results for reference stations that are in deep channels do not differ strongly from those that are nearby in shallower areas, in contrast to the results for oxygen (Task 7), as expected because chlorophyll variability is controlled mainly by near-surface processes less sensitive to bathymetric structure.

Interestingly, for the reference station farthest in to Greenwich Bay (GRB04), as with other reference stations in the southern locations, the rho values at stations in the northern Providence River and Seekonk River are weakly negative.

The weakly negative rho values at the farther south stations relative to northernmost reference stations, and vice versa, suggest there is some tendency for the timing of phytoplankton blooms and post-bloom declines in the northern and southern portions to occur with alternating timing.

The same rho values can be plotted as a function of distance between each station pair (Figure 11-3), to help quantify the decorrelation length scales. For simplicity the distances used are direct point-to-point lengths between stations, whether the straight line path between them is wholly in water or not, and it should be borne in mind that for some station pairs the line passes over portions of land. As expected, a high degree of scatter exists in the results. The decorrelation length (taken to be where rho decreases to 0.7) ranges from about 1 to 10 km, with most values between about 4 and 7 km, and an average of about 5 km. These distances are substantially shorter than the corresponding results for deep oxygen (Task 7). As expected, in contrast to results for oxygen (Task 7), decorrelation scales between stations with similar bathymetric depths (solid symbols in Figure 11-3; bathymetric depths not more than 5 m different) are not noticeably different than for station pairs with differing bathymetric depths; the difference between relatively deeper stations (e.g., WPS09) and relatively shallower stations in embayments (e.g., GRB04) is modest.

Finally, the results can also be presented with the stations ordered by the rank of their long-term mean chlorophyll concentration (Figure 11-4). Marginally more organized patterns are present, in comparison to the distance-based results of Figure 11-3.

Fixed site time series measurements only. Decorrelation temporal scales were examined using lagged autocorrelations of daily-mean chlorophyll concentration from time series records at individual fixed sites (Figure 11-5). Rho decreases to 0.7 over timescales between about 1.5 and 3 days, with most values near about 2 days. The timescales are substantially shorter than the results for deep oxygen (Task 7). The longest values occur at the farthest south station (TW), and differences among result from all the other stations are modest.

Combining spatial surveys and time series. Now the two datasets are used together, in context of knowledge that based on the spatial surveys the decorrelation length scale is typically about 5 km and based on the time series its decorrelation time scale is about 2 days.

Time series measurements on the day of each spatial survey can be used to further assess decorrelation spatial structure, using the map style presentation. Using the daily mean (midnight to midnight) time series results on the date of the spatial survey, results (Figure 11-6) show many similarities to the maps presented in Figure 11-3 above. Rho values are highest at spatial survey stations located nearest to the time series station, and decrease substantially at distances of a few km or more. The pattern of negative rho values noted above is similarly apparent.

The rho values for all pairs of spatial survey stations and fixed site stations fill a 77x11 matrix (Figure 11-7, left). The 77 spatial survey stations are ordered by their rank all-survey mean chlorophyll, in the two groups (median of deepest-sampled depth greater than or less than 4 m); the deeper stations are ranks 1-57 and the shallower stations are ranks 58-77 (Table 11-2). The order of the fixed site stations is

the same as in Figure 11-6: Phillipsdale (PD), Bullocks Reach (BR), Conimicut Point (CP), North Prudence (NP), Greenwich Bay (GB), Sally Rock (SR), Mount View (MV), Quonset Point (QP), Poppasquash Point (PP), T-Wharf (TW), and Mount Hope (MH). There is a general pattern for higher-magnitude rho values at the deeper group of spatial survey stations (vertical axis, 1-57) relative to the shallower group of spatial survey stations (vertical axis, 58-77). In addition, the pattern noted above for mostly negative rho values is apparent in two portions of the matrix: the spatial survey stations with the highest mean chlorophyll (vertical axis 1 to about 10), which are mostly very shallow stations at the northernmost reach of the bay or in Greenwich Bay, and the Phillipsdale fixed site station.

To investigate sensitivity of these results to the use of daily-mean time series results, they were repeated, but using the 15-minute resolution time series measurements nearer in time (instead of their daily mean) to the time of each individual spatial survey profile. Two cases were treated: the hourly mean, using the hour centered on the time of the profile; and the individual 15-minute value nearest to the time of the profile. Results for the hourly mean case show modest differences from the daily-mean results (Figure 11-7, right); results for the individual 15-min case (not shown) are very similar to the hourly-mean results, as expected given that variability on timescales as short as an hour is not particularly pronounced. Differences (Figure 11-7, right) of the hourly mean results relative to the daily mean results do not have a pronounced or systematic structure but are noticeably more strongly negative and more strongly positive at MV and QP (7 and 8 on horizontal axis), respectively. These differences are moderately more prominent than for oxygen (Task 7), which is consistent with the noted variability of chlorophyll on shorter timescales than oxygen noted above.

Finally, to explore both spatial and temporal variability while also using both datasets together, lagged correlations were computed between daily-mean time series measurements and spatial survey results. Results are complex with substantial scatter, with the BR fixed site station representative (Figure 11-8). They show rho values at zero lag that are either larger than for nonzero lags or not substantially different from them, and rho that generally decreases as the magnitude of the lag increases up to a few days, beyond which there can be additional peaks with increasing lag magnitudes. This generally corroborates results from spatial correlations and lagged temporal autocorrelations presented above.

Percent high-chlorophyll area. The percent high-chlorophyll area results are shown in Figure 11-9 for each individual survey, relative to the three threshold levels of 27.5, 14.0, and 8.0 $\mu\text{g L}^{-1}$. Two characteristic features are apparent: inter-annual variability, and a reduction to lower values over the past 5-10 years as treatment plant nutrient load reductions have taken effect. These two characteristics are broadly similar to some results of the fixed site analysis (for example, chlorophyll index results of Task 10) and reinforce that inter-annual variability of surface chlorophyll is not as tightly linked to variations in river runoff as is the case for near-bottom oxygen (Task 7). The recent decline in the percent area results occurred over several years, whereas for the Chlorophyll Index the decline was apparent over a shorter period.

Results for the 27.5 $\mu\text{g L}^{-1}$ threshold are typically in the 5-20% range for all years up to about 2013 and nearly always less than 10% since then. Results for the 14.0 $\mu\text{g L}^{-1}$ threshold up to about 2013 range from about 10% to 95% and are mostly in the 30% to 60% range, with more recent values nearly all less than about 40%. Results for the 8.0 $\mu\text{g L}^{-1}$ threshold ranged from about 25% to 100% through 2013, after which they have been between about 15% and 85%.

Regressions are used to help quantify the potential relationship between percent high-chlorophyll area from individual surveys and the daily-mean chlorophyll from time series on the date of each survey (Figure 11-10). In individual frames of Figure 11-10, for each spatial survey there is a group of three symbols aligned at the same x-axis value, the daily-average surface chlorophyll at that station on the date of the spatial survey. For clarity of the figures, the regression lines are not shown; the r^2 values are shown in Table 11-3, bold if $p < 0.05$. For the high threshold ($27.5 \mu\text{g L}^{-1}$) the regression has $p < 0.05$ at about half of the stations, while this is the case for all but a few stations for the intermediate ($14.0 \mu\text{g L}^{-1}$) and low ($8.0 \mu\text{g L}^{-1}$) thresholds. The r^2 values for $p < 0.5$ cases are highest for the high and intermediate threshold, and highest at stations NP and CP where they reach 0.44 and 0.45; for stations farther south they generally are less than 0.3. In comparison to the equivalent regressions for deep oxygen (Task 7), although a similar set of mostly-deeper stations has the strongest statistical relationships, these results show notably fewer $p < 0.05$ cases and notably lower r^2 values. This likely is a reflection of the shorter decorrelation length for chlorophyll compared to oxygen.

To investigate the potential utility of the new daily surplus-duration results, the regressions of Figure 11-10 were repeated but using the respective daily surplus-duration results on the horizontal axis instead of the daily-mean chlorophyll. The results (Figure 11-11) are generally similar to those in Figure 11-10 and Table 11-3, with slightly lower r^2 (not shown) at many stations. This suggests the new daily-resolution surplus-duration could be a useful metric in addition to the daily means. This was not the case for oxygen, likely because for oxygen a higher proportion of the observations does not exceed the threshold to contribute to the new metric, weakening the regressions.

To investigate the relationship between the two datasets at the time scale of entire growing seasons, the average of the percent high-chlorophyll area results from all spatial surveys each year were regressed (Figure 11-12, Table 11-4) against the seasonal Chlorophyll Index computed from the time series (Task 10). The regressions are generally weak with $p < 0.05$ at only a few stations (Table 11-4), where r^2 values are no higher than 0.38. For the high, intermediate, and low thresholds the regression is significant only at a single station which is different for each threshold (BR, NP, and CP respectively). This indicates the strength of the relationship between the seasonal Chlorophyll Index and the season-averaged percent high-chlorophyll area is not very robust. This contrasts the relatively stronger equivalent relationship found for deep oxygen (Task 7), and again likely reflects the shorter spatial and temporal decorrelation scales for surface chlorophyll as compared to deep oxygen.

Representativeness of each dataset relative to the other. The representativeness of the spatial survey temporal sampling, relative to the temporal variability captured in the time series records, was investigated as follows. Using all years of data at each fixed-site station, the histogram of all daily-mean chlorophyll values in the time series was compared to the histogram from daily-mean values only on dates of the spatial surveys (Figure 11-13). The date range used in each year was from mid-May to mid-Oct. The Kolmogorov-Smirnov test at most (7) of the stations indicated the distributions were not statistically different, suggesting the spatial surveys are representative. The test indicated differences at 4 of the 11 stations (PD, NP, MV, MH), with survey results biased higher than the time series.

The representativeness of the spatial sampling in the time series, relative to the spatial variability captured in the spatial surveys, was investigated in a similar way (Figure 11-14). Histograms were constructed using spatial surveys measurements only. One uses results from all stations and the other only uses the 11 spatial survey stations located nearest to the 11 fixed site stations. The statistical test indicates the distributions are not different.

The Kolmogorov-Smirnoff test has revealed mixed results regarding representativeness of each dataset for the characteristics of the other. However, the differences are generally difficult to discern, as descriptive summary statistics and plots of each dataset reveal similar patterns independently. Other measures to assess representativeness should be investigated and applied.

Summary. The spatial surveys dataset alone captures detailed spatial structure, with station-to-station decorrelation length scale of about 5 km on average. The time series dataset alone captures detailed temporal structure, with decorrelation time scale of about 2 days on average. The two datasets have been used together here in various additional ways to further explore and quantify these spatiotemporal patterns.

A pattern of weakly negative correlation coefficients was identified between locations at the far northern (Providence River Estuary) and most distant down-bay (southern parts of Upper West Passage and Upper East Passage) portions of the sampled area. This suggests there is some tendency for the timing of phytoplankton blooms and post-bloom declines in the northern and southern portions to alternate.

Percent high-chlorophyll area results have been regressed against daily-mean fixed site time series results on the dates of the spatial surveys. Regressions using daily surplus-durations give similar results. The relationships are generally strongest for deeper stations in and near the channels, with correlation coefficients reaching as high as about 0.45. At the timescale of the entire growing season, the annual-mean percent high-chlorophyll area is weakly correlated to the Chlorophyll Index (Task 10). These relationships are all substantially weaker than their equivalent for deep oxygen (Task 7), which likely reflects that the decorrelation scales (both spatially and temporally) for chlorophyll are shorter than for oxygen.

With respect to design of future monitoring, these results can be used to help choose what types of monitoring to prioritize, given the goals for temporal and spatial characteristics to be monitored.

Table 11-1. Ordering and ranks of spatial survey stations by decreasing long-term mean chlorophyll concentration

Column 1: Station
 Column 2: Median deepest sampled depth [m]
 Column 3: Average (all years, all surveys) near-surface chlorophyll a concentration [$\mu\text{g L}^{-1}$]
 Column 4: Rank within 77 stations (higher rank is lower chlorophyll concentration)

GRB05	3.0	37.81	1	UPB12	3.0	14.06	45
GRB02	2.5	36.12	2	EPN06	14.0	13.99	46
PRC08b	3.5	32.41	3	BBT1	5.0	13.90	47
PRT01	2.5	32.19	4	GRB10	3.0	13.73	48
PRC02	13.5	31.70	5	WPS11	6.5	13.73	49
PRC07	6.5	30.84	6	EPN05	13.8	13.59	50
PRS07	14.0	30.18	7	GRB07	8.0	13.30	51
SR01	3.5	30.16	8	RWU03b	7.0	13.19	52
PRT03	2.5	29.98	9	BH07	5.0	13.18	53
PRC05b	14.0	29.57	10	UPB03	12.5	13.10	54
PRT02	2.5	29.52	11	BH10	5.0	13.04	55
PRT04	2.0	29.42	12	GRB12	3.0	12.83	56
SR02	5.5	28.49	13	GRB09	11.5	12.45	57
BRT3	8.0	26.30	14	UPB04b	6.0	12.44	58
PRN01	12.5	26.00	15	UPB02	7.0	12.37	59
BRT2	13.5	25.96	16	UPB01	18.5	12.08	60
PRS03	13.0	25.78	17	EPN02	19.0	11.94	61
GRB01	3.0	25.70	18	EPN03	14.5	11.63	62
SR03	6.5	25.68	19	RWU05	19.0	11.46	63
BRT4	6.5	25.36	20	RWU03	7.0	11.32	64
PRN03b	13.0	24.81	21	EPN04	8.0	11.02	65
GRB_C	2.5	24.54	22	WPS10	9.0	10.96	66
PRN08	7.5	23.19	23	EPN08b	7.0	10.93	67
GRB03	3.5	22.62	24	WPS05	10.5	10.69	68
BRT6	2.5	22.40	25	WPS07	10.0	10.55	69
GRB04	4.5	22.35	26	WPS09	6.5	10.11	70
UPB10	14.5	21.05	27	WPS04	10.5	9.57	71
BRT1	5.0	20.26	28	WPS14	7.0	9.48	72
PRN07	8.0	20.16	29	WPS01	12.5	8.81	73
GRB_D	3.5	19.95	30	EPS01	21.5	8.70	74
GRB_B	2.5	19.71	31	WPS08	2.5	8.68	75
GRB_A	4.0	18.63	32	WPS13	6.0	8.20	76
UPB11	13.5	18.02	33	WPS02	11.5	7.46	77
GRB06	4.5	17.85	34				
BBT2	6.5	17.15	35				
BBT3	9.0	16.64	36				
UPB08	14.0	16.63	37				
PRS04	3.0	16.40	38				
UPB06	5.5	15.99	39				
UPB05	15.0	15.49	40				
UPB13	6.5	15.41	41				
UPB07b	8.5	15.35	42				
BH05	4.5	14.71	43				
UPB09b	5.0	14.52	44				

Table 11-2. Ordering and ranks of spatial survey stations by decreasing long-term mean chlorophyll concentration, in two groups: deeper/shallower stations, where the median of the deepest-sampled depths is greater than or less than 4m.

Column 1:	Station		Average (all years, all surveys) near-surface chlorophyll a concentration [$\mu\text{g L}^{-1}$]					
Column 2:			Deeper group: Rank within 57 stations in deeper group					
Column 3:			Shallower group: 57 + rank within 20 stations in shallower group					
Deeper			UPB04b	12.44	39	GRB12	12.83	76
PRC02	31.70	1	UPB02	12.37	40	WPS08	8.68	77
PRC07	30.84	2	UPB01	12.08	41			
PRS07	30.18	3	EPN02	11.94	42			
PRC05b	29.57	4	EPN03	11.63	43			
SR02	28.49	5	RWU05	11.46	44			
BRT3	26.30	6	RWU03	11.32	45			
PRN01	26.00	7	EPN04	11.02	46			
BRT2	25.96	8	WPS10	10.96	47			
PRS03	25.78	9	EPN08b	10.93	48			
SR03	25.68	10	WPS05	10.69	49			
BRT4	25.36	11	WPS07	10.55	50			
PRN03b	24.81	12	WPS09	10.11	51			
PRN08	23.19	13	WPS04	9.57	52			
GRB04	22.35	14	WPS14	9.48	53			
UPB10	21.05	15	WPS01	8.81	54			
BRT1	20.26	16	EPS01	8.70	55			
PRN07	20.16	17	WPS13	8.20	56			
UPB11	18.02	18	WPS02	7.46	57			
GRB06	17.85	19						
BBT2	17.15	20	Shallower					
BBT3	16.64	21	GRB05	37.81	58			
UPB08	16.63	22	GRB02	36.12	59			
UPB06	15.99	23	PRC08b	32.41	60			
UPB05	15.49	24	PRT01	32.19	61			
UPB13	15.41	25	SR01	30.16	62			
UPB07b	15.35	26	PRT03	29.98	63			
BH05	14.71	27	PRT02	29.52	64			
UPB09b	14.52	28	PRT04	29.42	65			
EPN06	13.99	29	GRB01	25.70	66			
BBT1	13.90	30	GRB_C	24.54	67			
WPS11	13.73	31	GRB03	22.62	68			
EPN05	13.59	32	BRT6	22.40	69			
GRB07	13.30	33	GRB_D	19.95	70			
RWU03b	13.19	34	GRB_B	19.71	71			
BH07	13.18	35	GRB_A	18.63	72			
UPB03	13.10	36	PRS04	16.40	73			
BH10	13.04	37	UPB12	14.06	74			
GRB09	12.45	38	GRB10	13.73	75			

Table 11-3. Regression r^2 between daily-mean fixed site surface chlorophyll and individual survey percent high-chlorophyll area, corresponding to Figure 11-10. Bold if $p < 0.05$.

Station	27.5 $\mu\text{g L}^{-1}$	14.0 $\mu\text{g L}^{-1}$	8 $\mu\text{g L}^{-1}$
PD	0.06	0.09	0.04
BR	0.26	0.2	0.13
CP	0.44	0.44	0.29
NP	0.15	0.45	0.38
GB	0.00	0.02	0.05
SR	0.04	0.12	0.1
MV	0.17	0.31	0.27
QP	0.02	0.25	0.28
PP	0.05	0.3	0.29
TW	0.01	0.24	0.17
MH	0.09	0.06	0.04

Table 11-4. Regression r^2 between seasonal Chlorophyll Index from fixed site time series and seasonally-averaged percent high-chlorophyll area from spatial surveys, corresponding to Figure 11-12. Bold values have $p < 0.05$. The missing value indicates the regression could not be computed because there were no non-zero values of the index.

