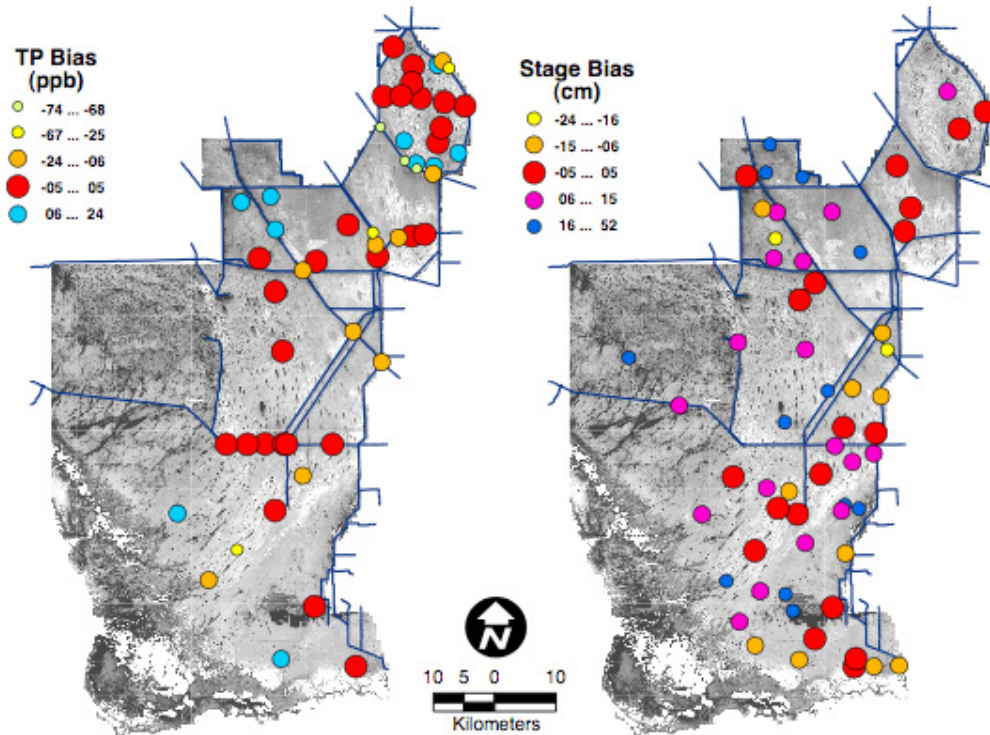
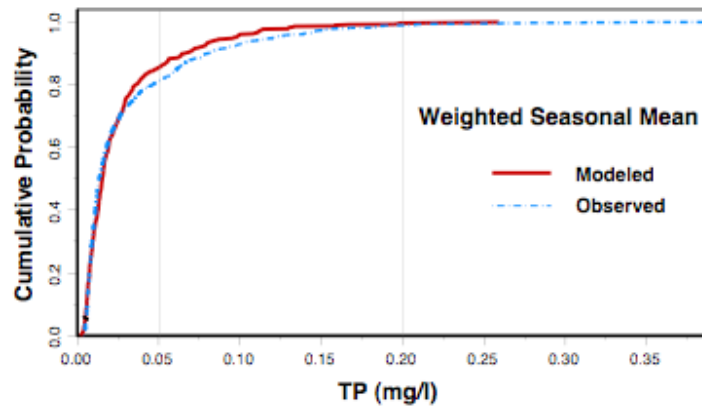


Calibration Performance of ELM v2.1a: 1979-1995 Water Quality and Hydrology



All stations



**Everglades Landscape Model Developers
Everglades Division, SFWMD**

Dec 21, 2002 Draft

<http://www.sfwmd.gov/org/wrp/elm/>

ELM Calibration Performance: Table of Contents

<i>Executive summary</i>	3
1. Background	6
ELM review	6
ELM versions	6
Observed data (targets)	6
Hydrology	6
Water quality.....	7
Observed data (boundary conditions).....	9
2. Calibration Performance Measures	10
Summary statistics	10
Stage heights	10
TP concentration - unweighted	10
TP concentration – seasonal means.....	11
Spatial distributions of summary statistics.....	11
Stage heights	11
TP concentration - unweighted	11
Temporal and cumulative distributions.....	11
Stage heights	11
TP concentration - unweighted	11
TP concentration - weighted	12
3. Synthesis	12
4. Literature cited	14
5. Table legends	15
6. Figure legends	15
7. Appendix: statistical methods	18
8. Tables	21
9. Figures	25

Executive summary

ELM objectives

One immediate objective, directly related to this report, is to apply the ELM in evaluations of hydrology and phosphorus water quality at regional and subregional scales for CERP¹ Projects, CERP RECOVER², and other Everglades Projects such as CSOP³. An updated version of ELM will soon provide other ecological performance measures across the regional model domain.

Document purpose

Following the initial two-month comment & response phase of the RECOVER Model Refinement Team (MRT) review of ELM, the request was made at the November 13 MRT meeting to provide additional numerical analyses of the performance of ELM in its calibration mode. In response, we expanded the statistical analyses of the comparisons between ELM and observed data (i.e., calibration performance measures for model “skill” assessment) for stage height and for Total Phosphorus (TP) concentration throughout the ELM domain. The information herein supplements the Oct 16, 2002 report⁴ “Agency/public review of ELM v. 2.1a: ELM developers’ response to reviews”.

Calibration performance

For general performance assessments, summary statistics including Bias and RMSE were used to assess the model calibration for predictive bias and accuracy. For the stage heights, the overall mean bias and RMSE for all monitoring stations were 6 cm and 23 cm, respectively, for predictions relative to observations. The overall mean bias and RMSE for surface water phosphorus concentration were -0.002 mg/l and 0.026 mg/l. When comparing seasonal means, the bias was 0.001 mg/l and the RMSE decreased to 0.013 mg/l.

Simulated stage heights explained 68% of variability in observed data. Compared to stage, the goodness of fit statistics for surface phosphorus concentration were lower, with an overall mean R^2 of 0.10 for the individual simulated & observed pairs. However, when weighted seasonal means were used, the average R^2 improved to 0.20.

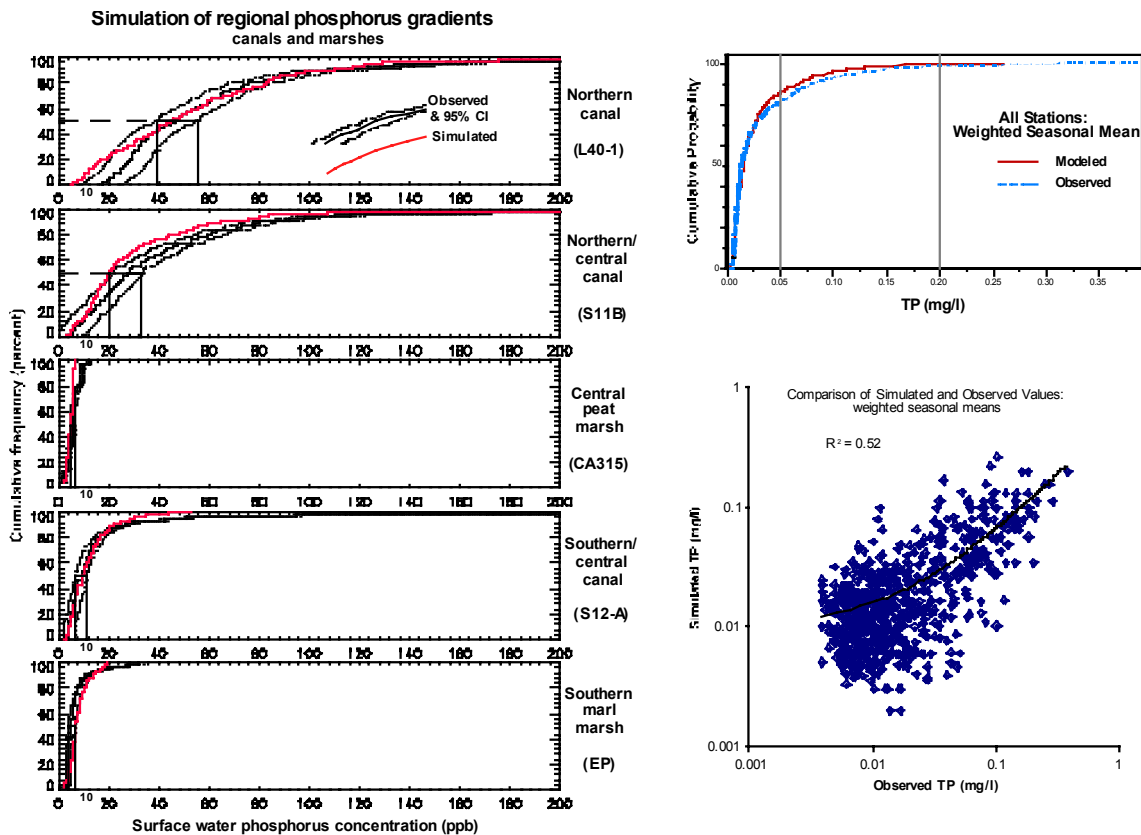
Such goodness of fit tests provide insight into the model capabilities, but measures of the magnitude of the model-observed deviations are critical to an evaluation of the efficacy, or “skill”, of the model in predictive mode. Overall differences between the model and observed data appeared to be within acceptable bounds for making water quality assessments on a regional basis. The spatial north-south trends in water quality, the monthly-seasonal dynamics, and many of the short-term nutrient pulses, were largely captured by the ELM simulation.