A (red in Figure 11-12): Chlorophyll Index rel. 17.6 $\mu\text{g l}^{-1}$, percent high-chlorophyll area rel. 27.5 $\mu\text{g l}^{-1}$
 B (blue in Figure 11-12): Chlorophyll Index rel. 9.4 $\mu\text{g l}^{-1}$, percent high-chlorophyll area rel. 14.0 $\mu\text{g l}^{-1}$
 C (green in Figure 11-12): Chlorophyll Index rel. 4.9 $\mu\text{g l}^{-1}$, percent high-chlorophyll area rel. 8.0 $\mu\text{g l}^{-1}$

Station	A	B	C
PD	0.25	0.13	0.00
BR	0.38	0.28	0.08
CP	0.04	0.25	0.35
NP	0.04	0.37	0.22
GB	0.22	0.2	0.07
SR	0.01	0.01	0.00
MV	0.33	0.3	0.04
QP	0.22	0.00	0.05
PP	0.01	0.03	0.28
TW	0.06	0.04	
MH	0.00	0.04	0.01

Figure 11-1. Rank of mean near-surface chlorophyll concentration at spatial survey stations.

a. All 77 stations (see Table 11-1). b. Subgroup of 57 stations with median near-bottom sample depth greater than 4m (see Table 11-2).

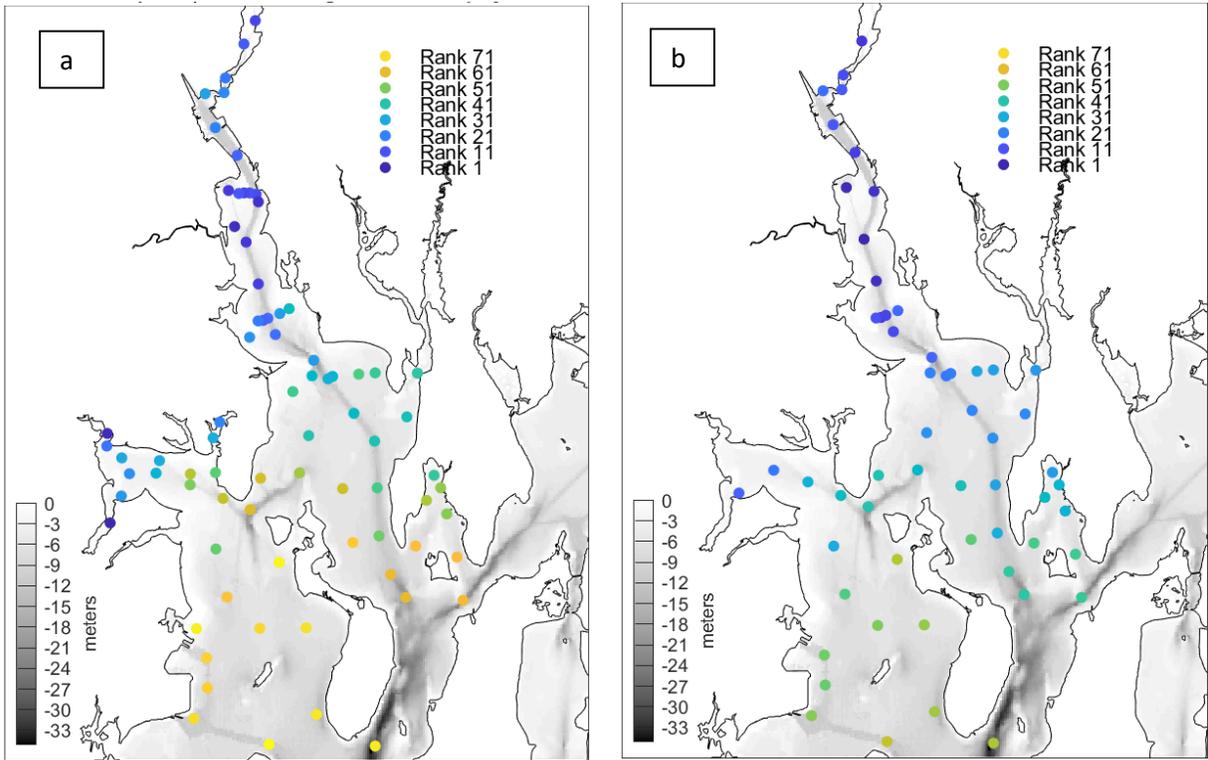


Figure 11-2. Spatial survey correlation results.

(a) Reference stations SR02, PRN01, PRS07, UPB10.

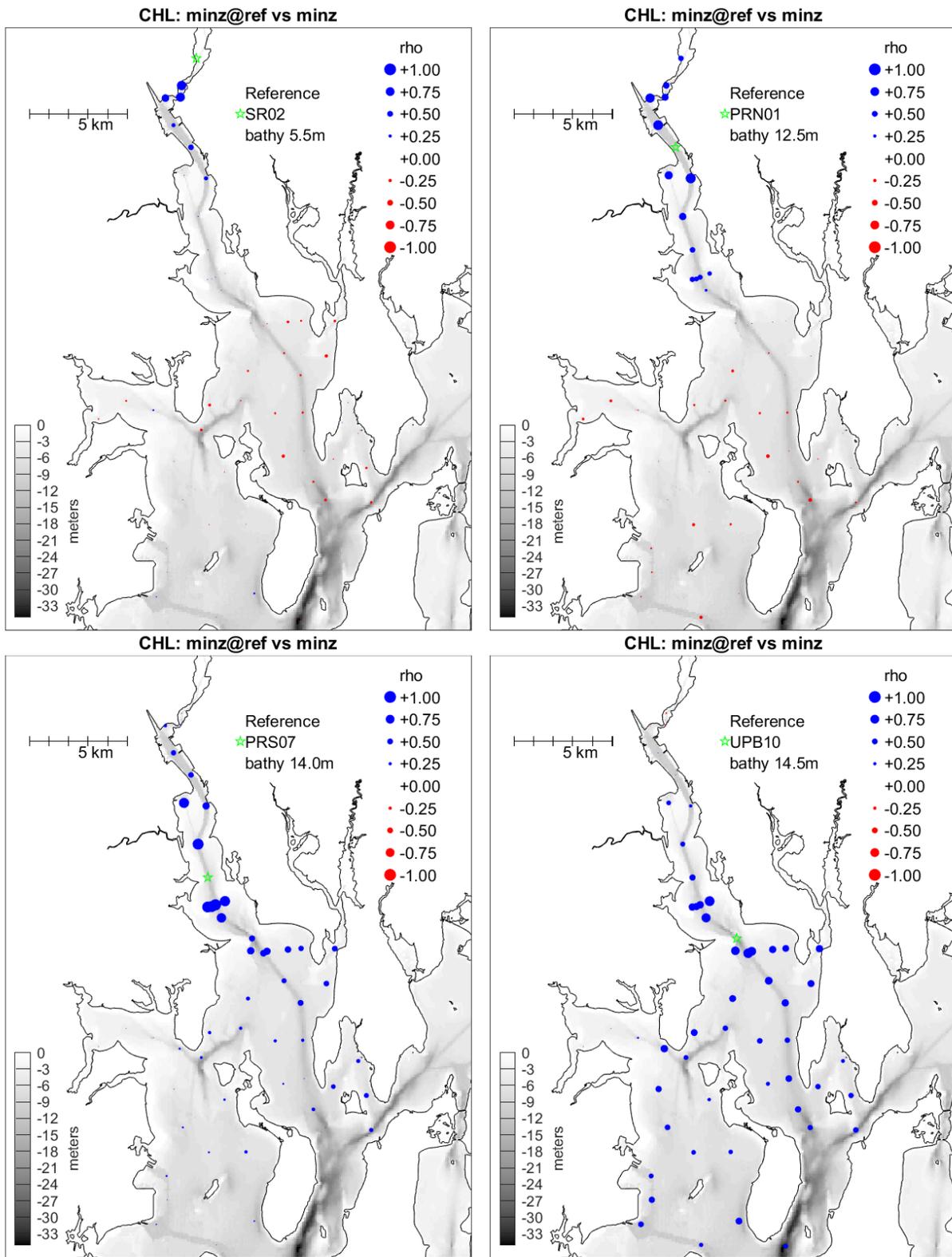


Figure 11-2. (b) Reference stations UPB08, GRB09, GRB04, WPS09.

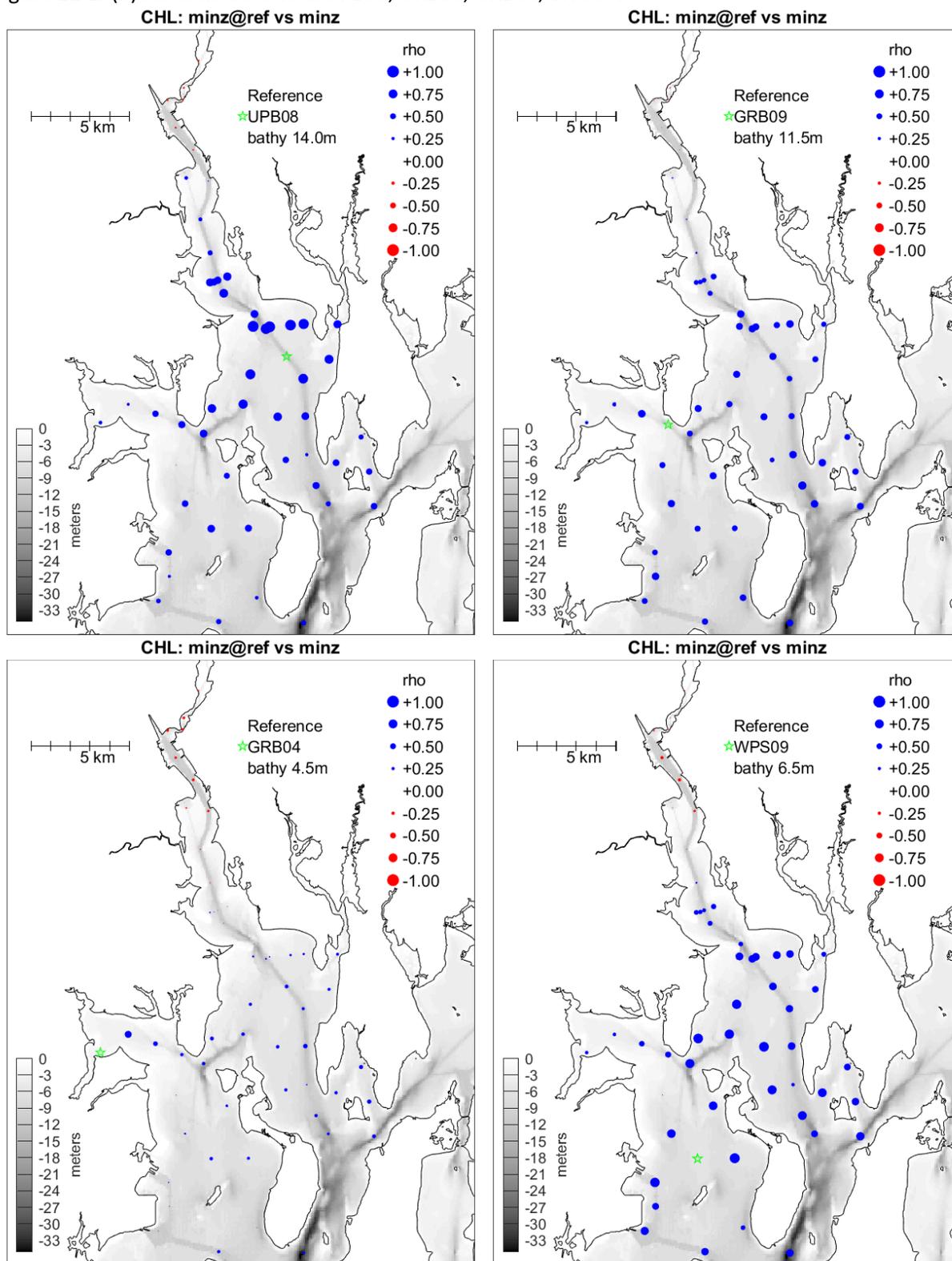


Figure 11-2. (c) Reference stations UPB06, EPN06, EPS01, RWU05.

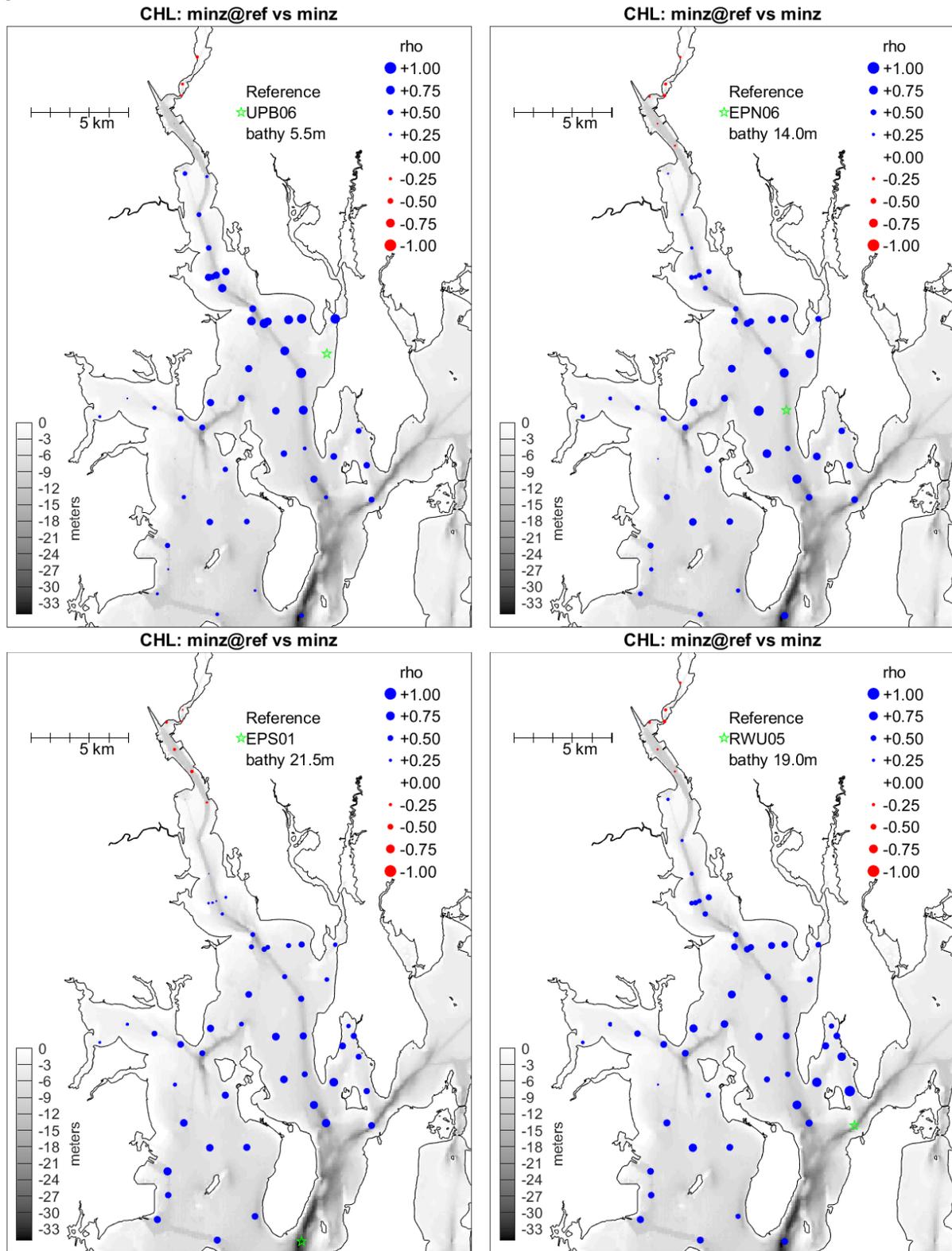


Figure 11-3. Spatial surveys, decorrelation length scales.

Correlation of temporal variability (all surveys, all years) at 12 reference spatial survey stations (listed in legend, with their bathymetric depth) to other 56 deeper spatial survey stations, as a function of distance between stations in the pair. At zero distance all values are 1. Solid symbols are station pairs for which the difference in bathymetric depths is less than 5 m. Lines connect solid symbols for stations WPS09 and GRB04, which are representative of deeper and shallower stations, respectively.

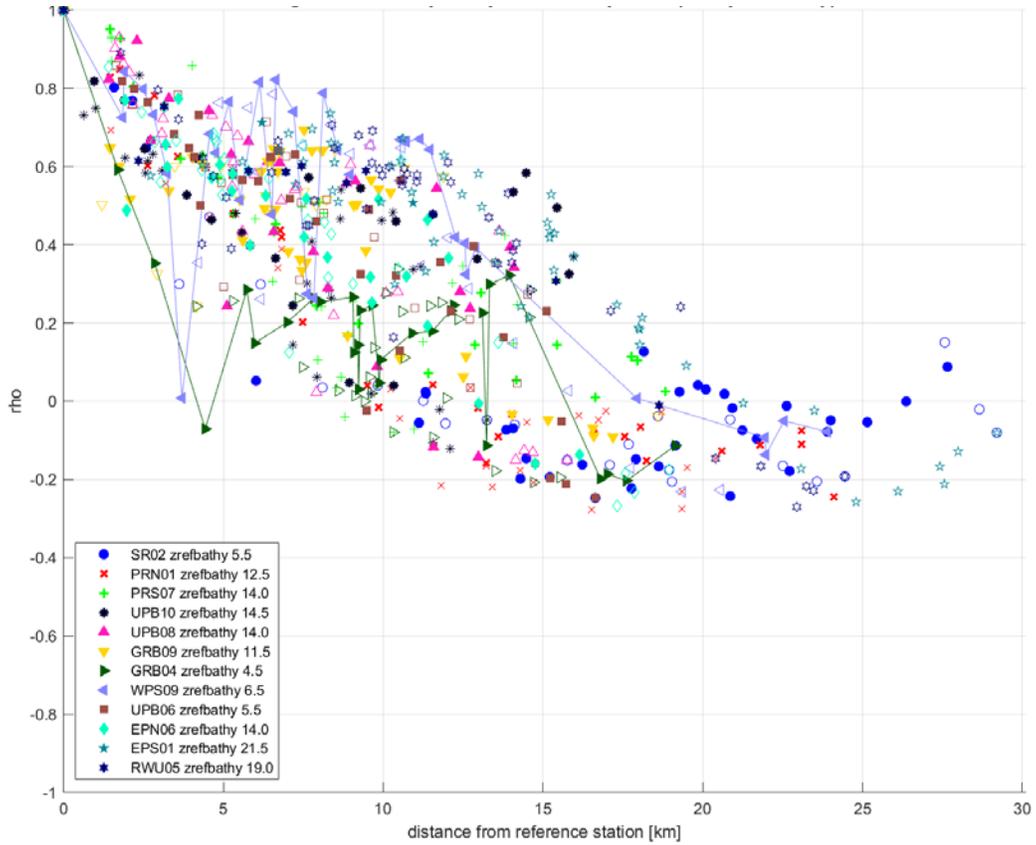


Figure 11-4. Same correlation results as in Figure 11-3 but here they are ordered on the horizontal axis by relative rank (shown in Figure 11-1) of station near-surface chlorophyll.

(Solid symbols are again station pairs for which the difference in bathymetric depths is less than 5 m.) The horizontal axis is the difference between the near-surface chlorophyll rank of each of the other 56 stations and the near-surface chlorophyll rank of the reference station. For example, reference station SR02 (blue solid circles) has near-bottom DO rank of 5 (Table 11-1) so there are 4 results that are on the negative portion of the horizontal axis; the station with the most extreme near-bottom DO rank of 57 (WPS08) appears farthest to the right at 52 (its rank difference relative to that of SR02) on the horizontal axis.

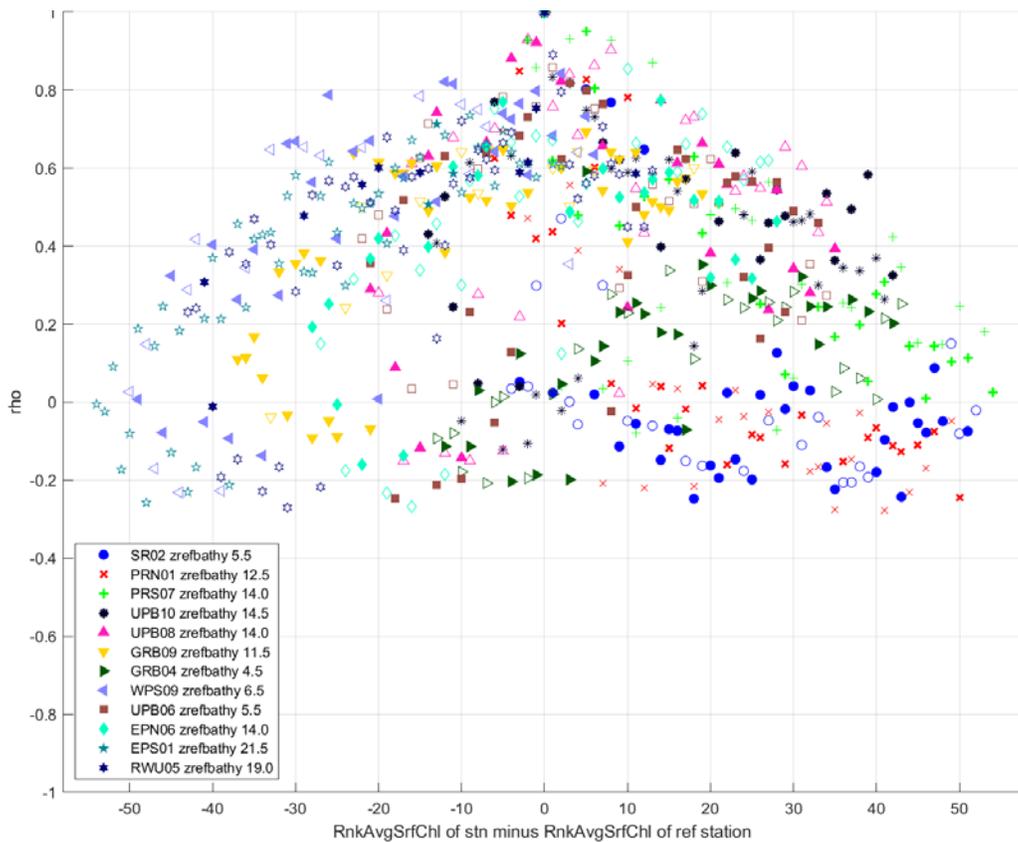


Figure 11-5. Decorrelation time scale for surface chlorophyll concentration from lagged autocorrelations of time series measurements at individual fixed site locations.

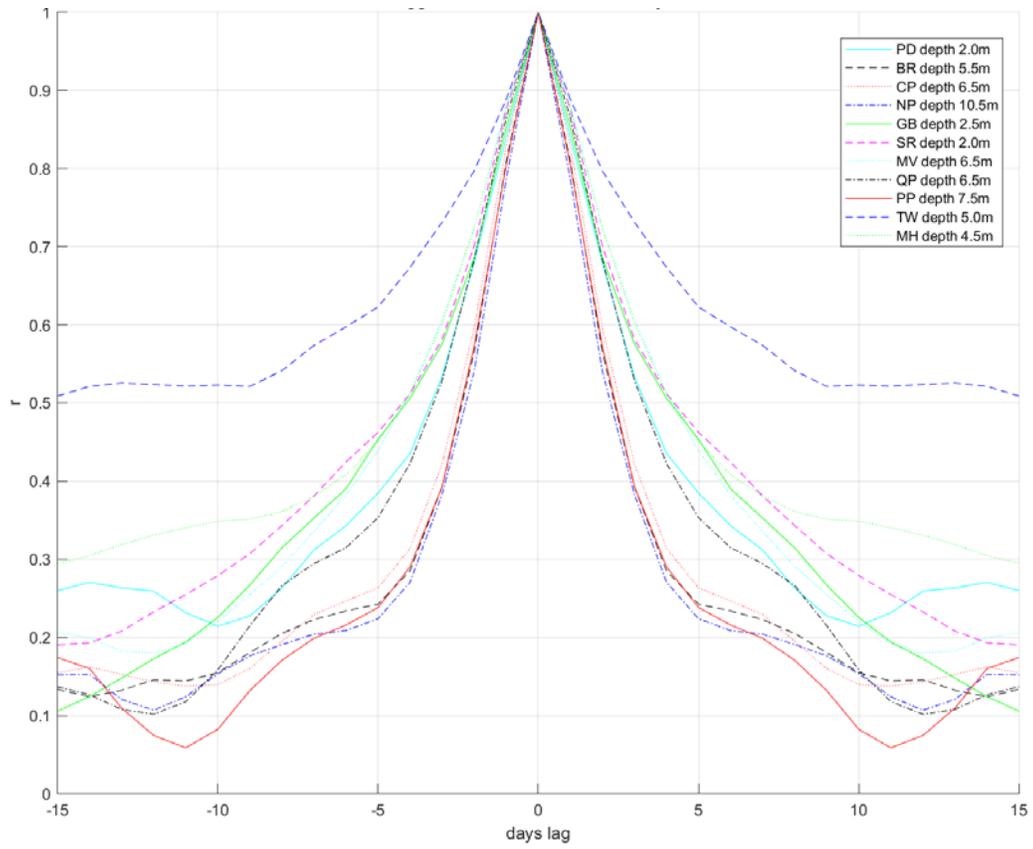


Figure 11-6. Correlations of spatial survey surface chlorophyll concentration and daily mean of 15-minute temporal resolution time series on date of spatial survey.

(a) Fixed site stations PD, BR, CP, NP.

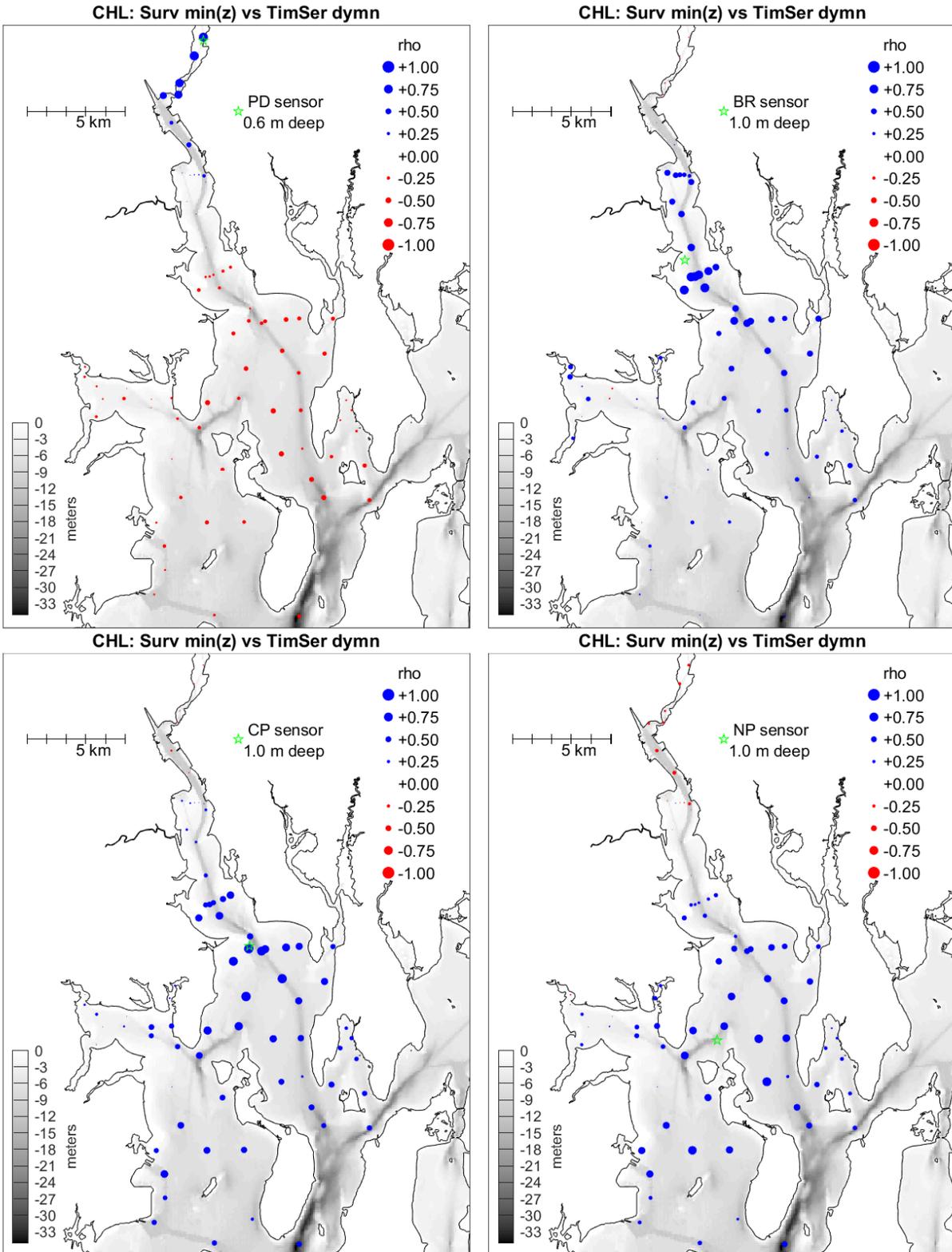


Figure 11-6. (b) Fixed site stations GB, SR, MV, QP.

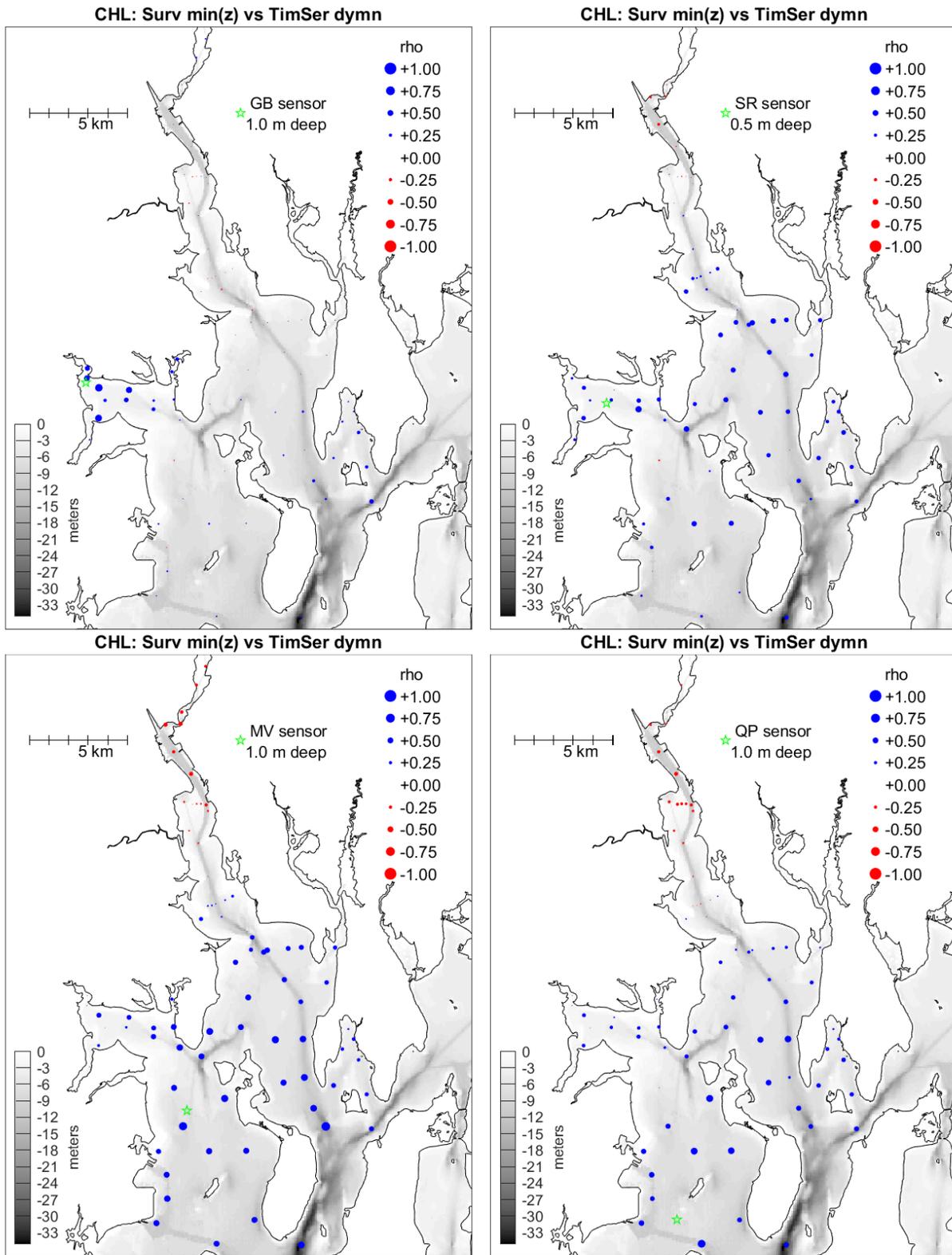


Figure 11-6. (c) Fixed site stations PP, TW, MH.

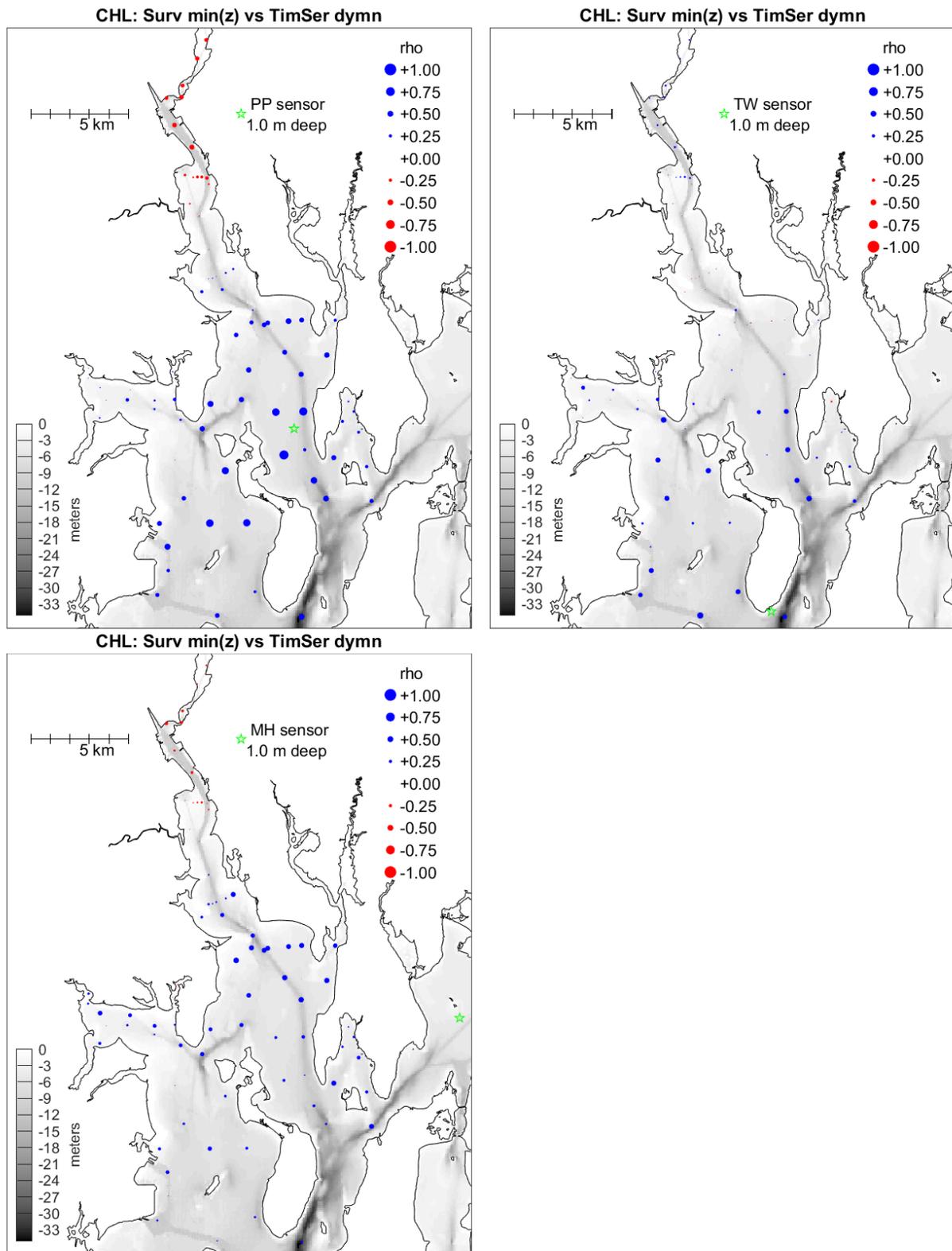


Figure 11-7. Correlation matrix of spatial surveys and time series.

Left: Matrix of rho values from results (shown in Figure 11-6) for correlation between spatial survey near-surface chlorophyll concentration and daily-mean of time series at fixed site station on the date of each spatial survey. Vertical axis is spatial survey station rank (Figure 11-1b), by decreasing (1 = highest) all-survey mean surface chlorophyll concentration, with deeper stations from 1 to 57 and shallower stations from 58 to 77. Horizontal axis is fixed site station, in order shown in Figure 11-6). Right: Difference from matrix at left, when hourly mean of time series, for hour centered on time of each spatial survey profile, is used instead of daily mean.