¹ Comprehensive Everglades Restoration Plan

² REstoration COordination and VERification

³ Combined Structural and Operational Plan

⁴ Available in the “News” section of the ELM web site <http://www.sfwmd.gov/org/wrp/elm/>



Summary figures: TP predictions: cumulative frequencies and correlations.

We used temporal aggregation to reduce the effects of random errors in observed data. Phosphorus concentration weighted by flow (water control structure stations) or by ponded water depth (marsh stations) is reflective of the relative importance of the total mass/volumes in the system, and can also reduce the influence of extreme values in some situations. When each simulated and observed seasonal mean surface water TP concentration at all monitoring stations were compared, simulated values explained more than 50% of variability in observed values. With further aggregation of seasonal means by each monitoring site, the R^2 increased to more than 0.60; i.e., 60% of the variance in observed values was explained by the simulation results. The cumulative frequency plot for weighted seasonal means for all sites also confirmed a good match of the simulated and observed data. The dry and wet seasonal means appeared to be a useful level of aggregation that was sufficient to minimize the influence of random error, while maintaining an appropriate temporal scale to account for wet/dry seasonal changes in surface water TP concentrations. Furthermore, at this level of aggregation, ELM clearly demonstrated the ability to predict overall seasonal mean phosphorus concentration changes with very good accuracy throughout the greater Everglades.

The various numerical analyses and visualizations in this document should demonstrate that the ELM is a useful predictive tool for hydrologic and water quality analyses in the Everglades. However, the strength of the ELM goes beyond merely predicting these “landscape drivers”. While we do not currently present Regional Performance Measures for the other ecosystem variables, these other ecological dynamics are key to understanding and evaluating management alternatives. The rates of growth and mortality of periphyton and macrophytes, the rates of peat accretion and oxidation, along with a number of other ecosystem processes, dynamically interact within the hydrologic

and biogeochemical cycles of the simulated ecosystems. We monitored these variables during this calibration process, ensuring that they remained within reasonable ranges. Importantly, these variable dynamics form the basis of changes to habitats that are defined by vegetation/periphyton community types and by soils. These changes to landscape attributes are the principal objectives of ELM simulations, and will form the primary basis for ELM ecological evaluations, as has been demonstrated for subregional scales in other publications⁵.

⁵ Available on web site at <http://www.sfwmd.gov/org/wrp/elm>

1. Background

ELM review

For a preliminary determination of the level of acceptance of the Everglades Landscape Model (ELM) for CERP⁶ application, the REstoration COordination and VERification (RECOVER) Model Refinement and Development Team (MRT) sponsored an inter-agency and public review of the ELM in fall 2002. The MRT initiated the review on August 7, 2002, comments on ELM were due in one month, and the ELM developers posted⁷ a ~90 page document on October 16 in response to reviewers' comments. At a November 13 MRT meeting, it was requested that we provide further statistical details of the calibration performance of the hydrologic (stage) and water quality (surface water phosphorus concentration) variables. The results herein provide such further information on the ELM v2.1a performance characteristics throughout the Everglades region for the calibration period 1979-1995.

ELM versions

The current version of ELM is the 2.1 family of updates to the model documentation, databases, and (some minor) codes. Version 2.1 was available in March 2000, and the code and data that drive the model are effectively unchanged from that time. Version 2.1a includes a large variety of enhancements to the model documentation and databases that were released⁸ for the initiation of the MRT review of ELM. In order to address two reviewers' comments, we made a small update (2.1a -> b) by coding an algorithm to evaluate numerical and actual dispersion (but which does not effect the output when switched off, as is the default case). To avoid confusion associated with incremental "version propagation", it is acceptable to refer to "**v2.1a**" as the **current version**, because the 2.1b update was minor and does not effect results in normal simulations.

Observed data (targets)

Hydrology

The hydrologic data used in this calibration assessment are water stage heights sampled during the period from January 1979 to December 1995. Using weekly⁹ observations with generally few missing data, the hydrologic data are of higher temporal quality, and the 60 stage gages generally have denser spatial distributions, than water quality observations. Figure 1 shows the spatial distribution of the stage monitoring locations used in the ELM calibration, and in the portion of the SFWMM domain within the greater Everglades. While there are subregions that should have denser spatial coverage in future monitoring plans, the current distribution provides a reasonable approximation of stage height differences along existing topographic gradients. The temporal sampling distribution is very appropriate relative to managed and natural stage variations at daily-

⁶ Comprehensive Everglades Restoration Plan

⁷ Available in the "News" section of the ELM web site <http://www.sfwmd.gov/org/wrp/elm/>

⁸ In Nov 2002, we found that we had erroneously released the "GlobalParms.xls" database with 5 parameters that did not coincide with those used in the "official" v2.1 calibration run whose results are posted on the ELM web site. These 5 parameters involved periphyton growth, mortality, and P uptake kinetics. Those parameters had been temporarily changed in an earlier, informal sensitivity analysis, but were never reverted back to the original values prior to posting the 2.1a database on Aug 7, 2002. Some changes in water column phosphorus will result if a run uses those altered (exploratory) parameters. All calibration results analyzed in this document use the "official" v2.1 calibration parameter set that corresponds to posted ELM v2.1 calibration results.

⁹ Stage observations are often recorded on an almost continuous basis, and daily stage means are generally available, with relatively few missing data for many monitoring stations, especially compared to water quality sampling.

weekly time scales, with missing data generally not an extremely important constraint in comparing simulated and observed data. Bias in sampling temporal changes in stage is generally not large relative to the modeling objectives, but there are some (relatively unknown) uncertainties associated with baseline elevation survey measurements.

For many statistics in hydrologic calibration analysis, we compare the ELM not only to observed data, but to the performance of the SFWMM¹⁰ v3.5, which is a managed hydrology simulation tool that has undergone extensive development and scrutiny over recent decades. While the ELM utilizes most of the same input data as used in the SFWMM, and has some similar hydrologic algorithms, the two models are largely independent of each other during calibration/validation simulations¹¹. Because of the ongoing application of the SFWMM for Everglades hydrologic evaluations, an important aspect of the ELM hydrologic calibration is comparison to results of the SFWMM.

Water quality

The water quality data used in this calibration assessment are Total Phosphorus (TP) concentrations sampled in the surface water column during the period from January 1979 to December 1995. Under past monitoring protocols, optimal sampling consisted of at most bi-weekly observations at most water quality monitoring stations. However, the monitoring at all stations often had much less frequent sampling resulting in months or years of missing data at irregular intervals. The spatial network of 57 monitoring locations is also relatively sparse compared to hydrologic monitoring (Figure 2). Directly related to these spatio-temporal constraints is the highly dynamic nature of phosphorus in wetland systems across both space and time. Because water column phosphorus concentrations can rapidly respond to a variety of “fast” biotic and abiotic processes on time scales from hours to days, these water quality data have higher spatial and temporal variability, and usually greater uncertainty, compared to observed hydrologic stages.

Nutrients respond rapidly in the environment along multiple pathways of canal and overland flows within the managed Everglades. There are many monitoring locations that should have denser spatial coverage in future monitoring plans in order to more effectively characterize the changing nutrient gradients; the current distribution in virtually all areas does not provide a complete understanding of the spatial differences along existing (and future) gradients. Considering the high sampling and processing cost, the (bi-weekly) temporal sampling frequency may be adequate to characterize long term trends in water quality variables that respond at daily-weekly time scales. However, the extensive multi-year or multi-month periods of missing data in the existing (1979-95) data set pose very significant constraints when comparing the simulated and observed water quality data. Moreover, bias in sampling temporal changes in water column nutrients is another significant source of uncertainty that is directly related to the low spatio-temporal sampling intensity. Because of the response of phosphorus to multiple environmental processes, some of which are stochastic or relatively unpredictable, infrequent water quality sampling is notoriously prone to relatively high data variability. Additionally, simulations of hydrologic flows compounds the uncertainty associated with any water quality simulation, as nutrients increase and decrease along the flow paths. Thus, compared to stage, the uncertainties in the observed water quality data, with low sample sizes relative to often broad data distributions, generally result in comparatively poor measures of the goodness-of-fit of instantaneous (paired observed and simulated) data points. As is generally the case with the existing data, it is frequently most informative to temporally aggregate water quality observations in order to evaluate monthly, seasonal, or annual trends in the simulated vs. observed data.

Flow-weighted mean:

¹⁰ <http://www.sfwmd.gov/org/pld/hsm/models/index.html>

¹¹ For CERP applications, the daily flows through each water control structure in the ELM simulation are driven by output of a SFWMM simulation run.

Water column phosphorus concentration is one of the important measures of water quality. But because phosphorus responds so rapidly in the environment, concentration alone may be an inadequate descriptor of the nutrient status of the system. The nutrient mass that is loaded into the system, and the mass that accumulates within a specific spatial area, are critical components of characterizing the magnitude of nutrient fluxes in the system. Nutrient mass loads are calculated from continuous measures (and summation) of water flow volumes and the associated nutrient concentrations. Related to a continuous nutrient load calculation is that of a flow-weighted nutrient concentration, which can be conceptualized simply as the sum of the mass loads divided by the sum of the flow volumes for all flow events¹². For a series of flow events at a particular water control structure, a flow weighted mean concentration is indicative of the relative “importance”, or weight, of each observed concentration relative to the potential mass loading that can impact the ecosystem. For example, for a given nutrient concentration at a water control structure, a relatively low flow would likely have less of a downstream impact compared to a higher flow (which would pass more nutrients into the receiving waters).

Depth-weighted mean:

Such flow weighted concentrations are routinely used in analysis of nutrients at water control structures in the Everglades, where flow observations are usually available. There is no direct equivalent for monitoring locations at marsh (or unstructured canal) flow locations. While ELM output provides flow and mass load calculations for overland (and groundwater) flows through an expanse of wetlands, equivalent flow observations in the field have been effectively unavailable for comparison. For some consistency with the flow weighted mean metric, and to potentially gain some insight into the relative nutrient masses in a depth-varying marsh, we investigated the use of a metric that parallels the flow-weighted concept. In this depth-weighted mean concentration, we weight the marsh nutrient concentration by the ponded surface water depth instead of the flow volume. The result is indicative of the total mass of phosphorus per marsh surface area (but does not reflect its immediate availability for biotic/abiotic processes).

Time series analysis and observed sample sizes:

Stage levels have been the focus of spectral analysis quite readily for several decades (Mandelbrot and Wallis 1969) and have repeatedly demonstrated a spectral exponent around two ($\beta \approx 2$) (see Appendix). This behavior holds true for the majority of weekly stage observations (mean $\beta = 1.62$, ± 0.045 S.E.) and ELM simulation results (mean $\beta = 1.85$, ± 0.025 S.E.). In Figure 3.1 this relationship is visually evident for the NP203 gauging station as the best-fit representation of the average system behavior. Both the observation and model data align fairly well with one another, as well as with published accounts of several other datasets. With stage power spectra results near 2 we can speculate that the behavior of this system is somewhat synonymous with fractional Brownian motion. As the weekly observations exhibit memory-effects to their neighboring weeks, this can be thought of as exhibiting mild-autocorrelation. Seasonal trends are also evident, shown as peaks that clearly prevail in both the observations and model results at frequencies between 21 and 24 weeks or ~ 5 -6 months.

While the temporal quality of (field) stage observations appears sufficient to characterize time-varying depths in the Everglades, the available field observations of water column TP concentrations appear to have a much more random or stochastic signal characteristic. This makes it problematic to infer temporal characteristics of sampling locations' nutrient status. In general, there is a lack of adequate observation data for time series evaluation over the period of record. While the ELM TP concentration power spectra are based upon daily output

$$^{12} Cf_w = \frac{\sum_{i=1}^n Q_i C_i}{\sum_{i=1}^n Q_i}, \text{ where } Cf_w \text{ is the flow-weighted mean concentration of the nutrient (mg L}^{-1}\text{) over the}$$

n flow events, Q_i is the flow volume (L) during event i , and C_i is the actual concentration of the nutrient during event i .

concentrations, the observed datasets vary considerably in their sampling frequency. In fact, the average time between sampling intervals when samples are taken is typically more than (and frequently much more than) a month, and the average number of samples for the 6209 day (17 yr) model run is ~131. This being said, it is quite understandable why the power spectra (Figure 3.2) of the observations are falling within the range of a random or stochastic process (mean $\beta = 0.07$, ± 0.065 S.E.). This is contrary to the model spectra (mean $\beta = 1.57$, ± 0.055 S.E.) that fall between the range of self-organized criticality and Brownian motion. We are not suggesting that the model spectra are necessarily appropriate to estimate the true power spectra for total phosphorus, but simply that in many cases, the sampling observations are of relatively poor temporal quality. As a note however, the model spectral exponents do coincide with recently discovered water quality trends seen in the more conservative daily concentrations of sodium and chloride tracers (Kirchner et al. 2000, 2001).

Furthermore, as described in later sections (Calibration Performance Measures and the Appendix), we developed an aggregate index of the variety of statistical metrics (Figure 4.1) considered for the calibration analysis. This Relative Performance Index allowed us not only to evaluate relative performance via one index, but provided an overall indication of the adequacy of observed sample sizes to provide useful comparisons with model results. This index suggested that simulated-observed comparisons were greatly affected by the number of observations per location (Figure 4.2). From this and other metrics, we can infer that a number of water quality monitoring locations had somewhat inadequate sample sizes for comparison with the model.

Observed data (boundary conditions)

The ELM uses a variety of time varying, daily observations of hydrologic and meteorological data for external forcings or boundary conditions of: water control structure flows, rainfall, windspeed, cloud cover, air temperature, and dew point temperature. For consistency among models, all of those data were provided to us from the developers of the SFWMM, and are the same data used to drive the SFWMM.

While the daily mean water control structure flow observations¹³ are generally available for all structures within the ELM domain, we also require the phosphorus concentrations associated with all inflows into the model domain. However, due to the limited sampling frequency of water quality data, an interpolation among data points must be used to fill in the extensive missing data. For the current model version, those daily inflow concentration data were calculated and provided to us by W. W. Walker¹⁴.

In preparing to update the calibration (and validation) period of record of ELM from 1979-1995 to 1979-2000, we obtained updated daily (1979-2000) flow and concentration data from C. Mo¹⁵. In comparing the new (retrieved in 2002) and old (retrieved in 1997) flow and concentration datasets for the period 1979-1995, we found significant differences between the old and new flow data. For some structures and some years, flows and loads were substantially different among datasets; in an example for the S-8 structure, a number of years had flow differences on the order of 10% or more among the datasets (Figure 5.1); some annual TP loads were different by more than 10 metric tons (Figure 5.2). Preliminary analysis showed that most of changes in TP load

¹³ Depending on the structure, these are often calculations based on flow rating curves and observations of head and tailwater stages.

¹⁴ Data as personal communication, W.W. Walker, Consultant to Department of Interior, 1997. His computer program generates daily phosphorus concentrations for all dates containing non-zero water control structure flows by selecting the appropriate (grab vs. autosampler) nutrient observations and performing a linear interpolation between available data.

¹⁵ Data as personal communication, C. Mo, SFWMD Department of Environmental Monitoring and Assessment, July 2002. The program originally developed by WW. Walker has continued to be modified, and was again used in this most recent data download from the SFWMD DBHYDRO database.

were directly related to the changes in flow data¹⁶. The changes to the DBHYDRO database between these data retrievals are currently being reviewed by various Departments at the SFWMD, and we will use the verified set of flow data for all water control structures that results from that review.

2. Calibration Performance Measures

No single statistic provides a complete understanding of the ELM performance, and this is particularly true with the water quality analysis due to the constraints imposed by spatio-temporal sampling distributions. In the case of highly dynamic variables (i.e., phosphorus) with coarse spatial sampling distributions, infrequent monitoring, and extensive missing data, it is particularly important to consider a variety of statistics that bring out different attributes of the observed and simulated data sets. We developed three fundamental sets of evaluation tools that vary in their level of temporal aggregation and spatio-temporal perspective: 1) **tables** of summary statistics that synthesize the entire period of performance for each station location; 2) **maps** of selected summary statistics that provide a spatial perspective on the performance throughout the region; and 3) **time series graphs** with confidence intervals at multiple scales of temporal aggregation, including cumulative frequency distribution plots.

Summary statistics

We combined all sampling locations for correlation analyses (Figure 6.1) that compared the simulated and observed values of (the weighted and the unweighted) TP concentrations to obtain an overall summary of the model skill in predicting TP concentration in the surface water (1979-95). A comparison of the cumulative frequency distribution of weighted seasonal means of modeled and observed data is presented in Figure 6.2. These figures provide strong evidence that the model appears to capture the observed distribution of concentrations very well overall.

A series of tables follow, summarizing various statistical measures for each point location, with each statistic indicative of the overall performance of the ELM predictions across the entire simulation time period (1979-95).

Stage heights

Table 1 summarizes the key goodness-of-fit and summary statistics for each stage monitoring location in the ELM domain. The Bias, R^2 , RMSE, Model Efficiency, and Theil's Inequality Coefficient U^2 statistics (see Appendix 1) are shown for the ELMv2.1a predictions vs. observed data and for the SFWMMv3.5 predictions vs. observed data.

TP concentration - unweighted

For (unweighted) surface water TP concentration data, Table 2 summarizes the key goodness-of-fit and summary statistics for each water quality monitoring location in the ELM domain. The Bias, R^2 , RMSE, Efficiency, and Theil's Inequality Coefficient U^2 statistics (see Appendix 1) are shown for the ELMv2.1a predictions vs. observed data. We also show the 95% Confidence Intervals of the observed data, the simulation mean, and the difference between geometric means of simulated and observed values

¹⁶ We have not thoroughly examined all of the recently retrieved data, as we are awaiting the final, verified data set prior to extensive analysis. However, some of the counter-intuitive differences, such as higher flow associated with lower load when comparing the datasets can be the result of changes in the presence/absence of non-zero daily flows in the two datasets.

TP concentration – seasonal means

For seasonal means of surface water TP concentration data, Table 3 summarizes the key goodness-of-fit and summary statistics for each water quality monitoring location in the ELM domain.

Spatial distributions of summary statistics

The following series of maps provide a perspective of the (presence or absence of) spatial patterns in ELM calibration performance. As previously discussed, there is no single statistic that provides a single “best” method of evaluating the complexities of hydrologic and water quality predictions from a model. In order to integrate the multiple statistics, we developed a Relative Performance Index that ranks (using equal weights) and combines six separate statistics. This Relative Performance Index allows visualization of any spatial pattern in relative (not absolute) performance of ELM. See the Appendix for descriptions of this, and other, statistics presented.