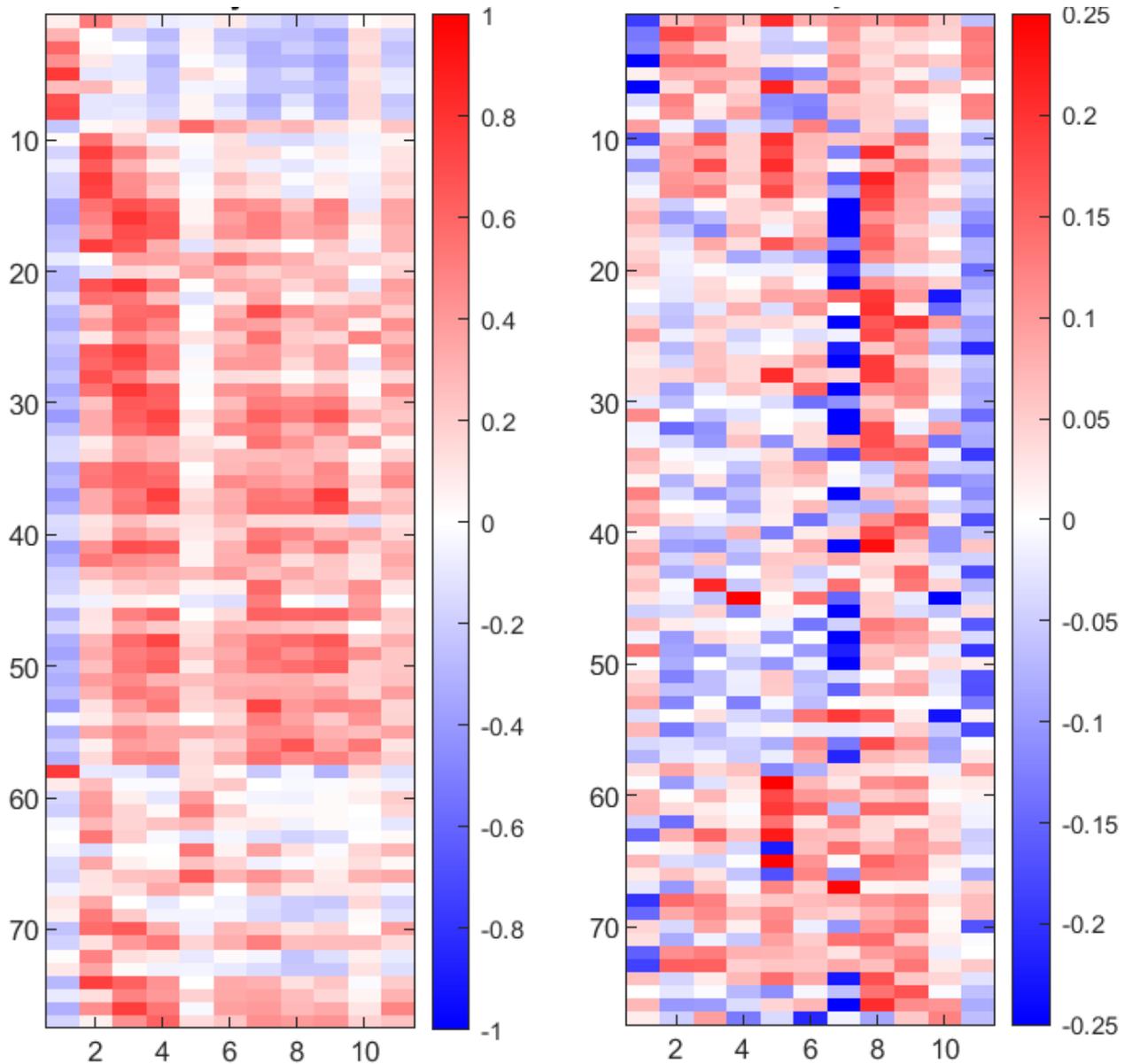


Figure 11-8. Lagged correlations between spatial surveys and time series.

Bullocks Reach (BR) site (5.5 m sensor depth) daily-mean time series measurements and spatial survey measurements, near-surface chlorophyll concentration. Correlation at each lag (positive or negative) is computed independently. Twelve representative spatial survey stations are included, as listed in the legend with their median near-bottom sample depth and distance in km from BR.

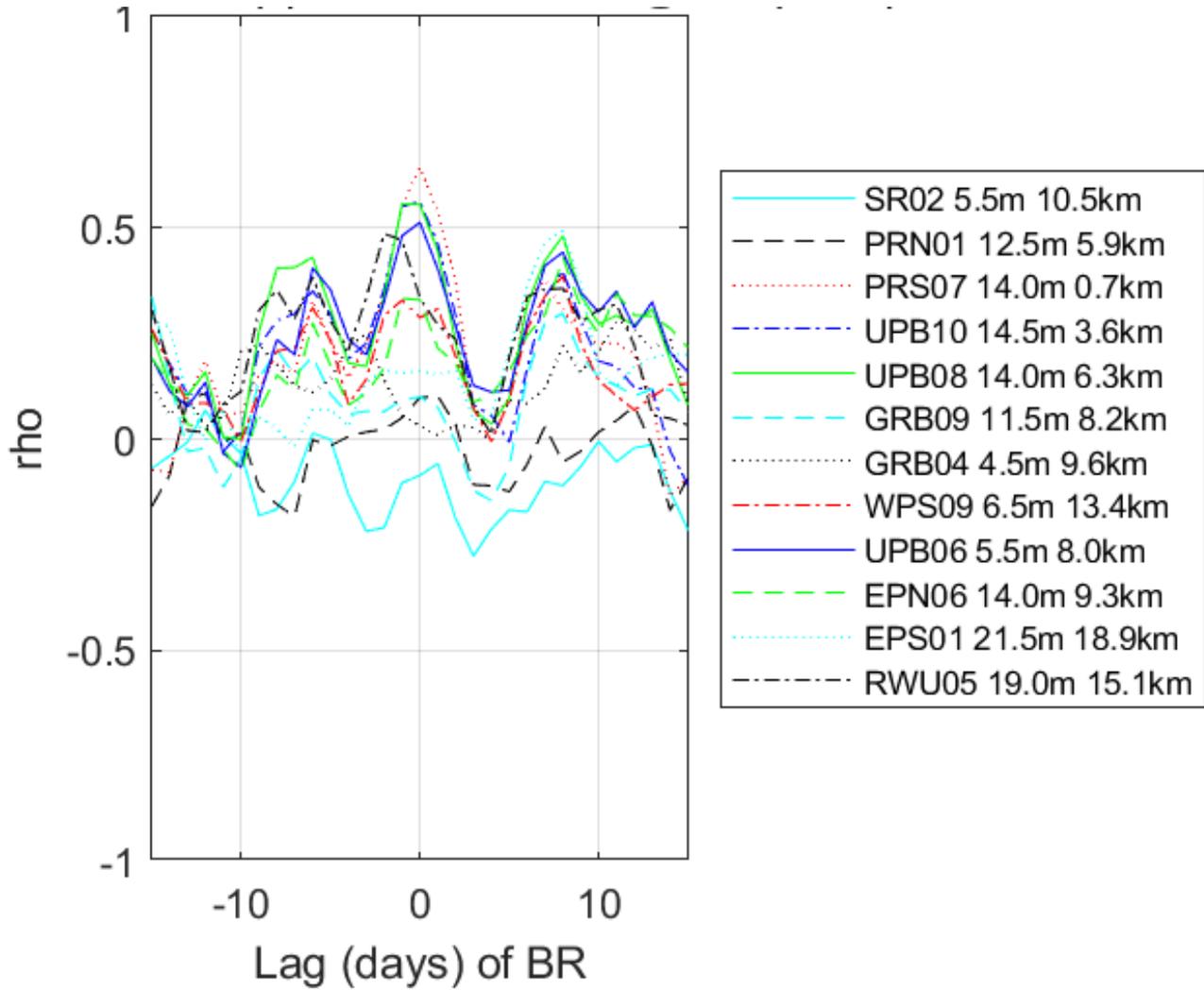


Figure 11-9 Percent high-chlorophyll area from individual spatial survey dates, thresholds $4.9 \mu\text{g L}^{-1}$ (green), $8 \mu\text{g L}^{-1}$ (blue), and $27.5 \mu\text{g L}^{-1}$ (red).

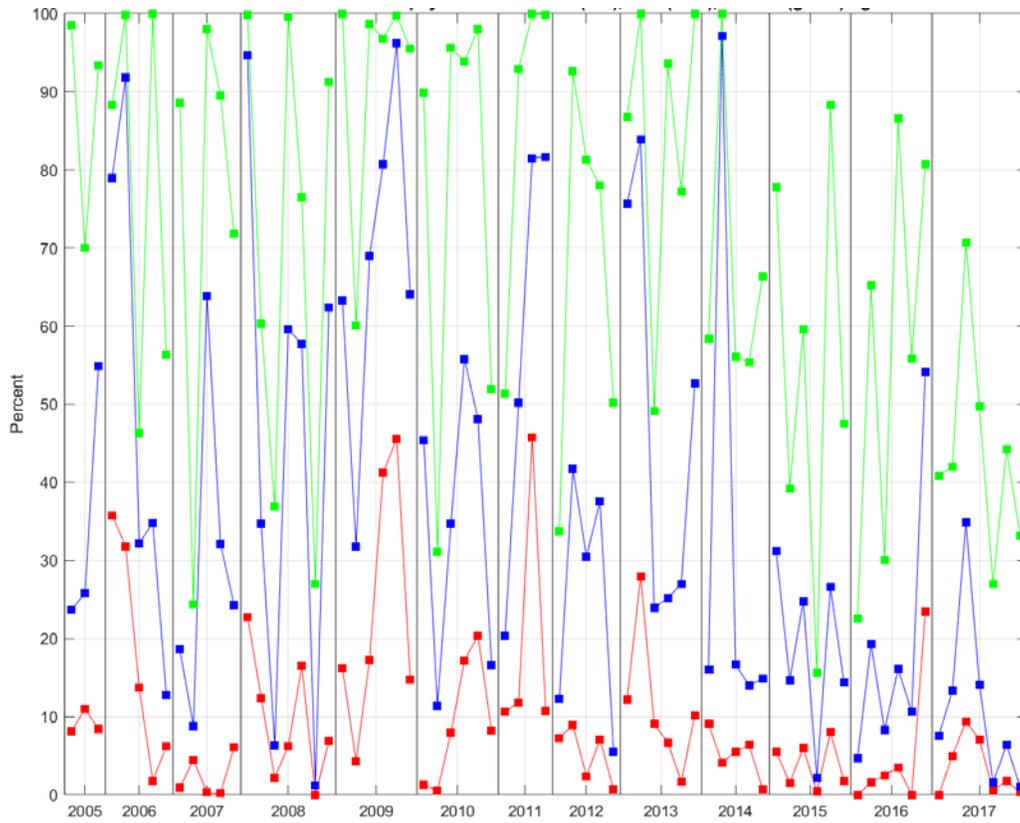


Figure 11-10. Percent high-chlorophyll area of spatial surveys vs time series.

Percent high-chlorophyll area on individual survey dates relative to the 27.5 (red), 14 (blue), and 8 (green) $\mu\text{g L}^{-1}$ thresholds, plotted as a function of daily-mean surface chlorophyll time series concentration at each of the 11 fixed site stations on the dates of the surveys. Note the horizontal axes ranges differ from station to station (zero to 100 $\mu\text{g L}^{-1}$ for PD, 75 $\mu\text{g L}^{-1}$ for BR and GB, and 40 $\mu\text{g L}^{-1}$ for other stations). Correlation coefficient results are in Table 11-3.

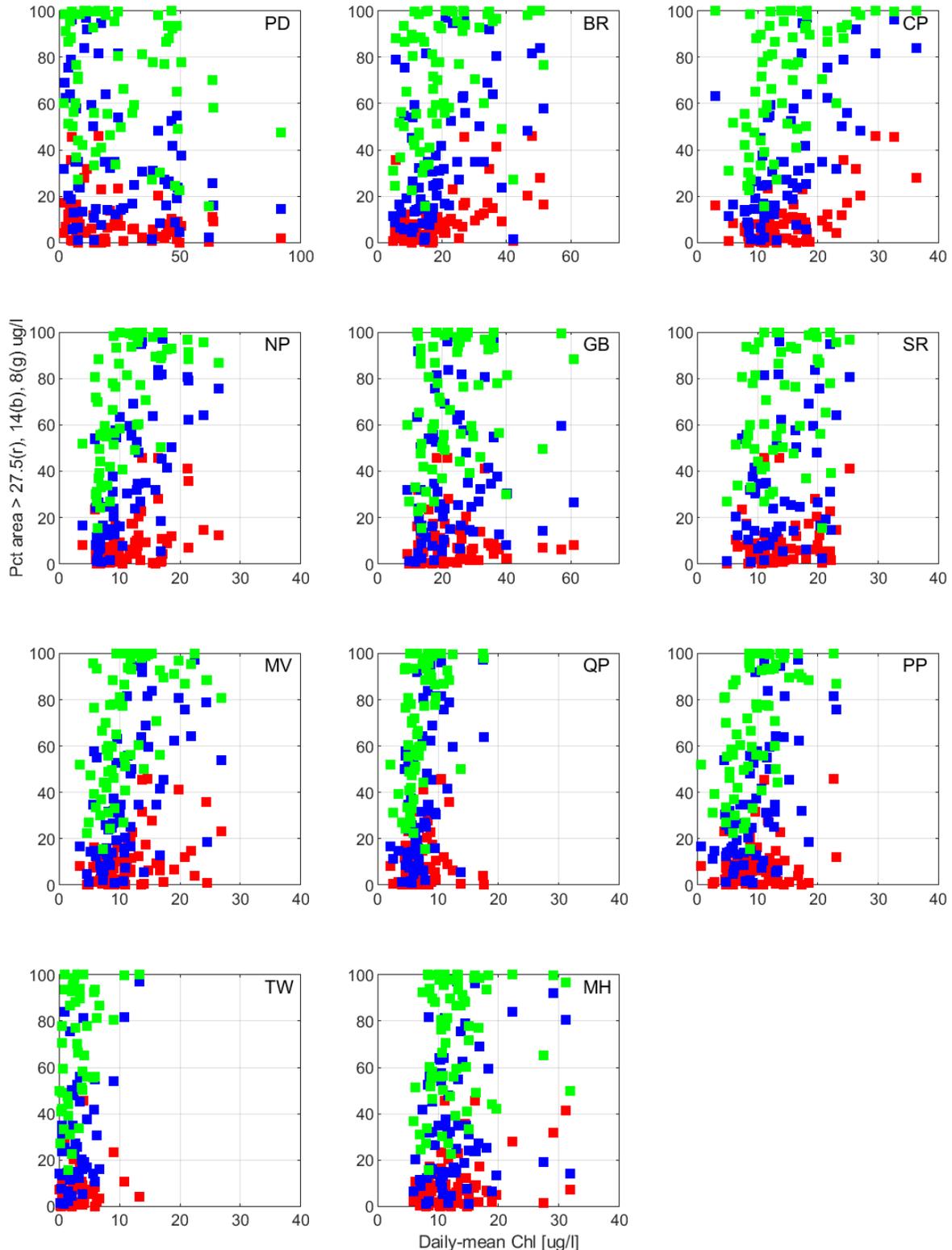


Figure 11-11. Same as Figure 11-10 but using daily surplus-duration.

horizontal axis is the daily surplus-duration of chlorophyll relative to 17.6 (red), 9.4 (blue), and 4.9 (green) $\mu\text{g L}^{-1}$ thresholds from Task 10. Note the horizontal axis ranges are different (zero to 95 $\mu\text{g L}^{-1}$ dy for PD, 60 $\mu\text{g L}^{-1}$ dy for BR and GB, and 35 $\mu\text{g L}^{-1}$ dy for other stations).

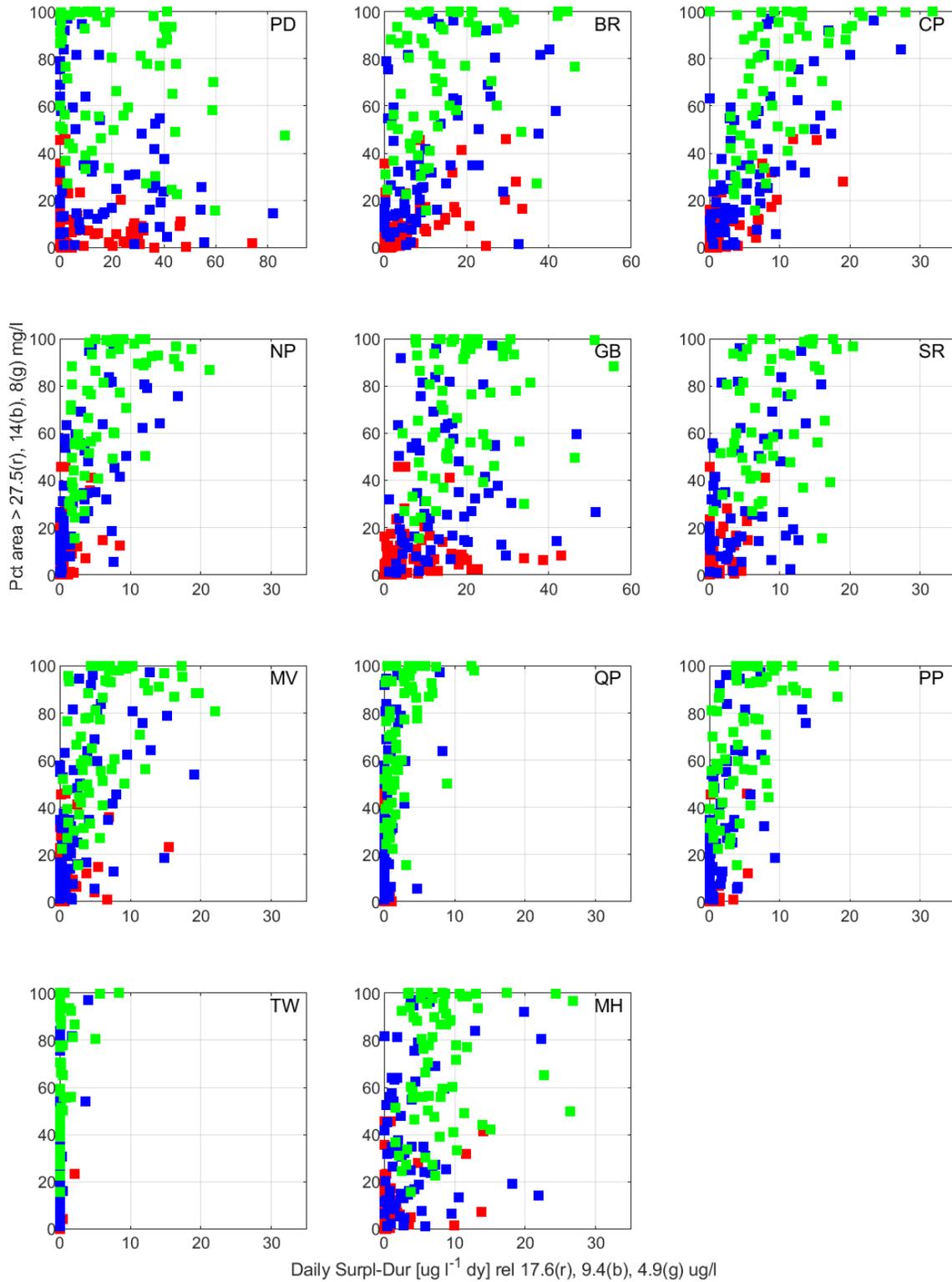


Figure 11-12. Seasonal Chlorophyll Index (time series) vs season-mean high-chlorophyll area (spatial surveys).

Chlorophyll Index as a function of seasonally-averaged percent high-chlorophyll area from spatial surveys. Red/blue/green symbols show Chlorophyll Index relative to 17.6/9.4/4.9 $\mu\text{g L}^{-1}$ threshold, in $\mu\text{g L}^{-1}$ day normalized by 5500/5500/4500 $\mu\text{g L}^{-1}$ day, vs seasonal-mean percent hypoxic area relative to 27.5/14.0/8.0 $\mu\text{g L}^{-1}$ threshold, normalized by 30%/75%/100%. Correlation coefficient results are in Table 11-4.