Stage heights

Figures 7.1-2 provide spatial perspectives on the performance of ELM hydrology: Figure 7.1 maps the Relative Performance Index; Figure 7.2 maps the Bias of the simulated data relative to observed data. Neither metric shows any consistent spatial trend in performance.

TP concentration - unweighted

For unweighted TP concentrations, Figures 8.1-6 provide spatial perspectives on the performance of ELM water quality. In Figure 8.1, the spatial coverage is shown, with somewhat high density in WCA-1, and sparse coverage in most of the rest of the system. The overall Relative Performance Index (Figure 8.2) does not appear to indicate any particular spatial pattern across the region. The RMSE and Bias statistics (Figure 8.3 and Figure 8.4) show that many locations throughout the region had predicted concentrations that were within 5 ppb or within 10 ppb of the observed data. Similarly, the Geometric Mean Difference (Figure 8.5) between simulated and observed data was frequently less than (the absolute value of) 5 ppb across space. Finally, the goodness-of-fit measure (Figure 8.6), while not as high as those for stage predictions, varied without clear trend across the region.

Temporal and cumulative distributions

The following series of plots show the distribution of simulated and observed data at varying levels of temporal aggregation. Particularly for the highly variable water quality data, these perspectives provide an important synthesis of the system dynamics.

Stage heights

Figures 9.1-60 show the match of the SFWMM, ELM, and observed data at multiple levels of temporal aggregation. In general, both models appear to capture the wide range of temporal variations in stage height over the 17 year period that encompassed extreme drought and flood conditions.

TP concentration - unweighted

Figures 10.1-57 provide a perspective of the temporal variability of raw, unweighted concentration data. While occasionally individual observation points show high extreme values, the model captures many of the trends due to changes in managed flows or drydowns, and demonstrates the regional gradient in surface water eutrophication from north to south.

TP concentration - weighted

Figures 11.1-52 generally demonstrate an even closer match of simulated to observed data due to weighting the TP concentrations by the flow through water control structures or the depth of the water in the marsh (although the latter weighting factor may not show as strong of a performance increase).

3. Synthesis

For general performance assessments, summary statistics including Bias and RMSE were used to assess the model calibration for predictive bias and accuracy. For the stage heights, the overall mean bias and RMSE for all monitoring stations were 6 cm and 23 cm, respectively, for predictions relative to observations (Table 1). The SFWMM had slightly better calibration at these same monitoring stations, with overall means of 1 cm Bias and 18 cm RMSE. The overall mean bias and RMSE for surface water phosphorus concentration were -0.002 mg/l and 0.026 mg/l (Table 2). When comparing seasonal means, these bias was 0.001 mg/l and the RMSE decreased to 0.013 mg/l.

The goodness of fit statistics (correlation coefficient- R^2 , model efficiency-EFF and Theil's inequality coefficient- U^2) described the agreement between simulated and observed data via a paired point to point comparison. For stage heights, simulated values from both ELM and SFWMM explained 68% of variability in observed stage heights (Table 1). There was also little difference between the two models in the U^2 statistics: 0.018 for ELM and 0.012 for SFWMM. The large difference in goodness of fit evaluated by model efficiency between the two model was principally due to sites in Big Cypress NP, which we had not previously evaluated. Compared to stage heights, the goodness of fit statistics for surface phosphorus concentration were lower, with an overall mean R^2 of 0.10 for the individual simulated & observed pairs (Table 2). However, when weighted seasonal means were used, the average R^2 improved to 0.20. The model efficiency and U^2 also exhibited an improved goodness of fit when analyzing weighted seasonal means.

Such goodness of fit tests provide insight into the model capabilities, but measures of the magnitude of the model-observed deviations are critical to an evaluation of the efficacy, or “skill” of the model in predictive mode. While there are some locations where the model predictions should be, and will be, improved in future versions, overall differences between the model and observed data appear to be within acceptable bounds for making water quality assessments on a regional basis. We do not attempt to summarize the multiple visualization plots of simulated-observed time series data, but perusal of such temporal dynamics, as one moves from the eutrophic northern sites to the southern regions of lower phosphorus concentrations, it should be clear that the spatial north-south trends, and the monthly-seasonal dynamics, are largely captured by the ELM simulation.

There are many factors affect water quality measurements, and ecological/water quality predictions of those dynamics. For example, while simulated phosphorus concentration is actually a mean concentration in one kilometer grid, the measured phosphorus concentration at a point location within the one kilometer grid may or may not represent the average condition and can be affected by many random effects. Because of the high variability of water quality data and random errors in it monitoring, an exact match between individual modeled and observed water phosphorus is difficult. When the number of observation is large, random samples do not increase bias (Scheaffer et al, 1986; Dixon and Garrett, 1993), thus the random errors can be canceled out by aggregation.

We used temporal aggregation to reduces the effects of random errors in observed data. Furthermore, phosphorus concentration weighted by flow (water control structure stations) or by ponded water depth (marsh stations) is reflective of the relative importance of the total mass/volumes in the system, and can also reduce the influence of extreme values in some situations. As noted, the average R^2 value for weighted seasonal means increased to 0.20 (Table

3) compared to 0.10 before aggregation (Table 2). When each simulated and observed seasonal mean surface water TP concentration at all monitoring stations were compared, simulated values explained more than 50% of variability in observed values (Figure 6.1). With further aggregation of seasonal means by each monitoring site, the R^2 increases to more than 0.60; i.e., 60% of the variance in observed values was explained by the simulation results. The cumulative frequency plot for weighted seasonal means for all sites also confirmed a good match of the simulated and observed data (Figure 6.2). The dry and wet seasonal means appeared to be a useful level of aggregation that was sufficient to minimize the influence of random error, while maintaining an appropriate temporal scale to account for wet/dry seasonal changes in surface water TP concentrations. Furthermore, at this level of aggregation, ELM clearly demonstrates the ability to predict overall seasonal mean phosphorus concentration changes with very good accuracy throughout the greater Everglades.

The various numerical analyses and visualizations in this document should demonstrate that the ELM is a useful predictive tool for hydrologic and water quality analyses in the Everglades. However, the strength of the ELM goes beyond merely predicting these “landscape drivers”. While we do not currently present Regional Performance Measures for the other ecosystem variables, these other ecological dynamics are key to understanding and evaluating management alternatives. The rates of growth and mortality of periphyton and macrophytes, the rates of peat accretion and oxidation, along with a number of other ecosystem processes, dynamically interact within the hydrologic and biogeochemical cycles of the simulated ecosystems. We monitor these variables during the calibration process, ensuring that they remain within reasonable values. Importantly, these variable dynamics form the basis of changes to habitats that are defined by vegetation/periphyton community types and by soils. These changes to landscape attributes are the principal objectives of ELM simulations, and will form the primary basis for ELM ecological evaluations, as has been demonstrated for subregional scales in other publications¹⁷.

¹⁷ Available on web site at <http://www.sfwmd.gov/org/wrp/elm>

4. Literature cited

- Dixon, P.M. and K.A. Garrett. 1993. Sampling ecological information: choice of sample size, reconsidered. *Ecol. Mod.* 68:67-73.
- Janssen, P.H.M., and P.S.C. Heuberger. 1995. Calibration of process-oriented models. *Ecol. Mod.* 83:55-66.
- Kirchner, J. W., X. Feng, and C. Neal. 2000. Fractal stream chemistry and its implications for contaminant transport in catchments. *Nature* **403**:524-527.
- Kirchner, J. W., X. Feng, and C. Neal. 2001. Catchment-scale advection and dispersion as a mechanism for fractal scaling in stream tracer concentrations. *Journal of Hydrology* **254**:82-101.
- Mandelbrot, B. B., and J. R. Wallis. 1969. Some long-run properties of geophysical records. *Water Resources Research* **5**:321-340.
- Mayer, D.G., and D.G. Butler. 1993. Statistical validation. *Ecol. Mod.* 68:21-32.
- McLeod, A. I., and K. W. Hipel. 1995. Exploratory spectral analysis of hydrological time series. *Stochastic Hydrology and Hydraulics* **9**:171-205
- Power, M. 1993. The predictive validation of ecological and environmental models. *Ecol. Mod.* 68:33-50
- Scheaffer, R.L., W. Mendenhall, and L. Ott. 1986. *Elementary Survey Sampling*. PWS-Kent, Boston, MA, 324pp.

5. Table legends

Table 1. Summary statistics for predictions of stage height (ELM v2.1a and SFWMM v. 3.5).

Table 2. Summary statistics for predictions of surface water TP concentration (ELM v2.1a).

Table 3. Summary statistics seasonal means of surface water (flow- or depth- weighted, and unweighted) TP concentration (ELM v2.1a).

6. Figure legends

Figure 1. Locations of stage gages used in ELM v2.1a calibration.

Figure 2. Locations of water quality monitoring stations used in ELM v2.1a calibration.

Figure 3.1. Power spectrum of observed and simulated stage data.

Figure 3.2. Power spectrum of observed and simulated TP concentration data.

Figure 4.1. Principal Components Analysis for a variety of statistical metrics that compare the observed and simulated TP concentration data.

Figure 4.2. The effect of the number of field TP observations on the overall Relative Performance Index.

Figure 5.1. Inflows through water control structure S-8, as determined from a 1997 and a 2002 database retrieval.

Figure 5.2. TP loads through water control structure S-8, as determined from a 1997 and a 2002 database retrieval.

Figure 6.1. R^2 correlation coefficient of seasonal mean surface water TP concentration at all monitoring stations (ELM v2.1a).

Figure 6.2. Cumulative frequency distributions of weighted seasonal means of simulated and observed surface water TP concentrations at all monitoring stations.

Figure 7.1. Spatial distribution of the Relative Performance Index of simulated stage heights (ELM v2.1a).

Figure 7.2. Spatial distribution of the Bias statistic for observed and simulated stage heights.

Figure 8.1. Spatial distribution of the number of field observations of surface water TP concentration (used in calibration of ELM v2.1a).

Figure 8.2. Spatial distribution of the Relative Performance Index for simulated surface water TP concentration.

Figure 8.3. Spatial distribution of the Root Mean Square Error of simulated surface water TP concentration relative to observed.

Figure 8.4. Spatial distribution of the Bias of simulated surface water TP concentration relative to observed.

Figure 8.5. Spatial distribution of the Geometric Mean Difference between observed and simulated surface water TP concentration.

Figure 8.6. Spatial distribution of the R^2 correlation coefficient between observed and simulated surface water TP concentration.

Figures 9.1-60. Observed and simulated water stage heights (ELM v2.1a and SFWMM v3.5): time series for all modeled and observed data; time series for seasonal means with confidence intervals & T-tests; and cumulative frequency distributions. The series of plots in these figures follow a common format for each stage gage location. The ELM (1979-95) weekly snapshot output is the continuous red line; the SFWMM (1979-95) weekly snapshot output is the continuous green line; the available daily observed data points are shown in black symbols. For each monitoring location, there are four plots:

- a) Raw data with **no temporal aggregation**.
- b) All data aggregated into arithmetic **mean values by wet and dry seasons** within water years; the continuous lines pass through mean of all simulated data points for each season; the mean of paired ELM-simulated & observed values are shown in boxes and diamonds, respectively; the 95% Confidence Interval of the paired means are shown by the "—" symbols in the red and black colors of the ELM-simulation and observed data, respectively; result of a T-test at $\alpha=0.05$, comparing ELM-simulated and observed data (significant="*", not significant="NS").
- c) All data aggregated into arithmetic **mean values by water years**, with the same statistics that are described in b).
- d) The **cumulative frequency distributions** of the simulated and observed (raw, un-aggregated) data; the 95% confidence interval for observed data is shown in the dashed black lines. Note that only paired simulated and observed data points are used.

Figures 10.1-57. Observed and simulated TP (unweighted) concentration: a) time series for all modeled and observed data (top); b) time series for seasonal means with confidence intervals & T-tests (middle); and c) cumulative frequency distributions (bottom). The ELM (1979-95) daily snapshot output is the continuous red line; the available daily observed data points are shown in black symbols. For each monitoring location, there are four plots of unweighted TP concentration data that utilize the same temporal aggregation and statistical methods described in Figure 9 for stage. Note: There are several sets of water quality observations within canals (usually associated with water control structures) that are within a single, unsegmented canal reach in the ELM simulation. We compare the simulated concentration (within that reach) to observations at each individual location, but also to the mean of the multiple observations within that reach, with the latter designated with the suffix "-M". This was done for the L-40, S-10, S-11, S-144-6, S-12 sets of observation points.

Figures 11.1-52. Observed and simulated TP (flow- or depth- weighted) concentration: a) time series for all modeled and observed data (top); b) time series for seasonal means with confidence intervals & T-tests (middle); and c) cumulative frequency distributions (bottom). The series of plots in these figures follow a common format for each water quality station location. The ELM (1979-95) daily snapshot output is the continuous red line; the available daily observed data points are shown in black symbols. For each monitoring location, there are four plots:

- a) Raw data with **no temporal aggregation**. (These plots are identical to those shown in the "Raw Data" plots of the Figure 10 series of TP concentrations, as they utilize unweighted, individual data points).
- b) All data aggregated into arithmetic **mean values by wet and dry seasons** within water years; same syntax and statistics described in Figure 9b series for stage.

- c) All data aggregated into arithmetic **mean values by water years**; same syntax and statistics described in Figure 9c series for stage.
- d) The **cumulative frequency distributions** of the simulated and observed seasonal means; the 95% confidence interval for observed data is shown in the dashed black lines. Note that only seasons containing both simulated and observed data are used.

7. Appendix: statistical methods

Time Series Analysis

Spectral analysis has proven to be one of the most effective and useful tools for exploring and examining time series data (McLeod and Hipel 1995). The root of this technique was spawned from the development of the Fourier transform (Mandelbrot and Wallis 1969). Specifically the Fourier series can be defined by the frequency representation of a signal's auto-covariance. It is estimated by direct integration of condensing an oscillating sine wave with a temporal signal. The result is a power spectrum of the amplitudes corresponding to incrementing frequencies. The strength of this technique is its ability to characterize the stochasticity and interdependence of a signal.

For decades researchers have been developing insight into the nature of these spectra (Mandelbrot and Wallis 1969). In a simple sense these plots can be directly evaluated by their linearity and slope. For instance, when plotted as the $\text{LOG}(\text{amplitude}^2) / \text{LOG}(\text{frequency})$, a random time-series or white-noise signal exhibits little or no negative slope ($m > -1$), while a self-similar or fractal time-series (e.g. Brownian Motion or hydrologic stage) would typically exhibit a relatively steep negative slope ($m = \sim -2$). The importance of the spectral slope has taken on the ubiquitous title of spectral exponent "β"; also commonly represented in the form $1/f^\beta$. In an attempt to address questions concerning the validity of our model's temporal results on a location-by-location basis, we calculated the power spectra for all of the stage and total phosphorus observed and model time-series data.

Bias

$$\text{Bias} = \frac{\sum (y - x)}{n}$$

Where x is observed values, y is model simulated values and n as the number of observations.

Bias is calculated as the mean differences between paired modeled and observed values. It is a measure of how biased overall values simulated by the model from the observed values. The bias should be as close to zero as possible.

Geometric Mean difference (gMean DIF)

$$\text{gMean DIF} = \sqrt[n]{y_1 y_2 \cdots y_n} - \sqrt[n]{x_1 x_2 \cdots x_n}$$

The geometric mean is the n th root of the product of the n observations. Like the arithmetic mean, geometric mean gives another indication of the central tendency of the data, specially the data is lognormal distributed.

Root Mean Square Error (RMSE)

$$\text{RMSE} = \sqrt{\frac{\sum (y - x)^2}{n - 1}}$$

Where x is the observed value, and y is the model prediction.

RMSE is the square root of the average values of the prediction errors squared. It measure the discrepancy between modeled and observed values on an individual level. It indicates predictive accuracy of the model. Because of the quadratic term, it gives greater weight to larger discrepancy than to smaller one.

Pearson product-moment correlation coefficient (R^2)

$$R^2 = \left(\frac{\sum (y - y_m)(x - x_m)}{\sqrt{\sum (y - y_m)^2 \sum (x - x_m)^2}} \right)^2$$

Where x_m is the mean of observed x as $(\sum x)/n$ and y_m is the mean of observed y as $(\sum y)/n$.

The R^2 measure the degree of linear association between x and y . It represents the amount of variability of one variable that is explained by correlating it with another variable. Depending on the strength of the linear relationships, the R^2 varies from 0 to 1 with 1 indicating the perfect fit.

Model Efficiency

$$\text{Eff} = 1 - \frac{\sum (y - x)^2}{\sum (x - x_m)^2},$$

Where x_m is the mean of the observed x , and y is the model prediction.

Like correlation coefficient, model efficiency is an another overall indication of goodness of fit (Mayer and Butler, 1993; Janssen, and Heuberger, 1995). It is the sum of squared prediction errors divided by the sum of squared deviation of observed values from the mean. A model efficiency of one indicates zero sums of squares of $(y - x)$, hence a perfect fit between modeled and observed values. A value of zero efficiency indicates the fit to $y = x$ is no better than $x = x_m$.

Theil's Inequality Coefficient (U^2)

$$U^2 = \frac{\sum (y - x)^2}{\sum y^2}$$

Theil's inequality efficiency was originally designed to evaluate econometric model forecasts and can be applied to environmental and ecological models as well (Power, 1993). When $y = x$ then $U^2 = 0$ and a perfect prediction is reached. Any Theil's coefficient less than one indicates that modeled values are an improvement relative to a naïve zero-change prediction.

T-Test

We use T-test to check the significance of the difference in means between paired modeled and observed values. The probability values (p) of the two means are equal are calculated and checked against an significant α level of 0.05.