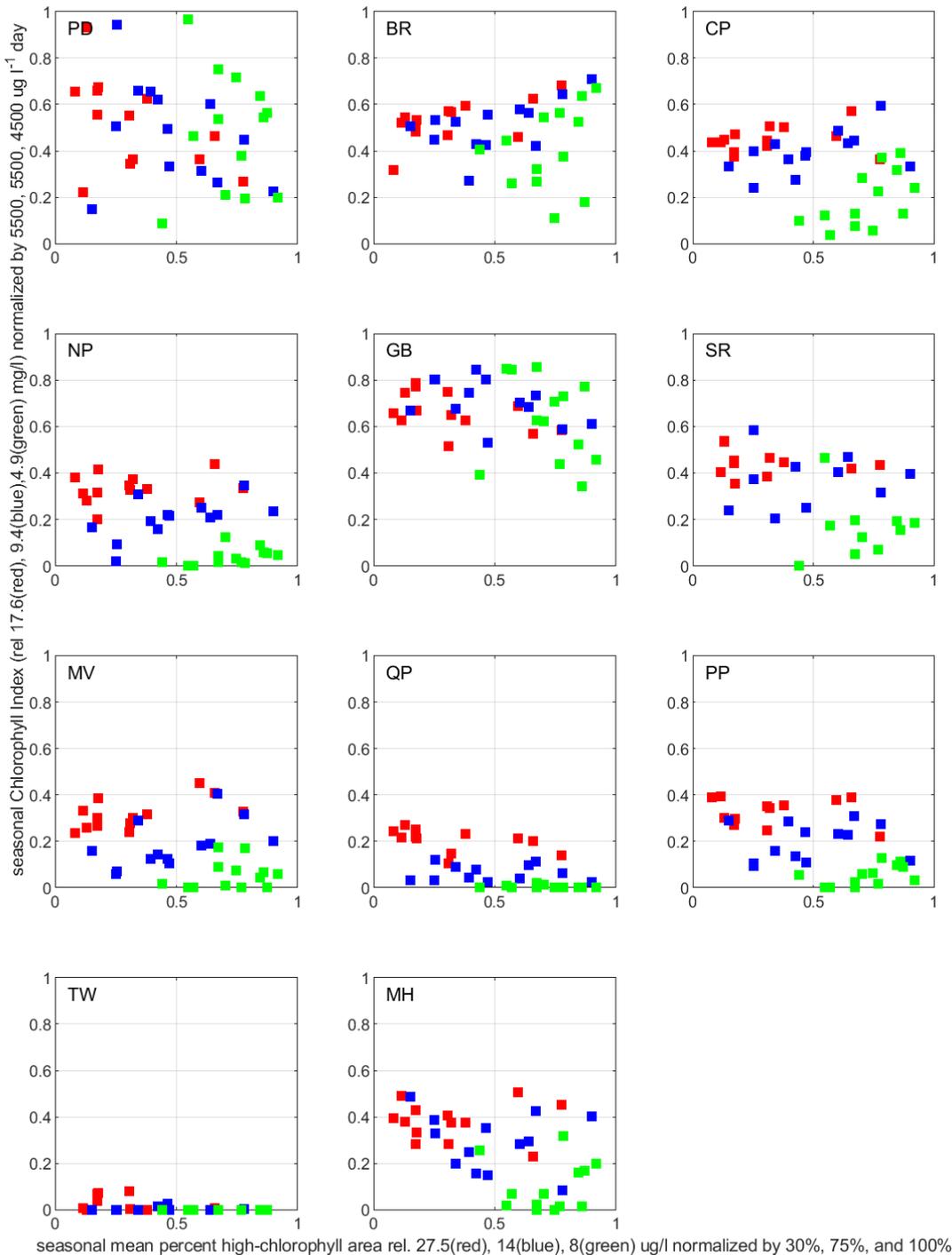


Figure 11-13. Histograms to gauge representativeness of spatial surveys.

Daily-mean near-surface chlorophyll concentration (horizontal axis) from fixed site time series using all years, May 15 through Oct 14: all dates sampled in time series (blue, left vertical axis) and only dates sampled by spatial surveys (orange, right vertical axis). Vertical axes are number of daily-mean values in each bin of DO ranges. Results of Kolmogorov-Smirnov test shown in upper right below station code.

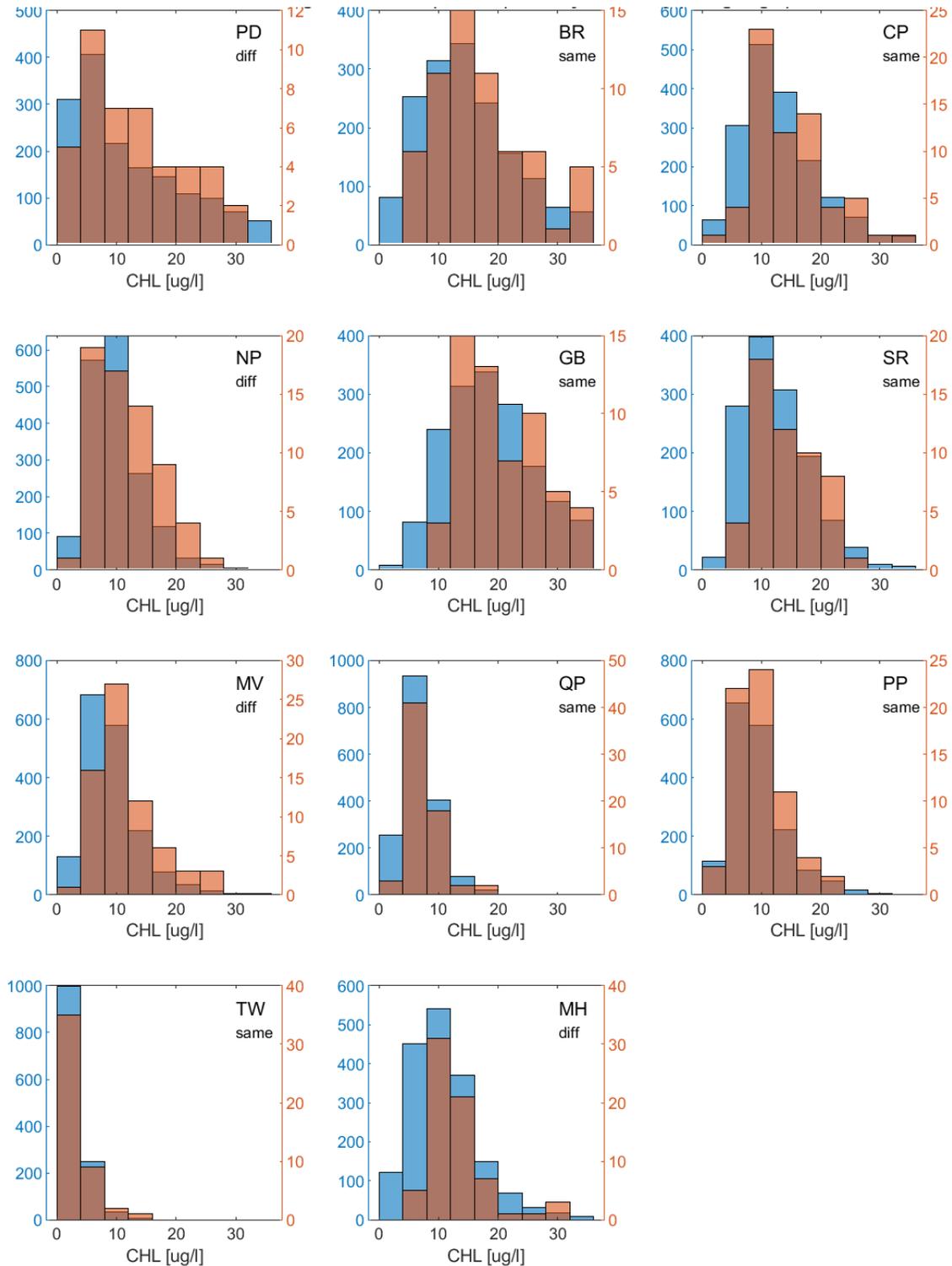
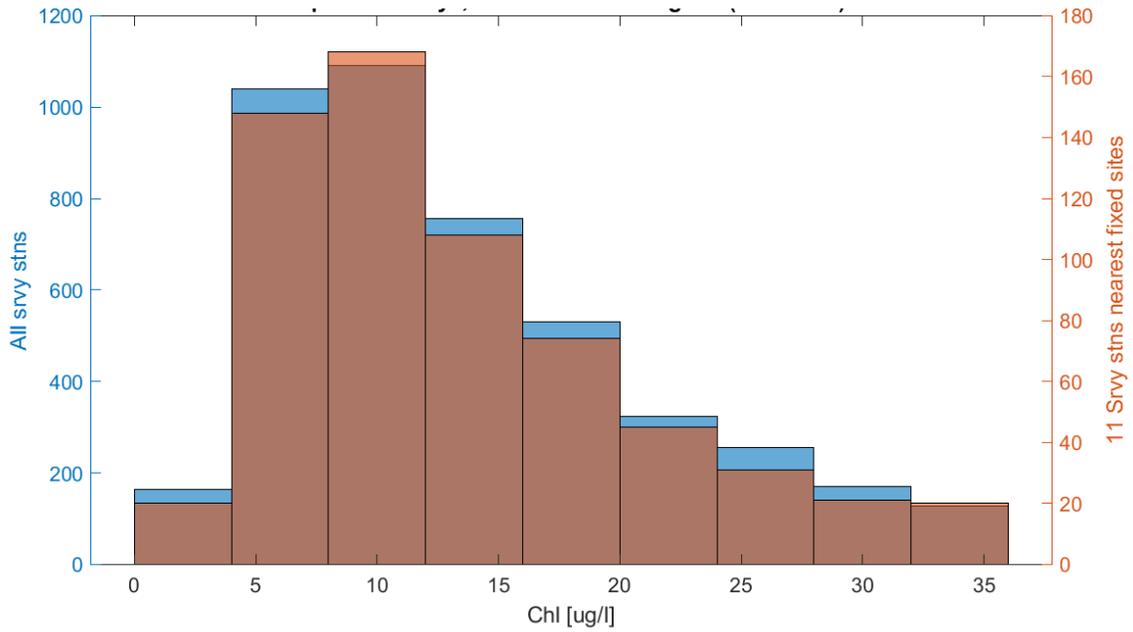


Figure 11-14. Histograms to gauge representativeness of time series.

Near-bottom chlorophyll concentration (horizontal axis) from spatial surveys using all years: all stations sampled (blue, left vertical axis) and only the 11 stations nearest to the fixed site station locations (orange, right vertical axis). Vertical axes are number of in each bin of chlorophyll ranges.



Task 12 Long-Term Trends in Temperature, Salinity, and Stratification

12.1 Scope

Objective. Assess long-term trends in NBFSMN temperature data, as called for in SNBIW temperature chapter. **Methodology.** Statistical trend identification (Weatherhead et al 1998, Leroy et al 2008) from near-surface and near-seafloor data. **Expected outputs and outcomes.** Number of years of data needed to establish trends will be determined. Warming rates and uncertainties will be quantified. The extent to which near-seafloor/near-surface trends differ indicates temperature influence on stratification trend.

12.2 Methods

The 15-minute resolution fixed site time series records for temperature and salinity were averaged at each station, using near-surface depth and near-bottom depth records individually, over the five monthly intervals available each season (May 15 to Jun 14, Jun 15 to Jul 14, Jul 15 to Aug 14, Aug 15 to Sep 14, and Sep 15 to Oct 14). The records for density stratification (difference between density at near-bottom depth and near-surface depth) was also treated. Caveats regarding weak confidence in density stratification, as noted above for the Phillipsdale station and in prior work for the Greenwich Bay station, apply similarly to this analysis. Months for which missing data periods occupied more than 40% of the time were omitted.

At each station and depth, the raw monthly-mean values from all years, for each of the five monthly intervals, were then averaged to give a mean seasonal cycle of five values. A de-seasoned record was computed by subtracting the mean seasonal cycle from the raw values of each year. Trend calculations were carried out on the de-seasoned values.

Two trend analysis methods were applied. The first method is the standard linear regression model. In addition to the rate or slope of the trend, and its 95% confidence interval based on the regression statistics, this method yields a p value and R^2 as metrics for the fit of the linear model.

The second method uses the autocorrelation of the unexplained variance or noise in the record, defined as the de-seasoned values minus the trend from the linear regression model. As explained by Phojanamongkolkij et al. (2014) the method was developed by Tian (1990), and Weatherhead et al. (1998) and Leroy et al. (2008) each improved on it with their own variants. This method yields confidence intervals for the trend slope, based on characteristics of the noise including its autocorrelation and magnitude relative to the estimated trend. It also gives an estimate for the minimum number of years sampling duration required for the calculation to have testing power sufficient to detect correctly, with probability 0.9, a trend that is at least twice as large as the noise. The estimate of years required to detect a trend can be compared to the actual sampling duration, as a useful gauge of confidence in the trend result. The fact that the NBFSMN time series data are only sampled during 5 months of the year is a constraint on the autocorrelation calculation. As a result the Leroy et al (1998) approach did not give meaningful results for this dataset. This also affected the Weatherhead et al (1998) approach, which as Phojanamongkolkij et al. (2014) noted is less general than Leroy et al (1998); the Weatherhead et al (1998) results are presented here, but should be viewed with caution.

See page 157 for information about supporting code and data files, and page 150 for how to use them.

12.3 Results

It is important to note at the outset that this trend analysis is based only on

- the fixed site time series measurements from 11 sites,
- the portion of the year they sample, from mid-May to mid-Oct, and
- the years 2001-2017 (for two of the 11 sites) or 2004/2005 to 2017 (all other sites).

The sites are representative mainly of deeper channelized areas of the bay, with some shallow embayment sites. The mid-May to mid-Oct period, during the 2001-2017 period, may well have trends that are not the same as those in other trend analyses that treat different time periods (for example wintertime conditions, or longer multi-decade records). Results from different datasets, collected at different locations and/or different time periods, that differ from those presented here would not imply any inconsistency or misinterpretation.

Temperature. The temperature records at Bullocks Reach are representative to demonstrate how the mean seasonal cycle is computed from the raw monthly values, then removed, and the difference used in the linear model trend calculation (Figure 12-1).

Trend results for temperature, at all stations and both near-surface and near-bottom depths, are summarized in Figure 12-2, from both computation methods. At the near-surface depth, the trend is for warming temperatures at a rate of about 0.3-0.4 °C/decade at most stations; Mount Hope Bay has a cooling trend of about 0.5 °C/decade, a magnitude as large as or larger than at most other sites but in the opposite sense, which will be discussed separately below. The warming rate is highest (about 0.6-0.7 °C/decade) at Sally Rock, but it has the shortest record duration and therefore the largest confidence interval. The rate is lowest (about 0.1-0.2 °C/decade) at Quonset Point, the only station for which the trend is not statistically significant, in the sense that its magnitude does not exceed the 95% confidence interval.

The p values of the linear model fit are greater than 0.05 at all stations except T Wharf, and the R² values are in the range of 0.01 to 0.08, typically about 0.04-0.05. The number of years of sampling required to detect the trend is higher than the number of sampled years at all sites, by at least a few years. In light of this, although the trend results are significant (at all sites except one) in the sense that they exceed the confidence interval, to place a high degree of confidence in their meaningfulness would not be strongly justified. Nonetheless they are consistent with prior work on near-surface temperature trends using independent datasets (e.g. SNBIW 2017, Chapter 1 on temperature, and references it cites), which lends them credence, suggesting they have merit and should not be dismissed as spurious.

At near-bottom depths all stations (including Mount Hope Bay) show warming and the rate of warming is faster, between about 0.6 and 1.0 °C/decade at most stations, than at near-surface depths. The minimum rate (about 0.25 °C/decade) is at Sally Rock, which is the only station where the trend is not larger than the confidence interval, and the maximum rate (about 1.1-1.2 °C/decade) is at the Phillipsdale and Bullocks Reach stations. More than half of the p values are less than 0.05, the R² values are in the range of 0.01-0.22, and the years required to detect a trend is comparable to or less than the sampled years at many stations and larger than it by several years only at two stations. This suggests the trend results for near-bottom depths are more meaningful than for near-surface depths, which is

consistent with the expectation that variability is higher near the surface due to more vigorous atmospheric influences there.

Salinity. Trend results for salinity, at all stations and both near-surface and near-bottom depths, are summarized in Figure 12-3. At the near-surface depth the trend is for increasing salinity at all stations, at a rate of typically about 1-2 PSU/decade (maximum about 4.2 PSU/decade at Phillipsdale and minimum about 0.7 PSU/decade at North Prudence), which is larger than the confidence interval at all stations. The p values are less than 0.05 at all stations except for Bullocks Reach, the R^2 values are in the range of about 0.04-0.23, and the years required to detect a trend are slightly less than the years sampled at many stations while higher than them by up to several years at other stations. These metrics suggest more confidence can be placed in the trend results for near-surface salinity than for near-surface temperature.

At the near-bottom depth, trends in salinity are also increasing, at rates larger than the confidence interval for all stations except Bullocks Reach where there is effectively no trend. The rates at most stations are about 0.5-1 PSU/decade, with a maximum of about 1.8 PSU/decade at Phillipsdale, thus notably lower overall when compared to the rates of salinity increase at near-surface depths. At Bullocks Reach the confidence in the trend is very weak, with p value greater than 0.05, R^2 near zero, and the number of years to detect a trend more than an order of magnitude longer than the sampled duration. For the other stations, the confidence is comparable to the shallow results, with the p value less than 0.05 at all but Phillipsdale, R^2 values ranging from about 0.03 to 0.29, and the years required to detect a trend ranging from slightly less than the sampled years to between a few and several years higher.

Stratification. Trend results for stratification at all stations are summarized in Figure 12-4. The trend is negative (decreasing stratification) and larger than the confidence interval at all stations. The rates range between decreases of about 0.2 and 2 kg m^{-3} per decade, averaging about 1 kg m^{-3} . The highest rates (all greater than 1 kg m^{-3} per decade) are at the Phillipsdale, Mount Hope, Poppasquash Point, and Conimicut Point stations and the lowest rates (less than 0.25 kg m^{-3} per decade) are at the North Prudence and Greenwich Bay stations. About half of the p values are less than 0.05, the R^2 values range from about 0.02 to 0.23, and compared to the number of sampled years the number of years required to detect a trend is about the same or slightly less for a few stations and up to a few years longer for all other stations except North Prudence and Greenwich Bay where it is many years longer.