Relative Performance Index

To address the question of relative performance throughout the spatial domain of the model, we synthesized an aggregated statistical index for each Total Phosphorus and stage monitoring location based on the results of their statistical comparison with field observations. Through the plethora of statistical analyses we performed on these datasets, we arrived at the question of: Which statistical indices were contributing novel information? Therefore, we wanted to select a collection of statistical analyses that captured the full range of variation that coexisted among the datasets, while excluding indices that exhibit redundancy to one another. The first step to selecting such key statistics was to capture the correlation / variance that existed among our suite of analyses. Using principal components analysis (PCA) (Figure 4.1) we were able to clearly define five essential groups of statistics. As demonstrated by different hues of color in Figure 4.1, the groups were: (blue) temporal aggregation significant difference tests, (green) parametric and non-parametric paired comparison tests, (yellow) error and offset tests, (orange) calculated differences between descriptive statistics, and (red) time-series based differences and efficiency. From these five groups we selected six key statistics (demonstrated by their darker shades in Figure 4.1; seasonally

aggregated T-test, R^2 , RMSE, difference in the geometric mean between model and observations, bias, and efficiency) to synthesize an index of relative model performance. We simply ranked each of the comparison locations based on their quality respective of each statistic (e.g. a location with an $R^2 = 0.95$ would have a better rank than one with an $R^2 = 0.5$). We then summarized the ranks for each location to produce an overall assessment of relative performance (Figures 7.1 and 8.2). Hence, this allowed us to grasp the power of several statistical techniques via one index, but also provided an overall indication of the adequacy of observed sample sizes to provide meaningful comparisons with model results. This index additionally suggested that the ability of field point observations to be compared with model results are highly affected by the number of observations per location (Figure 4.2). From this and other metrics, we can infer that a number of water quality monitoring locations had somewhat inadequate sample sizes for comparison with the model.

Outlier TP observations

Observed TP concentration observations: outliers that were not included in this calibration analysis.

STATION	DATE	TP	COMMENTS (from S. Hill, SFWMD Environmental Monitoring and Assessment)
C123SR84	09/29/94	262	TP value very high relative to historical values & associated parameter values. Color value of 144 may affect TP measurement
L7	01/29/79	1415	TP value very high relative to historical values & associated sample parameter values. Site no longer sampled
LOX16	05/16/95	78	TP value very high relative to historical values & associated sample parameter values. Color value of 70 may affect TP measurement
LOX5	02/08/95	80	TP value very high relative to historical values & associated sample parameter values. Color value of 100 may affect TP measurement
P33	03/06/89	546	TP & TN values very high relative to historical values. Possibly due to TSS of 180, Color of 91 & turbidity of 106.
P36	03/06/89	1137	TP & TN values very high relative to historical values. Possibly due to turbidity of 50, color of 69 & tss of 118.
S10C	07/25/90	3435	TP & TN values very high relative to historical values. Possibly due to turbidity of 63, color of 282 & tss of 145.
S10D	08/31/82	1347	TP & TN values very high relative to historical values. Possibly due to turbidity of 38, color of 219.
S10E	02/19/91	484	TP value high relative to historical values. Possibly due to turbidity of 115 & color of 105.
S10E	05/10/93	493	TP value high relative to historical values. Color of 105 may affect TP concentration measurement.
S12B	03/04/85	593	TP value high relative to historical values.
S12B	02/04/92	484	TP value high relative to historical values.

8. Tables

Following this page are three summary tables.

Table 1. Stage summary statistics.

Location	OBS n	ELM vs OBS					WMM vs OBS				
		Bias	R ²	RMSE	EFF	U ²	Bias	R ²	RMSE	EFF	U ²
		meter		meter			meter		meter		
1-7	416	0.06	0.73	0.16	0.33	0.001	0.00	0.71	0.15	0.44	0.001
1-8T	368	0.04	0.67	0.23	0.06	0.002	0.11	0.73	0.19	0.35	0.001
1-9	413	0.00	0.72	0.15	0.50	0.001	0.08	0.72	0.17	0.35	0.001
2A-159	162	-0.05	0.63	0.22	0.51	0.003	-0.06	0.60	0.22	0.52	0.003
2A-17	816	-0.04	0.65	0.24	0.43	0.004	-0.06	0.76	0.18	0.69	0.002
2A-300	662	-0.05	0.56	0.23	0.46	0.004	-0.06	0.70	0.20	0.63	0.003
3-34*	131	-0.09	0.84	0.16	-1.70	0.000	-0.12	0.51	0.19	-0.68	0.000
3-69*	200	0.52	0.86	0.52	-10.02	0.004	0.49	0.84	0.50	-3.88	0.003
3-71	204	-0.09	0.68	0.14	0.35	0.004	-0.03	0.34	0.15	0.31	0.004
3-76	198	-0.07	0.66	0.12	0.46	0.003	-0.08	0.23	0.19	-0.37	0.007
3A-10*	735	0.06	0.64	0.34	0.51	0.001	0.00	0.61	0.19	0.61	0.000
3A-11	695	-0.24	0.78	0.34	-1.25	0.011	-0.22	0.74	0.27	-0.42	0.007
3A-12	769	0.10	0.66	0.25	0.25	0.006	0.04	0.62	0.20	0.51	0.004
3A-2	429	0.06	0.59	0.26	0.53	0.006	0.02	0.58	0.25	0.57	0.006
3A-28	755	0.29	0.83	0.31	-0.19	0.010	0.02	0.84	0.13	0.78	0.002
3A-3	616	0.16	0.87	0.22	0.68	0.005	0.12	0.84	0.20	0.74	0.004
3A-4	609	0.10	0.84	0.17	0.75	0.003	0.01	0.85	0.13	0.84	0.002
3A-9*	662	-0.02	0.83	0.16	0.82	0.000	-0.07	0.82	0.16	0.78	0.000
3A-NE	762	0.07	0.68	0.25	0.59	0.006	0.07	0.52	0.28	0.48	0.007
3A-NW	779	-0.07	0.63	0.25	0.38	0.005	-0.02	0.64	0.20	0.62	0.003
3A-S	615	0.01	0.85	0.20	0.44	0.004	0.03	0.87	0.11	0.81	0.001
3A-SW	643	0.11	0.82	0.17	0.49	0.003	0.02	0.86	0.12	0.74	0.002
3B-2*	50	-0.23	0.63	0.35	0.18	0.002	-0.07	0.66	0.14	-0.09	0.000
3B-SE	492	0.03	0.56	0.31	0.46	0.022	0.11	0.67	0.26	0.60	0.016
ANGEL*	559	0.18	0.64	0.31	0.46	0.002	-0.07	0.68	0.20	0.31	0.001
EP12R*	256	-0.13	0.68	0.14	-3.26	0.001	-0.05	0.59	0.08	-0.01	0.000
EP9R*	235	0.03	0.69	0.11	0.63	0.000	0.11	0.56	0.14	-1.08	0.000
EPSW	389	-0.09	0.77	0.13	-0.38	0.123	-0.03	0.61	0.08	0.46	0.039
G1502	808	0.11	0.57	0.28	0.39	0.022	0.02	0.69	0.21	0.67	0.013
G3273*	555	0.15	0.67	0.26	0.39	0.001	0.10	0.81	0.18	0.68	0.001
G3353*	466	-0.03	0.67	0.18	0.57	0.001	0.02	0.76	0.11	0.71	0.000
G618*	794	0.10	0.60	0.18	0.02	0.000	0.05	0.72	0.13	0.52	0.000
G620	615	0.11	0.80	0.16	0.57	0.006	0.07	0.81	0.12	0.73	0.004
HOLEY1	235	0.23	0.64	0.26	-0.53	0.005	0.06	0.65	0.14	0.57	0.001
HOLEY2	233	0.19	0.67	0.23	-0.11	0.004	0.00	0.39	0.19	0.31	0.003
HOLEYG	240	0.24	0.55	0.29	-1.48	0.005	0.09	0.53	0.16	0.20	0.002
MonRd*	140	0.43	0.65	0.49	-11.90	0.004	0.21	0.65	0.30	0.29	0.001
NESRS1	726	0.02	0.48	0.15	0.43	0.005	-0.03	0.51	0.14	0.47	0.005
NESRS2	714	0.09	0.63	0.18	0.39	0.008	0.02	0.65	0.15	0.61	0.005
NESRS3	536	0.07	0.60	0.26	0.29	0.017	-0.03	0.71	0.17	0.70	0.008
NP202*	751	-0.06	0.81	0.12	0.71	0.000	-0.07	0.84	0.12	0.70	0.000
NP203*	665	0.00	0.79	0.10	0.77	0.000	-0.04	0.84	0.09	0.78	0.000
NP205	781	0.05	0.67	0.19	0.64	0.010	0.03	0.72	0.17	0.70	0.008
NP206	714	0.14	0.57	0.29	0.45	0.028	0.08	0.75	0.21	0.71	0.016
NP207	100	-0.05	0.79	0.14	-0.35	0.122	0.04	0.61	0.17	-1.06	0.126
NP33	771	0.04	0.55	0.16	0.42	0.006	-0.05	0.73	0.12	0.67	0.004
NP34	762	0.10	0.70	0.23	0.29	0.064	-0.06	0.67	0.18	0.56	0.062
NP35	783	0.19	0.69	0.25	-0.95	0.106	0.13	0.64	0.18	-0.04	0.070
NP36	785	0.04	0.47	0.18	0.38	0.020	0.01	0.56	0.15	0.55	0.015
NP38	770	0.08	0.70	0.19	-0.03	0.109	0.04	0.74	0.13	0.56	0.059
NP44	743	0.34	0.68	0.42	0.07	0.101	0.04	0.80	0.20	0.79	0.035
NP46*	699	-0.06	0.63	0.17	0.59	0.001	-0.10	0.71	0.17	0.55	0.001
NP62*	624	0.12	0.72	0.20	0.32	0.001	-0.10	0.76	0.18	0.36	0.001
NP67	734	0.03	0.71	0.13	0.63	0.037	-0.07	0.80	0.14	0.55	0.060
NP72*	669	0.37	0.66	0.44	-0.67	0.004	0.03	0.79	0.22	0.78	0.001
ROTTs	395	0.03	0.62	0.15	0.25	0.002	-0.13	0.64	0.20	-0.24	0.003
RUTZKE	473	-0.12	0.48	0.35	-1.36	0.086	0.08	0.64	0.17	0.47	0.016
SHARK	633	-0.02	0.68	0.15	0.64	0.004	-0.02	0.60	0.16	0.58	0.005
TAMI40*	790	0.11	0.55	0.29	-14.55	0.001	-0.10	0.82	0.17	0.60	0.000
THSO	428	0.05	0.71	0.18	0.59	0.036	-0.14	0.85	0.18	0.60	0.056
Mean	546	0.06	0.68	0.23	-0.48	0.018	0.01	0.68	0.18	0.35	0.012
Max	816	0.52	0.87	0.52	0.82	0.123	0.49	0.87	0.50	0.84	0.126
Min	50	-0.24	0.47	0.10	-14.55	0.000	-0.22	0.23	0.08	-3.88	0.000

Table 2. Unweighted TP concentration summary statistics.

NAME	OBS	Bias	R ²	RMSE	EFF	U ²	gMean
	N						DIF
		mg/l		mg/l			mg/l
C123SR84	54	-0.007	0.06	0.040	-1.25	0.62	-0.011
CA311	31	0.003	0.04	0.004	-2.50	0.24	0.003
CA315	29	-0.001	0.09	0.004	-0.71	0.54	-0.001
CA32	26	0.000	0.26	0.005	-1.17	0.44	0.000
CA33	27	0.024	0.06	0.031	-9.50	0.57	0.023
CA34	29	0.001	0.01	0.008	-0.12	0.59	0.002
CA35	22	0.021	0.04	0.028	-4.77	0.59	0.022
CA36	30	0.022	0.54	0.027	-0.40	0.24	0.026
CA38	30	0.002	0.11	0.013	-0.22	1.24	0.004
COOPERTN	115	0.000	0.02	0.006	-1.32	0.28	-0.001
EP	52	0.001	0.10	0.007	0.03	0.49	0.001
L40-1	99	-0.009	0.13	0.058	0.00	0.68	-0.009
L40-2	101	-0.027	0.17	0.057	-0.23	0.66	-0.031
L7	60	0.012	0.00	0.097	-0.57	0.92	0.006
LOX10	17	0.002	0.08	0.011	-0.28	0.58	0.004
LOX11	22	-0.004	0.03	0.006	-1.17	1.47	-0.003
LOX12	23	0.013	0.12	0.014	-19.95	0.48	0.013
LOX13	25	-0.003	0.00	0.006	-0.54	0.97	-0.002
LOX14	25	0.014	0.08	0.016	-18.06	0.46	0.014
LOX15	25	0.018	0.00	0.023	-18.50	0.58	0.017
LOX16	23	0.016	0.00	0.019	-10.14	0.52	0.016
LOX3	17	0.001	0.11	0.015	-0.33	0.96	0.004
LOX4	16	0.024	0.01	0.027	-105.73	0.62	0.022
LOX5	19	-0.002	0.19	0.006	-0.71	0.61	-0.001
LOX6	23	0.005	0.03	0.009	-2.69	0.39	0.005
LOX7	23	0.002	0.09	0.005	-2.41	0.21	0.002
LOX8	23	-0.002	0.06	0.005	-0.73	0.57	-0.001
LOX9	19	-0.003	0.04	0.008	-0.15	1.15	-0.001
NE1	75	-0.006	0.05	0.009	-0.26	2.74	-0.004
P33	88	-0.002	0.02	0.008	-0.48	1.02	-0.001
P34	66	0.008	0.14	0.010	-2.60	0.37	0.009
P35	82	-0.008	0.01	0.023	-0.14	7.20	-0.004
P36	85	-0.027	0.21	0.074	0.22	120.37	-0.007
P37	44	0.011	0.09	0.014	-0.18	0.50	0.012
S10A	94	0.001	0.15	0.049	-0.81	0.41	-0.005
S10C	100	-0.034	0.29	0.066	0.02	0.91	-0.026
S10D	172	-0.062	0.47	0.087	-0.12	1.76	-0.053
S10E	46	-0.061	0.05	0.085	-1.14	2.74	-0.058
S11A	164	0.007	0.05	0.040	-1.26	0.68	0.007
S11B	178	-0.011	0.24	0.047	0.14	0.97	-0.007
S11C	222	-0.023	0.30	0.046	0.04	1.05	-0.020
S12A	381	-0.004	0.07	0.029	0.06	3.81	0.000
S12B	388	-0.001	0.06	0.022	0.03	2.13	0.000
S12C	399	-0.002	0.04	0.018	-0.05	1.50	-0.001
S12D	395	-0.002	0.07	0.015	-0.10	0.95	-0.002
S144	166	-0.005	0.01	0.025	-0.44	2.05	-0.002
S145	196	-0.002	0.00	0.020	-0.44	1.43	-0.001
S146	167	-0.003	0.01	0.021	-0.55	1.44	-0.001
S151	217	-0.007	0.10	0.025	-0.07	1.22	-0.007
S31	74	-0.010	0.05	0.024	-0.18	3.72	-0.008
S333	389	-0.004	0.06	0.017	-0.06	1.27	-0.003
TSB	60	0.000	0.02	0.010	0.82	0.49	0.006
Mean	101	-0.002	0.10	0.026	-4.07	3.43	-0.001
Max	399	0.024	0.54	0.097	0.82	120.37	0.026
Min	16	-0.062	0.00	0.004	-105.73	0.21	-0.058

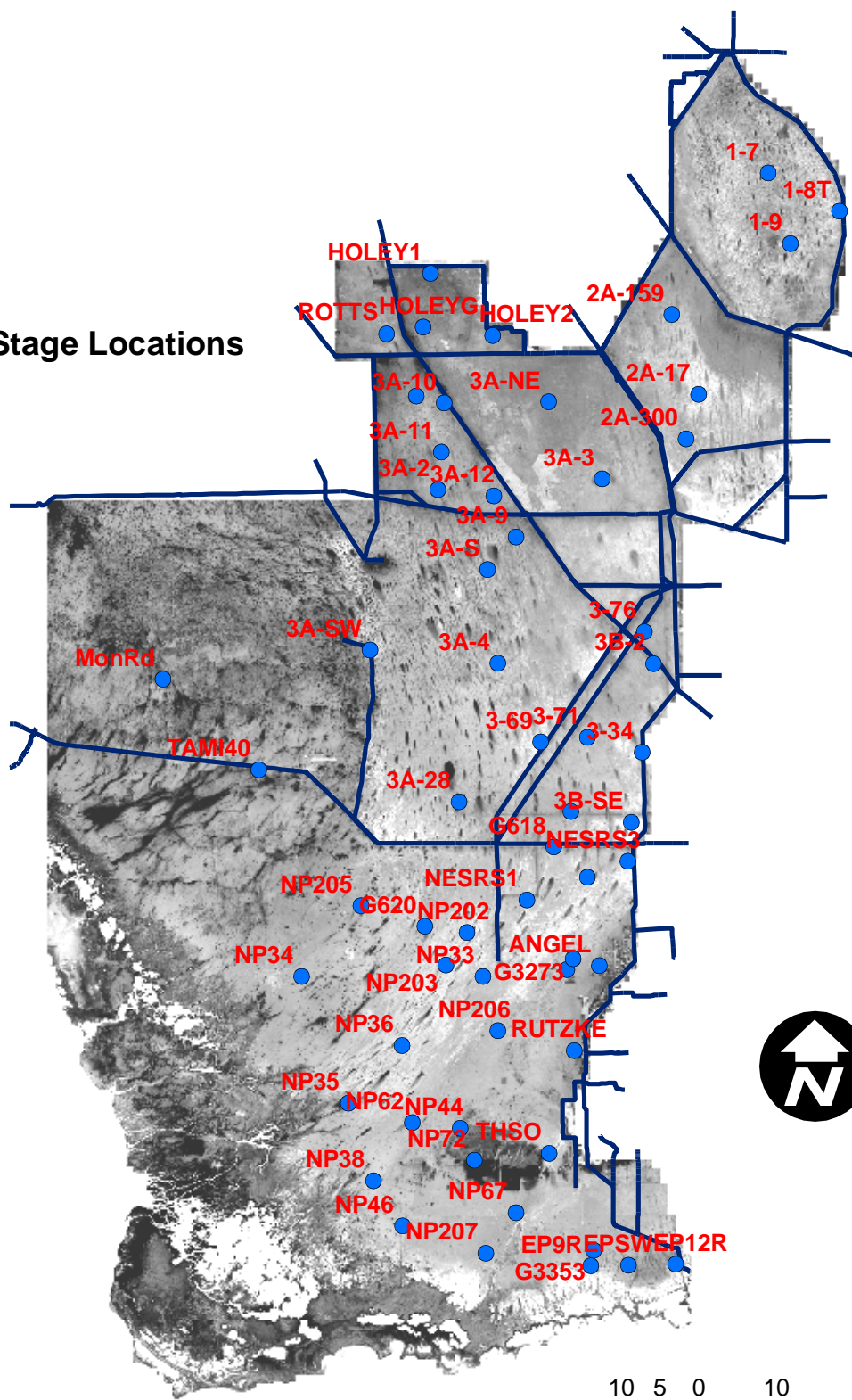
Table 3. Seasonal mean TP concentration summary statistics.