The trends in temperature and salinity each contribute, independently and in the same sense, to the trend of decreasing stratification. The near-surface temperature trend consists of warming at a rate that is about 0.5 °C/decade slower than the near-bottom temperature warming trend, on average across all stations (Fig 2). This corresponds to a decrease in stratification of about 0.1 kg m^{-3} per decade, because a property of the seawater equation of state is that the density changes by roughly 1 kg m^{-3} for each 5 degrees of temperature change. The near-surface salinity trend is an increase that is about 0.9 PSU faster than the near-bottom salinity trend increase, on average across all stations (Figure 12-3). By the equation of state, this corresponds to a decrease in stratification of about 0.9 kg m^{-3} per decade. The two contributions to decreasing stratification, 0.1 kg m^{-3} due to temperature and 0.9 kg m^{-3} due to salinity, together account for the observed 1 kg m^{-3} per decade rate of decrease in stratification. Salinity is the dominant contributor (Figure 12-5). The larger influence of salinity, relative to temperature, on long-term trends in density stratification is consistent with the conclusions of Codiga (2012).

Summary and discussion. Trends in near-surface temperatures, with the exception of the Mount Hope Bay station which is discussed separately below, show warming at a rate of about 0.5 °C/decade on average bay-wide. This result is in a similar range as trends found by previous investigators using other records that are longer, from different locations in the bay, and/or from different time periods (e.g., Oviatt, 2004, Pilson 2008, NBEP 2009, Nixon et al 2009, Smith et al 2010, Fulweiler et al 2014, SNBIW 2017). This supports interpreting the trends described here as meaningful, despite that for some parameters at some stations the statistical measures provide only modest or weak support. The results for required sampling duration to detect trends should be viewed with caution, due to limitations of the method for partial-year sampling as noted above, but they suggest that after another 5-10 years have been added to the fixed site records (by the mid-2020s) they will be sufficiently long to detect trends confidently in all cases, as long as the general nature of the variability and trends do not change markedly.

At all stations across the bay the near-bottom waters are warming, at a rate of about 1 °C/decade on average bay-wide, which is roughly 0.5 °C/decade faster than the near-surface waters are warming. The long-term trend results identified here appear to be the first that have been reported for near-bottom bay waters. The finding that deep waters are warming faster seems sufficiently unexpected to generate questions about whether it may be an artifact of the field measurement and/or analysis methods. However, it is difficult to imagine a means by which all 11 fixed site stations would give spurious or strongly biased results. The time series records are 15-minute resolution and not subject to the fair-weather sampling bias Fulweiler et al (2014) noted could be an issue for deep measurements made from research vessels. While the uncertainties in the shallow and deep trends are large enough that the deep rates are distinguishable from the shallow rates only at some of the stations, the deep rate is higher than the shallow rate at all stations except one (Sally Rock, for which the record is shortest so trend results are weakest).

Trends for salinity include long-term increases at both the near-surface depths and near-bottom depths at all stations (about 1.6 PSU/decade, and 0.7 PSU/decade, respectively, on average bay-wide). The statistical measures are generally stronger for these trends than for the temperature trends.

The trend for stratification is a decrease of about 1 kg m⁻³ per decade, on average bay-wide. Both temperature and salinity contribute to the decrease; near-surface waters are warming slower and become saltier faster, compared to the respective trends in near-bottom waters, so for both parameters the magnitude of the difference between deep and shallow depths decreasing (Figure 12-5). Due to the seawater equation of state, the trend in salinity is of dominant importance to the stratification trend.

To discern the processes responsible for the observed trends in temperature, salinity, and stratification would require an in-depth examination of the heat and salt budgets for various subdomains of the bay. This would involve assessing how they are influenced by a number of complex interacting factors including air-sea interaction, river flow influences, trends in offshore coastal waters entering the bay, advection by estuarine exchange flow, and vertical mixing in the water column. Such an investigation is not attempted here, so discussing the causative factors for the trends identified is somewhat speculative. However, certain observations can nonetheless be made, and some likely conclusions drawn, as follows.

The main cause for increasing near-surface temperatures is understood (e.g. SNBIW Chapter 1) to be air-sea interactions and the influence of warming air temperatures. It is more difficult to ascertain what is

responsible for the observed near-bottom waters warming faster than near-surface waters. In addition to air-sea interaction, influences on bay temperatures include temperatures of river water and offshore waters that enter the bay. River water more strongly influences near-surface conditions due to its buoyancy, while offshore water is more saline and more strongly influences near-bottom conditions as it enters mainly in the deep inward limb of the estuarine exchange circulation (e.g. Codiga 2012). Air-sea interaction cannot be the main factor responsible for the near-bottom waters rate of warming, because its influence is mediated by near-surface waters that are warming more slowly. The most plausible explanation for the deep warming appears to be warming of subsurface waters offshore, in Rhode Island Sound and on the adjacent shelf. While recent work by numerous investigators (e.g., Kavanaugh et al 2017, Wallace et al 2018, others) has shown that shelf waters are in fact warming, they have examined different time periods and not necessarily the Rhode Island Sound waters entering the bay, so it is unclear whether the rates observed are sufficiently high to explain the warming rate of the deep bay identified here. Influences of the deep warming are influencing fish populations and fisheries (Langan et al., 2020).

The cooling trend for near-surface water at the Mount Hope Bay station is sharply anomalous compared to the other stations, but carries comparable statistical strength. It suggests that the heat budget for that waterbody is influenced by a different combination of processes than the rest of the bay. Fan and Brown (2003) and Swanson et al (2006) demonstrated that the Brayton Point Power Station has played a role. A feasible hypothesis is that Mount Hope Bay temperatures have been impacted by changes to the operations at the Brayton Point Power Station in recent years. Prior to 2011-2012, the power plant used bay water for cooling and thus was a source of heat for the bay. At that time a newly built set of cooling towers became operational, which reduced this heat source. In June 2017 plant operations were discontinued, so the heat source is no longer affecting the bay. Therefore, a plausible and self-consistent interpretation of the observed cooling trend is that it is due to this history of changes at the power station, which reduced a source of heat to the bay at about 5-6 years in to the 13-year record analyzed and eliminated it during the final year treated.

Bay salinities are controlled to a large extent by river runoff, in inverse relationship (e.g. Codiga 2012), and a consistent interpretation of the observed trends is that they are due to the reduced runoff that has occurred during the many of the last several years of the record treated. Note that the observed increasing salinity trend is opposite to the sense expected from recognized long-term region-wide trends occurring in precipitation, and therefore river runoff, which are increasing (e.g., SNBIW Chapter 2). The two results are not inconsistent; the long-term regional precipitation and river flow increases span multiple decades, and relatively short periods within them include decreases. The range of years treated here includes, during its latter portion, such a period of decline, and it has influenced the overall trend determined in salinity. With respect to the trend of more weakly increasing salinity in near-bottom waters, it is reasonable to assign most probable responsibility also to river runoff, because due to its buoyancy its effect would be expected to be stronger on near-surface waters than near-bottom waters. However, trends in salinity of offshore waters can be expected to influence deep bay conditions to a similar extent as trends in temperature of offshore waters, and as deduced above this could be an important factor.

The trend in stratification, while related directly the differences in trends of shallow and deep temperature and salinity, is influenced by vertical mixing as primarily controlled by wind conditions and internal dynamics of the estuarine exchange circulation. All other processes being the same, if there was

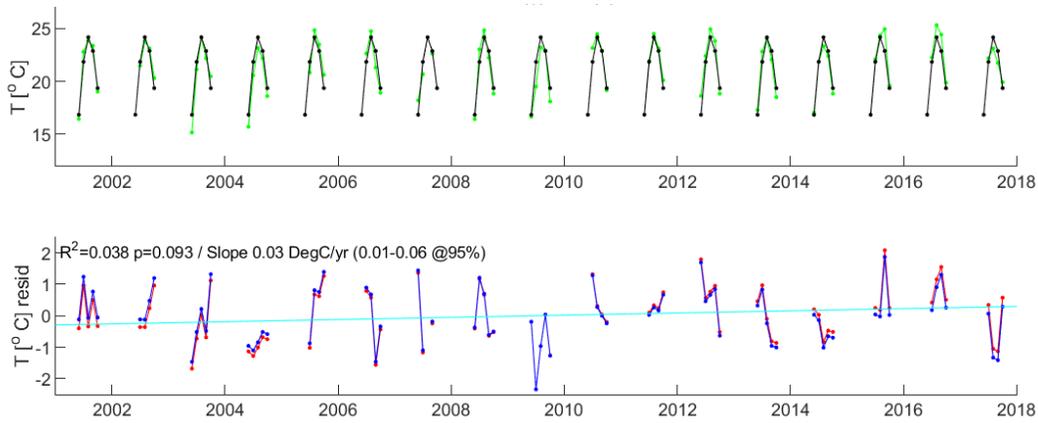
a long-term increase in vertical mixing it would drive decreased stratification, and associated decreases in the magnitude of the deep vs shallow temperatures and salinities. The trends observed can be explained as this very pattern, but only if superposed on trends of warming and increasing salinity throughout the water column. That the rates of decrease in magnitude of vertical difference in salinity and in temperature are roughly in proportion to their contributions to the stratification is somewhat suggestive that a long-term trend of increased vertical mixing may be playing a role in this way. But it will be a major challenge to discern whether this is the case, and if it is, identify the roles of wind and estuarine exchange circulation dynamics. It seems more likely that increased salinities due to reduced river flows in recent years is the dominant process, with trends in vertical mixing not being a major factor, but further work is required.

Relevance to hypoxia. The long-term trends for increasing temperatures and decreasing stratification have implications for hypoxia. Warming temperatures will increase metabolic rates and can be expected to increase chlorophyll growth, drive faster bacterial consumption of oxygen, and therefore tend to exacerbate hypoxia. In contrast, decreasing stratification will reduce the tendency for bottom waters to be isolated from oxygenated shallower waters, weakening one factor known to intensify hypoxia. The stratification trend is a decrease of 1 kg m^{-3} per decade. As noted in Codiga (2012) and Tasks 5 and 6 above, the difference in stratification between strongly stratified and weakly stratified years is at least a few kg m^{-3} . This suggests that substantial effects of the decreasing stratification trend on hypoxia would require up to a decade to start becoming important. The stratification trend is mostly driven by the trend in salinity, which as noted is in the sense opposite to that expected based on longer-term regional trends for precipitation and river runoff, due to dry conditions in the last several years of the records analyzed. Given this situation there is no strong expectation that the decreasing stratification trend will continue for another decade or longer. In this context it is unlikely that the trend in stratification will be of major importance to hypoxia, for example as compared to the effect that decreased nutrient loading appears to be having as described in earlier report sections. For these reasons, the long-term warming trend is considered more important to hypoxia than the trends in salinity or stratification.

Figure 12-1. Raw time series, removal of mean seasonal cycle, linear regression.

Representative example (Bullocks Reach). Near-surface depth and near-bottom depth records are shown in the upper and lower sections, respectively. The top frame shows the raw monthly-mean values (green) and the mean seasonal cycle repeated each year (black). The bottom frame shows the deseasoned record (red), the trend result from the linear model (cyan), and the deseasoned and detrended unexplained variance or noise (blue) used by the Weatherhead et al (1998) method.

Near-surface



Near-bottom

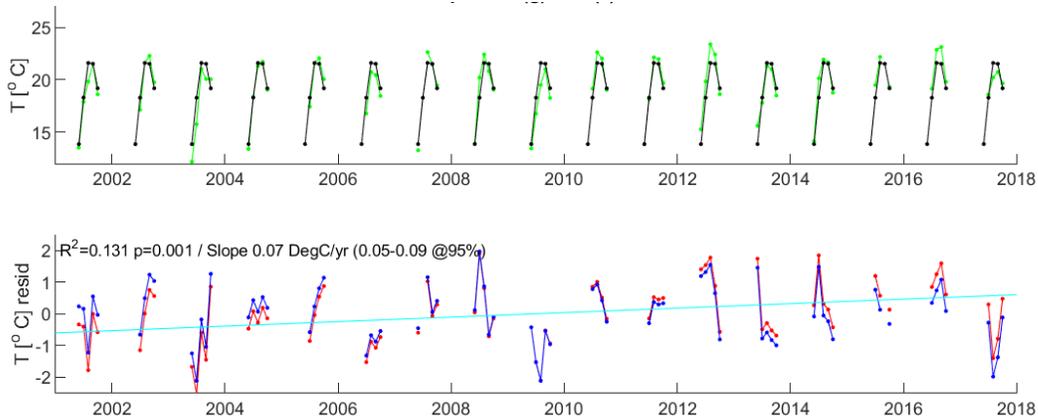


Figure 12-2. Trend results for temperature.

Upper two frames are for near-surface depth; lower two frames are for near-bottom depth. Results from the standard linear regression model are shown in the upper frame (trend magnitude with 95% confidence intervals) and lower frame (R^2 ; symbol is solid if p value < 0.05). In the lower frame is the duration of the record (solid bar) adjacent to the duration required to detect the trend (open bar).

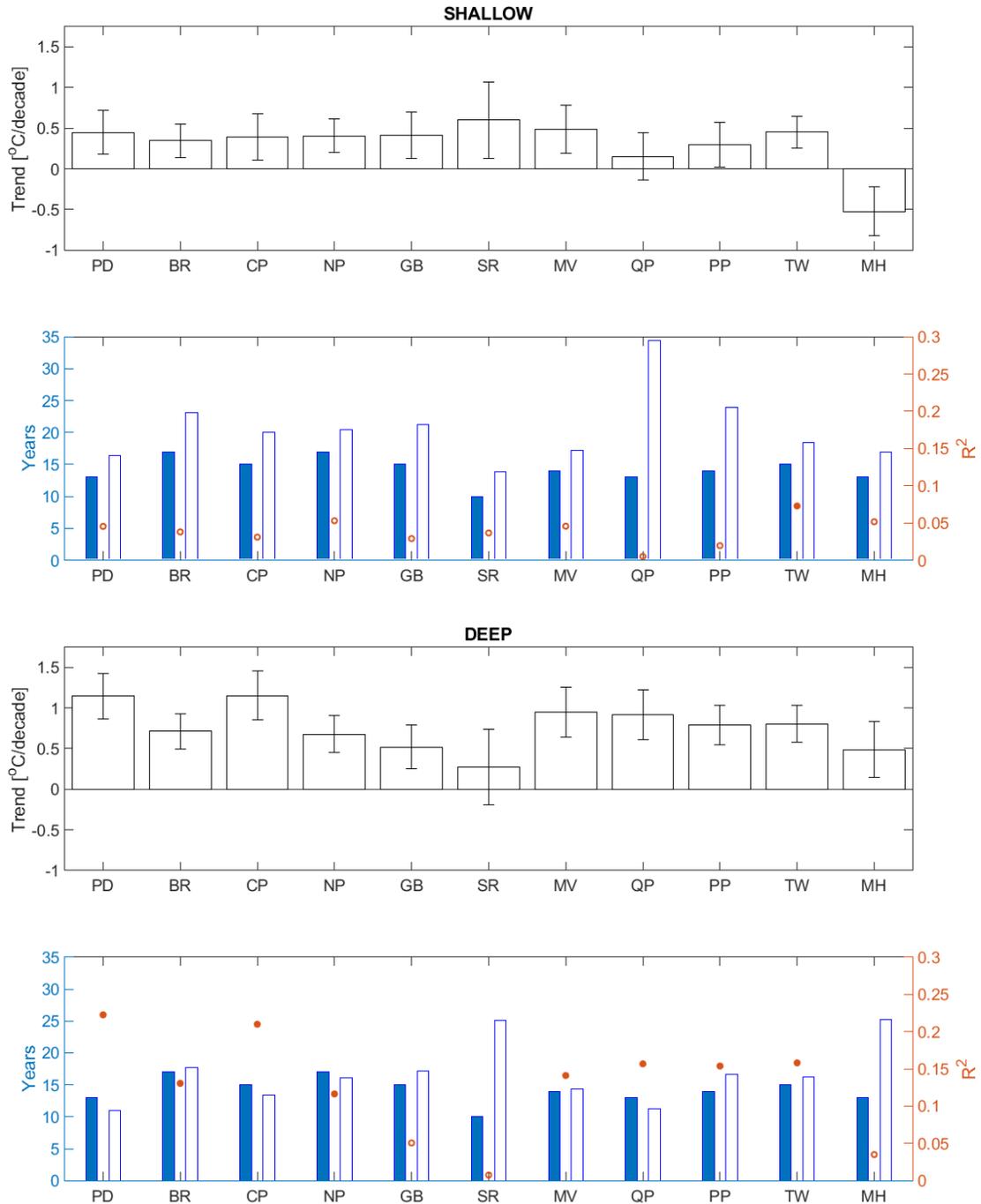


Figure 12-3. Trend results for salinity. Presented as in Figure 12-2.

For the deep Bullocks Reach record, the duration required to detect a trend (open bar, bottom frame) is 569 years.

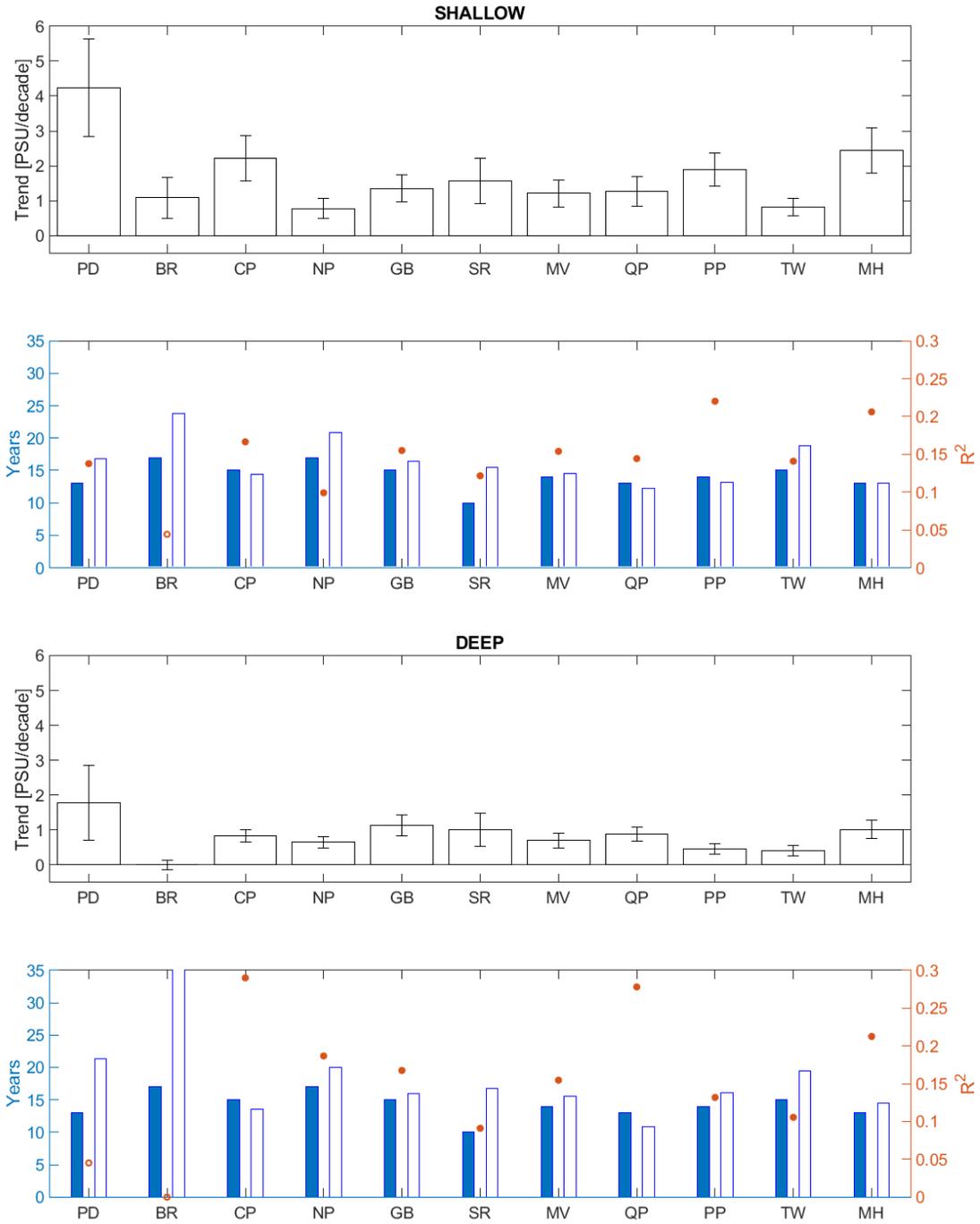


Figure 12-4. Trend results for stratification. Presented as in Figure 12-2.

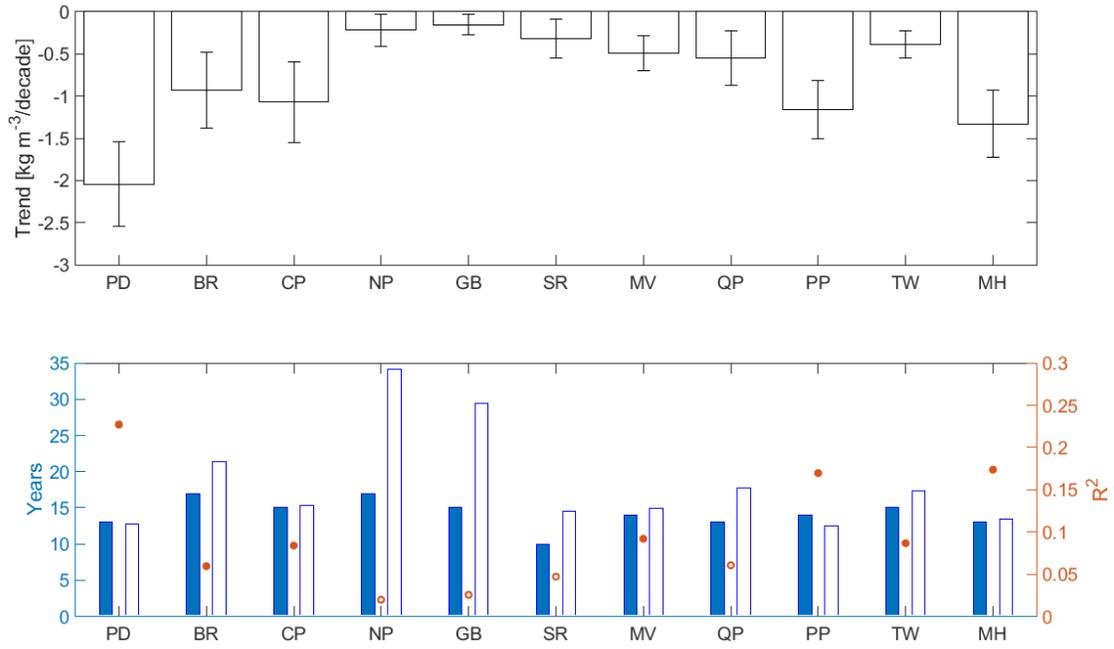
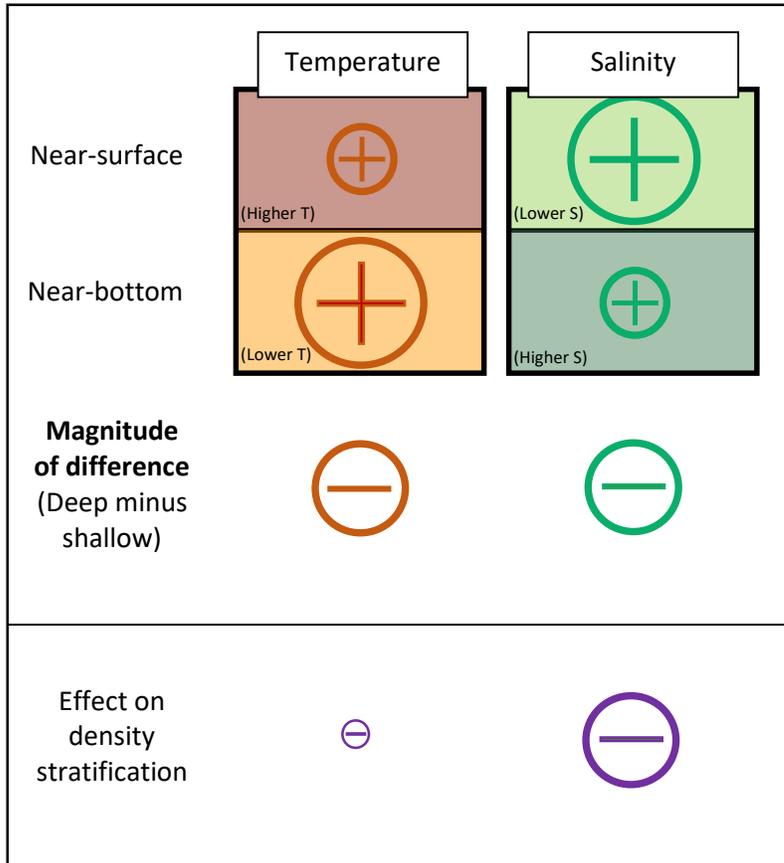


Figure 12-5. Schematic summary of trends.

Trends (plus and minus symbols in circles) observed in temperature, salinity, their vertical differences, and their effects on the trend in stratification. In the lower left corner of the four main colored boxes of the diagram, the long-term mean conditions are noted in parentheses. For both temperature and salinity, the difference between the shallow and deep trends contribute to a long-term trend for declining stratification; the magnitude of the salinity effect on density dominates that of temperature.



All Tasks: Data and Code Files Documentation

What follows here is a task by task explanation of the data and code files used. All data and code files are in a single project repository accessible online at <https://figshare.com/s/7d51f2540df6638a4552>. There are subfolders **Task1**, **Task2**, etc under the main folder of the repository, for tasks with associated data and code files.

The code is in Matlab and generally structured in a series of “modules”. Each module has its own folder with identifying name indicating the task the module accomplishes (for example to manipulate data and save to disk newly calculated quantities; or to read in data and create a series of plots from it). In this folder is the main script, usually named the same as the main folder or similarly, and subfolders with some or all of the following names: **in**, **sub**, **out**, **plt**, and **diary**. The **in** subfolder holds configuration files containing information that is read in and used by the main script, for example input file locations. The **sub** subfolder holds functions or subscrips that the main script calls, if any. The **out** subfolder holds data files the module creates and saves to disk, if any. The **plt** subfolder holds plot files the module creates and saves to disk, if any. The **diary** folder holds logs of commands executed during each run of the main script.

There is a set of utility functions, many of which facilitate, and depend on, the above subfolder and configuration input file structure for each module:

- **setpath.m**: This is the only utility that must be run manually. Run it prior to running the main script of a module, after changing directory in to the main folder for the module. It modifies the Matlab path to add the paths of the main module folder and all its subfolders (and remove from it the paths of the folders in the prior module that was run, if any). Subsequent runs of the main script do not require running setpath again, if the working directory has not been changed.
- **wipe.m**: This clears the workspace and deletes all figures. Every main script starts by calling it.
- **modprelims**: This is used in the main script to keep track of the run name and other diagnostics. It also provides convenient access to a few commonly used constants.
- **inputsdiary.m**: This is used to read information from configuration files stored in the **in** subfolder, and save a log of commands made to a file in the **diary** subfolder.
- **getparams.m**: This is called by inputsdiary.m to read each individual configuration file.
- **figport.m**: Function to create 8.5”x11” size figure window, portrait orientation.
- **figland.m**: Function to create 8.5”x11” size figure window, landscape orientation.
- **tri_var_ss.m**: Function that applies triangle weight running mean as low-pass filter to time series data, and subsamples the results. Used during some data processing stages.

All of these utility functions are used by every module and therefore must be located in a folder that is on the Matlab path, in order for a module to successfully execute; the only exceptions are the last three (figport.m, figland.m, tri_var_ss.m), which are required only if the module uses them. In the repository, these utility files are in the folder **Utilities**.

1. Phillipsdale Landing Time Series Data Reduction

The data files and supporting code for this task are grouped with, and a subset of, the data files and supporting code for Task 4 on oxygen, and Task 9 on chlorophyll.

2. Development of Nitrogen Load Time Series

Results files

- **Daily2001to2017TNloadTNconcQflow18NBsources.xlsx**
 - The first sheet in this spreadsheet provides the metadata.
- **Daily2001to2017TNloadTNconcQflow18NBsources.mat**
 - This is a matlab binary storage file with the same contents as in the spreadsheet.

Raw input files (in spreadsheet format with flow and concentration observations)

- Treatment plants:
 - **Bristol, East Greenwich, East Providence, Jamestown, Newport, Quonset, and Warren:** File **FromDEM20180619.xlsx**, provided by Heidi Travers at Rhode Island Department of Environmental Management (RI DEM). The file also contains Fields Point and Bucklin Point data; these were used but only for the period prior to early 2004 (see below).
 - **Somerset:** File **URI_MA_WWTFHistoricNutrient-cac.xlsx**, provided by Heidi Travers (RI DEM). File also contains Fall River data but they were not used (see Fall River below).
 - **Field's Point and Bucklin Point:** Files **Field's Point Plant Data_Mar 2004-2017.xlsx** and **Bucklin Point Plant Data_Mar 2004-2017.xlsx**, from Narragansett Bay Commission (NBC), provided by Heidi Travers (RI DEM). The files include left-censored values, denoted with prefix string "LT" (less than). The censored values are a small portion of the dataset so it was determined acceptable for the analysis to treat them by substitution of the censoring level. Therefore, to make it easier to read values in to Matlab, a search and replace operation was applied manually to the files to remove all "LT" strings and the results saved as **Field's Point Plant Data_Mar 2004-2017_removedLTs.xlsx** and **Bucklin Point Plant Data_Mar 2004-2017_removedLTs.xlsx**, which are the files actually used by the analysis.
 - **Fall River:** 204 files **YYYY-MM.XLS**, where YYYY is year 2001 through 2017 and MM is month 01 through 12, provided on request by Jeanne Wordell of the Fall River treatment plant.
- Rivers:
 - Files **YYYY NBC River and Nutrients Data.xlsx**, for years YYYY = 2005 through 2017, obtained from NBC via download from <http://snapshot.narrabay.com/WaterQualityInitiatives/NutrientMonitoring>.

Supporting code (a series of matlab modules)

- WWTFs (in **WWTFs** folder)
 - Seven WWTFs in Rhode Island (in **RI7** folder)
 - Modules **Inventory**, to make diagnostic plots showing the temporal coverage/resolution of flow and concentration measurements, and **RI7LoadCalc**, for needed data manipulations to create the load timeseries.
 - Fields and Bucklin (in **FieldsBucklin** folder)
 - Modules **FPBP1**, for needed data manipulations to create the load timeseries, and **FPBP2**, to save files and plots in a standard format used by all 18 sources.
 - Fall River (in **FallRiver** folder)
 - Modules **Inventory**, to make diagnostic plots showing the temporal coverage/resolution of flow and concentration measurements, and **FRLoadCalc**, for needed data manipulations to create the load timeseries.
 - Somerset (in **Somerset** folder)

- Modules **Inventory**, to make diagnostic plots showing the temporal coverage/resolution of flow and concentration measurements, and **SomersetLoadCalc**, for needed data manipulations to create load timeseries.
- Rivers (in **Rivers** folder)
 - Modules **Inventory**, to make diagnostic plots showing the temporal coverage/resolution of flow and concentration measurements, and **RivLoadCalc**, which does the needed data manipulations to create the load timeseries.
- Ungauged/riparian area (matlab module **UngaugedLoad**), which uses river flow and area ratios and computes the load.
- All sources (in folder **AllSources**)
 - Module **Summary1**, which makes various plots of results for all 18 sources and saves the data from all sources to a single mat file and a single xlsx file (“Results files” above).

3. Retrieval and Reduction of Wind, Tidal Range, and Sea Level Data

Winds. The Matlab module is called `nc2wndprms`. It retrieves the winds, computes a range of associated parameters from them-- eastward wind component, northward wind component, wind speed, wind direction, wind speed cubed (relevant to wind mixing), wind components directed towards the 16 cardinal directions i.e. toN, toNNE, toNE, ... For each such parameter it computes the daily mean and a low-pass filtered version (12-hour half-width triangle weight), in addition to the raw instantaneous NARR wind. All results are saved to the binary Matlab file `WindsNARR.mat`. A spreadsheet named `WindsNARR.xlsx` was generated, by a script `mat2xlsx`, to include the same contents as `WindsNARR.mat`. The first sheet of the spreadsheet provides all metadata for the other sheets.

Data files: `WindsNARR.mat`, `WindsNARR.xlsx`.

Supporting code: `nc2wndprms`, `mat2xlsx`.

Tidal Range. Module `tidprms` saves all fields the binary Matlab file **TidalRange.mat**. A spreadsheet named **TidalRange.xlsx** was generated, by a script `mat2xlsx`, to include the same contents as **TidalRange.mat**. The first sheet of the spreadsheet provides all metadata for the other sheets.

Data files: `TidalRange.mat`, `TidalRange.xlsx`.

Supporting code: `tidprms`, `mat2xlsx`.

Sea Level. Module `slvlprms` saves all fields to the binary Matlab file `slvlprms.mat`. A spreadsheet named `slvlprms.xlsx` was generated, by a script `mat2xlsx`, to include the same contents as `slvlprms.mat`. The first sheet of the spreadsheet provides all metadata for the other sheets.

Data files: `SeaLevel.mat`, `SeaLevel.xlsx`.

Supporting code: `slvlprms`, `mat2xlsx`.

4. Update SNBIW Oxygen Results with Phillipsdale

Data files: The modified corrected files are in a folder **ModifiedCorrected**, with a subfolder for each of the 11 stations using 4-letter codes (phdl, bulr, cnpt, npru, mtvw, qnpt, popt, twhf, mthp, grby, and srck). In each such subfolder there is an xlsx file with filename `XXYYYY.xlsx` where `XX` is the two-letter code for the station (PD, BR, CP, NP, MV, QP, PP, TW, MH, GB, and SR) and `YYYY` is the four-digit year of the data it contains. The metadata (parameters, units, source contact person, etc) is included in each spreadsheet; the files are all “corrected” versions of the data from Heather Stoffel (URI GSO), with minor modifications to the formatting in order to make them compatible with the Matlab-based processing

scripts applied to them. In the **Plots** folder is HypoxiaIndex.xls, containing data from Figure 4-1, as generated by the script mat2xlsx.m included under supporting code.

Supporting Code:

Script “Location.m”. A stand-alone script in which the station location information is kept, and a simple map plot is made.

Module “xls2mat”. For each of the XYYYYY.xlsx modified corrected files, it creates a matlab data file XYYYYYqh.mat, containing the raw data in matlab structures (“qh” refers to the quarter-hourly temporal resolution), and some metadata including the latitude-longitude coordinates of the station, the bathymetric depth at the station, and the depths of the shallow and deep sensors.

Module “rawplot”. For each station it uses the XYYYYYqh.mat file from the xls2mat module and makes a one-page plot, with one frame for each year of data, for oxygen and for chlorophyll. These are for visual inspection. They annotate the percent of time between mid-May and mid-Oct that has good quality data for each year.

Module “MWT”. Uses the XYYYYYqh.mat files from the xls2mat module. For each, if percent good data is 55% or higher, it applies the moving window trigger (MWT) algorithm to determine events relative to the three thresholds (1.4 mg L⁻¹, 2.9 mg L⁻¹, 4.8 mg L⁻¹) and save the statistics of events in output files named with format XYYYYY-Txpy-May15Oct15.mat where xpy is 1p4, 2p9, or 4p8 respectively. It also generates plots showing sections of timeseries, from each year of data at each station, during which there were events, with the identified events marked in color for easy visual inspection.

Module “merge”. Uses output files from the MWT module and merges all stations and all years. It creates one file for results relative to each of the three thresholds: T1p4-May15Oct15-min55.mat, T2p9-May15Oct15-min55.mat, and T4p8-May15Oct15-min55.mat.

Module “stats”. Uses the three output files from the Merge module and computes the statistics necessary to estimate uncertainties based on data gaps, saving them in three output files: T1p4-May15Oct15-min55-2wk1dy-90th10th.mat, T2p9-May15Oct15-min55-2wk1dy-90th10th.mat, and T4p8-May15Oct15-min55-2wk1dy-90th10th.mat.

Module “metrics”. Uses the output from the Merge and Stats modules and computes the seasonal Hypoxia Index values, with their uncertainties, saving them in output files T1p4-May15Oct15-min55.mat, T2p9-May15Oct15-min55.mat, and T4p8-May15Oct15-min55.mat.

Module “IndStns”. Uses the output from the Metrics module, together with the ‘wet’/’normal’/’dry’ designations of years based on analysis of basin-wide river flows in Task 6 below, and creates Figures 4-1, 4-3, and 4-5.

Module “StnGrps”. With the same inputs as IndStns, creates Figures 4-2, 4-4, and 4-6.

5. Inter-Annual Variability of Hypoxia and Stratification

Supporting Code:

Module “strt”. This module reads in the temperature and salinity data and computes the density and density difference (stratification), as well as other auxiliary parameters such as the temperature difference and salinity difference. saving the results as output files (one per station-year).

Module “rawplot”. This module uses the output from the strt module to generate a one-page plot for each station showing density and density difference for all years.

Module “MergeStrt”. This module pulls the density difference from the output files of the strt module, organizes them in to a format equivalent to the format of the hypoxia index results saved by IndStns, and saves results for all stations and all years in a single file.

Module “HypVsStrt1”. This module uses the output of IndStns (from Task 4) and MergeStrt to generate the plots.

6. Drivers of Inter-Annual Hypoxia Variability

Supporting code:

Rivers. Supporting code is in folder “RivNldStrtHyp”. A series of Matlab modules is used, named ‘1GetData’, ‘2and5DiagnosticPlots’, ‘3FillGaps’, ‘4AddMaskerchuggHardig’, ‘6Baywide’, ‘WetDryInterm’, ‘rivprms’, and ‘mat2xlsx’. The ‘1GetData’ module retrieves the raw data from the US Geological Survey website, and aligns all the records to a common time grid. The ‘2and5DiagnosticPlots’ module acts on the output of ‘1GetData’ or ‘4AddMaskerchuggHardig’ and creates a series of diagnostic plots. The ‘3FillGaps’ module fills gaps shorter than 3 days by linear interpolation and fills longer gaps using regression against a nearby river. The ‘4AddMaskerchuggHardig’ module computes the Maskerchugg and Hardig flows from the measured Hunt flow and wappends them. The ‘6Baywide’ module computes the river mouth flows, the bay-wide flow, and the ungauged flow. The ‘WetDryInterm’ module computes and applies the thresholds for designating individual years as wet, dry, or intermediate. The ‘rivprms’ module creates and appends 12-hr subsampled records. The ‘mat2xlsx’ module saves the data to an Excel spreadsheet file with the final results, Rivers.xlsx.

Output files

Data files: Rivers.mat, Rivers.xlsx.

In the Plots folder are spreadsheets with data from Figures 6-2 and 6-3.

Supporting code: 1GetData, 2and5DiagnosticPlots, 3FillGaps, 4AddMaskerchuggHardig, 6Baywide, WetDryInterm, rivprms, mat2xlsx.

Hypoxia. Supporting code is in folder “RivNldStrtHyp”. The module Hyp computes the bay-wide hypoxia index, for each of the three thresholds, averaged over stations BR-CP-NP-MV-QP-PP-TP-MH. A series of matlab modules computes the surplus-duration calculation of stratification: 1MWT (computes surplus-duration for each individual station-year), 2merge (groups all station-years together), 3stats (computes statistics used to correct for missing data gaps), 4metrics (computes gap-corrected season-cumulative surplus-durations), and Strat (averages across stations BR-CP-NP-MV-QP-PP-TP-MH). A series of modules NldStrtQvHyp1, NldStrtQvHyp2, and NldStrtQvHyp3 does final calculations and plotting. NldStrtQvHyp1 calculates the river flow, mean stratification, and TN load over the four month ranges and makes a series of plots, including Figures 6-3 and 6-4. NldStrtQvHyp2 does the same but uses the median instead of the mean, for stratification. NldStrtQvHyp3 does the same but uses the season-cumulative surplus-duration (output of the Strat module), instead of the mean or median, for stratification. The results in Figure 6-5 are plotted by module AuxPlt.

The data underlying Figure 6-2 are in spreadsheet file RiversWetIntDry.xls.

7. Time Series and Spatial Survey Datasets Together: Oxygen

Starting with the spreadsheet file from Prell, In Excel, “export to text” was used, so the data could be read from the resulting ascii file in to Matlab while self-consistently handling the date formatting. Corrections were made as follows:

- Some times on 6/8/2010 at station WPS14 were corrected to be “13:00” as they were “13”.

- The years in some date strings were corrected to be “2014” or “2015” instead of “14” or “15”:
 - Some of 7/8, 7/23, 8/5, 8/21 and 9/3 of 2014, and 9/8/2015.
- Some “BBT3” strings were corrected (from “BBT3 ”, extra space at end).
- Some “UPB10” strings were corrected (from “UPB 10”, extra space in middle).
- Some “EPS01” strings were corrected (from “ESP01”).
- Station “UPB-NP” was dropped as it was sampled only one day on one year.

In the table constructed from the spreadsheet the only values missing were on 7/23/2014 at EPS01, WPS01,02,04,05,07,08,09,13,14, and UPB01.

A series of Matlab modules was used. ‘ReduceBoth’ performs data reduction on the spatial surveys (including the rank ordering) and time series datasets and stores the results in a matlab file ‘SrvysAndTimSer.mat’, which is used as the input to most additional calculations. The map of ranks is created by module ‘basicmaps’. Calculations of spatial correlations (spatial surveys only) are done by module ‘SpatCorrCalc’, results of which are plotted by the ‘MapPlts’, ‘DistPlts’, and ‘RnkPlts’ modules. The lagged autocorrelations were calculated and plotted by module ‘TemporalCorr1’. Lagged correlations between spatial survey and time series datasets were computed by module ‘LaggedCorrCalc’ and plotted by modules ‘CorrMaps1’, ‘RhoMatrix’, and ‘RvsLag’. Percent hypoxic area calculations and plotting are done by modules ‘Scat’ and ‘RgrsVsIndx’. Distributions used to assess representativeness were computed and plotted by modules ‘SrvTmpl’ (for spatial surveys) and ‘TimserSptl’ (for time series).

Input files: ‘DO MASTER 2005-2017.xlsx’ (spatial surveys profiles), ‘MASTER HYPOX 05-17 WLP.xls’ (percent hypoxic areas), ‘narrbay_bathy_ngdc.mat’ (bathymetry used in plots), ‘narrbaycoast.mat’ (coastline used in plots), ‘80th-May15Oct15-min55.mat’ (chl thresholds, see Task 9).

A spreadsheet file for data in Figure 7-2, generated by the mat2xlsx.m script, is in the Plots folder.

8. Toward Improved Chlorophyll Time Series Using Grab Samples

Time series of *in situ* chlorophyll fluorescence.

Data files: The modified corrected files are the folder **ModifiedCorrected** under the supporting code for Task 4, as described above for that task including the script **Location.m** and modules **xls2mat** and **rawplot**.

Grab samples.

Raw data for grabs was obtained in files “20200506_Chlorophyll Grabs for Dan Codiga.xlsx” from NBC, “CHL.grab samples.2006-2017.KH.xlsx” from URI/GSO, and “TW.surface2003-2017.csv” from URI/GSO containing NBNERR results. A module “Reduction” reads in the data from the three spreadsheets, organizes it to a single table of values for station, date, time, and chlorophyll values, computes initial statistics in Table 8-1, and saves results in file “grabs.mat”. A module “rawplotgrabs” creates plots of time series chlorophyll fluorescence (processed as described in Task 4), in the same way as the rawplot module of Task 4, with grab sample results (from grabs.mat) superposed, one page per station showing all years (Figures 8-1 to 8-11). A module “Scat” creates (a) an output file containing both the data in grabs.mat and time series data corresponding to each grab sample, and (b) scatterplots of the time

series and grab results (Figures 8-12 to 8-14). Finally `mat2xlsx.m` creates a spreadsheet `GrabsAndSondes.xlsx` containing the values of the file `GrbSnd.mat`.

9. Update SNBIW Chlorophyll Results with Phillipsdale

A series of Matlab code modules, building from those in earlier tasks and using as its input the in situ time series chlorophyll fluorescence data described in section 8 above, carried out the calculations and generated the figures as follows. The label “SS” in the module names refers to the fact that these calculations use Site Specific thresholds in the event-based analysis.

- 0ThreshCalcSS: Compute 80th percentile thresholds.
- 1MWT_SS: Compute events and event characteristics.
- 2mergeSS: Merge events results from multiple sites in to one file.
- 3statsSS: Compute statistics of events, needed for gap-based estimation of uncertainties in index.
- 4metricsSS: Compute index results including uncertainties.
- 5IndStnsSS: Create Chlorophyll Index figure showing individual station results.
- 6StnGrpSS: Create Chlorophyll Index figure showing means of station groups.

The data underlying Figure 9-1 are in spreadsheet file `ChlIndexSiteSpecific.xls`, generated by `mat2xlsx`.

10. Refine Chlorophyll Index

A series of Matlab code modules, building from those in earlier tasks and using as its input the in situ time series chlorophyll fluorescence data described in section 8 above, were written to carry out the calculations and generate the figures as follows. The label “BW” in the module names refers to the fact that these calculations use thresholds that are applied bay-wide (same threshold at all stations; in contrast to the site-specific thresholds used in Task 9) in the event-based analysis.

- 0ThreshCalcBW: Compute thresholds. The results for the 20th, 50th, and 80th percentile are used in the results section below. In addition the 33rd, 67th, 90th, and 95th percentiles were computed but they were not used as the 20th, 50th and 80th were considered sufficient.
- 1MWT_BW: Compute events and event characteristics.
- 2mergeBW: Merge events results from multiple sites in to one file.
- 3statsBW: Compute statistics of events, needed for gap-based estimation of uncertainties in index.
- 4metricsBW: Compute index results including uncertainties.
- 5IndStnsBW: Create Chlorophyll Index figures showing individual station results.
- 6StnGrpBW: Create Chlorophyll Index figure showing means of station groups.

The data underlying Figure 10-2 are in spreadsheet file `ChlIndexBayWide.xls`, generated by `mat2xlsx`.

11. Time Series and Spatial Survey Datasets Together: Chlorophyll

Data and code files in this task are mostly included in those for Task 7 described above. There is one additional input file, ‘CHL%AREA 05-17 WLP.xlsx’ (percent high-chlorophyll areas), and the chlorophyll-based thresholds file ‘80th-May15Oct15-min55.mat’ from Task 10 is used; there are chlorophyll-focused

versions of the modules `RegrVsIndx` and `Scat`. The `mat2xlsx` code saves spreadsheet `ChlorophyllSpatialCorrelations.xls`, which is the data in Figure 11-1, in the `Plots` folder.

12. Long-Term Trends in Temperature, Salinity, and Stratification

Module “Reduction” computes monthly means from the 15-minute resolution time series files. Module “Monthly” applies trend computations to monthly mean values and creates a series of plots (example, Figure 12-1). Module “SummaryPlots” creates summary plots (Figures 12-2 to 12-4). The data underlying Figures 12-2 to 12-4 are in the `DataFiles` folder, as generated by `mat2xlsx.m`.

References Cited

- Codiga, D.L., Stoffel, H., Deacutis, C., Kiernan, S., & Oviatt, C., 2009. Narragansett Bay Hypoxic Event Characteristics Based on Fixed-Site Monitoring Network Time Series: Intermittency, Geographic Distribution, Spatial Synchronicity, and Interannual Variability. *Estuaries and Coasts*, 32(4), 621-641.
- Codiga, D.L. 2012. Density stratification in an estuary with complex geometry: Driving processes and relationship to hypoxia on monthly to inter-annual timescales. *J. Geophys. Res.* 117. C12004. DOI 10.1029/2012JC008473.
- Codiga, D.L. 2020. Daily-Resolution 2001-2017 Time Series of Total Nitrogen Load to Narragansett Bay (RI/MA USA) from Bay-Wide Treatment Facility and Watershed Sources. NBEP Technical Report. NBEP 20-231B. URL: <https://figshare.com/s/95abe0296139515ffb88>. DOI: 10.6084/m9.figshare.12573851.
- Durant, D. and K.B. Raposa. 2010. Water quality, nutrients, and meteorology at the Narragansett Bay National Estuarine Research Reserve: 2008 annual report. Narragansett Bay Research Reserve Technical Reports Series 2010:2. <http://nbnerr.org/wp-content/uploads/2016/12/2010-DurantRaposa-NBNERR-Tech-Series-2010.2.pdf>
- Fan, Y and Brown, W, 2006. On the heat budget for Mt Hope Bay. *Northeastern Naturalist*. 13 (Special Issue 4): 47-70.
- Fulweiler, RW, A.J. Oczkowski, K.M. Miller, C.A. Oviatt, M.E.Q. Pilson, 2015. Whole truths vs. half truths – And a search for clarity in long-term water temperature records, *Estuarine, Coastal and Shelf Science*, Volume 157, Pages A1-A6, ISSN 0272-7714, <https://doi.org/10.1016/j.ecss.2015.01.021>.
- Kavanaugh, M. T., Rheuban, J. E., Luis, K. M. A., & Doney, S. C., 2017. Thirty-three years of ocean benthic warming along the U.S. Northeast Continental Shelf and Slope: Patterns, drivers, and ecological consequences. *Journal of Geophysical Research: Oceans*, 122, 9399– 9414. <https://doi.org/10.1002/2017JC012953>
- Kellogg, D.Q., 2017. Long-term Trends in River Flow to Narragansett Bay. Coastal and Estuarine Research Federation, Biennial Conference. Providence, RI.
- Kellogg, D.Q., 2018. Freshwater flow to Narragansett Bay: An analysis of long-term trends. EPA Southern New England Program, Healthy Communities Grant Program.
- Langan, JA, MC McManus, DR Zemeckis, and JS Collie, 2020. Abundance and distribution of Atlantic cod (*Gadus Morhua*) in a warming southern New England. *Fish. Bull.* 118:145-156.
- Leroy, S.S., J. G. Anderson, J. Dykema, and R. Goody, 2008. Testing climate models using thermal infrared spectra. *J. Climate*, 21, 1863–1875, doi:10.1175/2007JCLI2061.1.
- Mesinger, F., G. DiMego, E. Kalnay, K. Mitchell, P.C. Shafran, W. Ebisuzaki, D. Jovic, J. Woollen, E. Rogers, E.H. Berbery, M.B. Ek, Y. Fan, R. Grumbine, W. Higgins, H. Li, Y. Lin, G. Manikin, D. Parrish, and W. Shi. 2006. North American Regional Reanalysis. *Bulletin of the American Meteorological Society* 87 (3): 343–360. doi:10.1175/BAMS-87-3-343.

NBEP, 2009. Currents of Change: Environmental Status & Trends of the Narragansett Bay Region. Final Technical Draft. Narragansett Bay Estuary Program.

NBEP, 2017. State of Narragansett Bay and Its Watershed. Technical Report. Narragansett Bay Estuary Program. Providence, RI.

Nixon, S, Fulweiler, W, Buckley, B, Granger, S, Nowicki, B, Henry, K., 2009. The impact of changing climate on phenology, productivity, and benthic–pelagic coupling in Narragansett Bay. *Estuarine, Coastal and Shelf Science*. 82. 1-18. 10.1016/j.ecss.2008.12.016.

Oviatt, C.A. The changing ecology of temperate coastal waters during a warming trend. *Estuaries* 27, 895–904 (2004). <https://doi.org/10.1007/BF02803416>.

Oviatt, C., L. Smith, J. Krumholz, C. Coupland, H. Stoffel, A. Keller, M. C. McManus, and L. Reed, 2017. Managed nutrient reduction impacts on nutrient concentrations, water clarity, primary production, and hypoxia in a north temperate estuary. *Est. Coast. Shelf Sci.*, 199, 25-34. <https://doi.org/10.1016/j.ecss.2017.09.026>.

Phojanamongkolkij, N., Kato, S., Wielicki, BA, Taylor, PC, Mlynczak, MG, 2014. A comparison of Climate Signal Trend Detection Uncertainty Methods. *J. Climate*. 27, 3363-3376. DOI: 10.1175/JCLI-D-13-00400.1

Pilson M.E., 2008. Narragansett Bay Amidst a Globally Changing Climate. In: Desbonnet A., Costa-Pierce B.A. (eds) *Science for Ecosystem-based Management*. Springer Series on Environmental Management. Springer, New York, NY

Prell, W.L., Murray, D.W., Howell, P., 2015, Reanalysis of the Narragansett Bay Spatial Survey (Day Trippers) Water Column Data for Temperature, Salinity, Density, and Dissolved Oxygen (2005 to 2013), Narragansett Bay Estuary Program, NBEP-15-128. Providence, RI.

Prell, W.L, Murray, D.W., Carlson, L., and Thieleman, E., 2016, Update and integration of the 2014 and 2015 data into the Spatial Survey data set and generation of decade-long (2005-2015) trends of DO and Chlorophyll distribution in mid and upper Narragansett Bay for the Narragansett Bay Estuary Program’s The State of Our Watershed Report. Narragansett Bay Estuary Program, NBEP-16-132. Providence, RI.

RIDEM, 2014. Quality Assurance Project Plan, Narragansett Bay Fixed-Site Monitoring Network. Rhode Island Department of Environmental Management. Providence, RI. <http://www.dem.ri.gov/pubs/qapp/nbfsmn.pdf>

Ries, K. G. 1990. Estimating surface-water runoff to Narragansett Bay, Rhode Island and Massachusetts, U. S. Geol. Surv., Water-Resour. Invest. Rep., 89-4164, 1–49.

Smith, L. M., Whitehouse, S., & Oviatt, C. A. 2010. Impacts of Climate Change on Narragansett Bay. *Northeastern Naturalist*, 17(1), 77-90. doi: 10.1656/045.017.0106SNBIW

Swanson, C., H. S. Kim, and S. Sankaranarayanan, 2006. Modeling of temperature distributions in Mount Hope Bay due to thermal discharges from the Brayton Point Station. *Northeast. Nat.*, 13, 145–172, doi:10.1656/1092-6194(2006)13[145:MOTDIM]2.0.CO;2.

Tiao, G. C., and Coauthors, 1990: Effects of autocorrelation and temporal sampling schemes on estimates of trend and spatial correlation. *J. Geophys. Res.*, 95, 20 507–20 517, doi:10.1029/JD095iD12p20507.

Ullman, D.S., C. Kincaid, C. Balt, and D.L. Codiga, 2019. Hydrodynamic modeling of Narragansett Bay in Support of the EcoGEM ecological model. GSO Technical Report 2019-01. University of Rhode Island, Graduate School of Oceanography. Narragansett, RI.

USEPA, 2015. National Coastal Condition Assessment 2010 (EPA 841-R-15-006). Office of Water and Office of Research and Development. Washington, DC. <http://www.epa.gov/national-aquatic-resource-surveys/ncca>

Wallace, EJ, Looney, LB, and Gong, D 2018. Multi-decadal trends and variability in temperature and salinity in the Mid-Atlantic Bight, Georges Bank, and Gulf of Maine. *Journal of Marine Research*, 76,163–215.

Weatherhead, E. C., and Coauthors, 1998: Factors affecting the detection of trends: Statistical considerations and applications to environmental data. *J. Geophys. Res.*, 103, 17 149–17 161, doi:10.1029/98JD00995.

YSI, 2012. 6-Series Multiparameter Water Quality Sondes User Manual. Yellow Springs International. Yellow Springs, Ohio. <https://www.ysi.com/File%20Library/Documents/Manuals/069300-YSI-6-Series-Manual-RevJ.pdf>

Acknowledgements

Little to none of this report could have been possible without many individuals generously sharing their datasets, expertise, and scientific as well as administrative guidance. Though an incomplete list, I would like to thank in particular the following:

David Borkman (RIDEM)
Mark Brush (VIMS)
Crystal Charbonneau (RIDEM)
Chris Deacutis (RIDEM)
Richard Freisner (NEIWPC)
Mike Gerel (NBEP)
Joe Habarek (RIDEM)
Kristin Huizenga (URI)
Q Kellogg (URI)
Jason Krumholz (McLaughlin)
Eliza Moore (NBC)
David Murray (Brown)
Candace Oviatt (URI)
Bill Patenaude (RIDEM)
Warren Prell (Brown)
Heather Radcliffe (NEIWPC)
Lew Rothstein (URI)
Courtney Schmidt (NBEP)
Heather Stoffel (URI)
Heidi Travers (RIDEM)
Dave Ullman (URI)
Molly Welsh (NBC)
Jeanne Wordell (Veolia)

Although the information in this document has been funded by the EPA, it has not undergone the EPA's publications review process and therefore, may not reflect the views of EPA and no official endorsement is inferred. The viewpoints expressed do not necessarily represent those of NEIWPC or EPA. Mention of trade names, commercial products, or causes do not constitute endorsement or recommendation for use.

This report is dedicated to Richard Maurice Codiga (1937-2019).