Site	n	Seasonal Mean					Weighted Seasonal Mean				
		R ²	Bias	RMSE	EFF	U ²	R ²	Bias	RMSE	EFF	U ²
			mg/l	mg/l				mg/l	mg/l		
C123SR84	11	0.05	0.004	0.038	-2.49	0.65	0.05	0.004	0.034	-1.81	0.53
CA311	2		-0.002	0.009				-0.003	0.001		
CA315	2		0.002	0.007				0.001	0.000		
CA32	2		-0.001	0.008				-0.001	0.002		
CA33	3		-0.017	0.011				-0.019	0.009		
CA34	2		0.000	0.003				0.000	0.002		
CA35	2		-0.020	0.012				-0.022	0.006		
CA36	2		-0.023	0.022				-0.022	0.019		
CA38	2		-0.002	0.011				-0.004	0.007		
COOPERTN	6	0.60	0.001	0.006	-15.00	0.26	0.60	0.001	0.001	0.12	0.01
EP	12	0.26	-0.002	0.010	-11.87	1.20	0.11	-0.001	0.009	-4.23	0.90
L40-1	14	0.02	-0.001	0.033	-0.24	0.23	0.02	-0.001	0.024	0.34	0.13
L40-2	14	0.09	0.017	0.049	-2.56	0.34	0.09	0.017	0.036	-0.93	0.18
L7	10	0.14	-0.023	0.047	-1.49	0.34	0.14	-0.023	0.027	0.21	0.11
LOX10	3		-0.001	0.008				-0.002	0.004		
LOX11	3		0.004	0.004				0.003	0.003		
LOX12	3		-0.012	0.002				-0.012	0.000		
LOX13	3		0.003	0.003				0.003	0.001		
LOX14	3		-0.014	0.005				-0.014	0.002		
LOX15	3		-0.018	0.007				-0.018	0.005		
LOX16	3		-0.016	0.007				-0.016	0.004		
LOX3	3		0.000	0.011				-0.002	0.010		
LOX4	2		-0.022	0.004				-0.022	0.001		
LOX5	3		0.004	0.006				0.004	0.004		
LOX6	3		-0.004	0.006				-0.005	0.002		
LOX7	3		-0.001	0.006				-0.001	0.001		
LOX8	3		0.002	0.005				0.002	0.002		
LOX9	3		0.005	0.007				0.003	0.004		
NE1	14	0.33	0.006	0.009	-1.09	0.39	0.32	0.005	0.007	-0.82	0.31
P33	15	0.09	0.002	0.009	-5.24	0.67	0.06	0.001	0.007	-4.33	0.57
P34	13	0.43	-0.008	0.008	-13.10	0.92	0.17	-0.009	0.008	-8.27	1.00
P35	15	0.00	0.008	0.016	-0.49	0.57	0.04	0.006	0.011	-0.76	0.52
P36	15	0.05	0.028	0.030	0.04	0.38	0.45	0.014	0.020	-0.17	0.47
P37	14	0.04	-0.012	0.009	-3.60	0.94	0.22	-0.012	0.008	-4.58	1.00
S10A	19	0.24	0.002	0.033	-0.49	0.21	0.23	-0.016	0.021	0.67	0.08
S10C	19	0.27	0.037	0.064	-0.85	0.33	0.44	0.035	0.066	-0.05	0.26
S10D	22	0.31	0.060	0.072	-1.25	0.30	0.67	0.076	0.077	-0.02	0.18
S10E	8	0.00	0.050	0.101	-7.31	0.66	0.07	0.069	0.099	-4.15	0.58
S11A	25	0.40	-0.013	0.010	0.78	0.09	0.68	-0.025	0.003	0.98	0.01
S11B	24	0.52	0.008	0.014	0.89	0.05	0.63	0.000	0.015	0.87	0.06
S11C	25	0.57	0.018	0.034	-0.80	0.31	0.65	0.017	0.020	0.75	0.08
S12A	25	0.11	0.006	0.009	0.80	0.11	0.02	-0.002	0.004	0.38	0.11
S12B	25	0.10	0.005	0.008	0.83	0.09	0.00	-0.003	0.004	0.34	0.11
S12C	28	0.16	0.004	0.005	0.72	0.07	0.01	0.002	0.005	0.89	0.05
S12D	28	0.10	0.004	0.005	0.55	0.08	0.01	0.002	0.004	0.36	0.08
S144	21	0.01	0.003	0.009	0.24	0.20	0.02	0.004	0.011	0.22	0.27
S145	25	0.00	0.000	0.006	0.51	0.14	0.01	-0.001	0.005	0.67	0.10
S146	22	0.04	0.001	0.008	0.27	0.19	0.00	-0.001	0.007	0.60	0.15
S151	27	0.20	0.009	0.016	0.10	0.23	0.11	0.009	0.017	0.18	0.23
S31	8	0.00	0.010	0.021	-0.85	0.57	0.08	0.018	0.024	-2.31	0.55
S333	29	0.07	0.006	0.006	0.60	0.08	0.12	0.008	0.007	0.49	0.11
TSB	15	0.02	-0.002	0.017	-0.53	0.80	0.01	-0.004	0.012	-0.87	0.77
MEAN	11.65	0.17	0.002	0.017	-2.10	0.38	0.20	0.001	0.013	-0.84	0.32
MAX	29.00	0.60	0.060	0.101	0.89	1.20	0.68	0.076	0.099	0.98	1.00
MIN	2.00	0.00	-0.023	0.002	-15.00	0.05	0.00	-0.025	0.000	-8.27	0.01

9. Figures

Following this page are Figures 1 through 11.

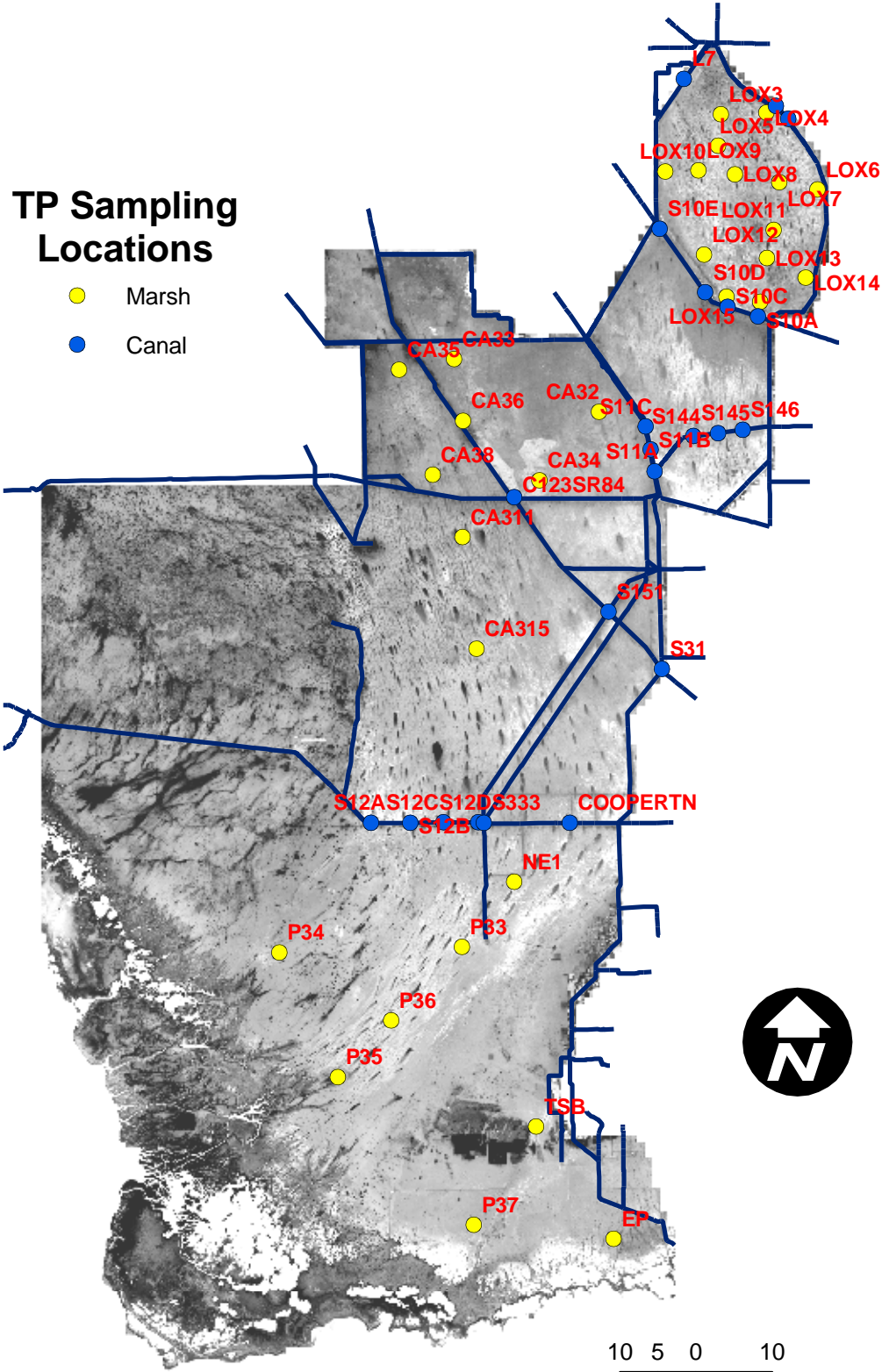
● Stage Locations



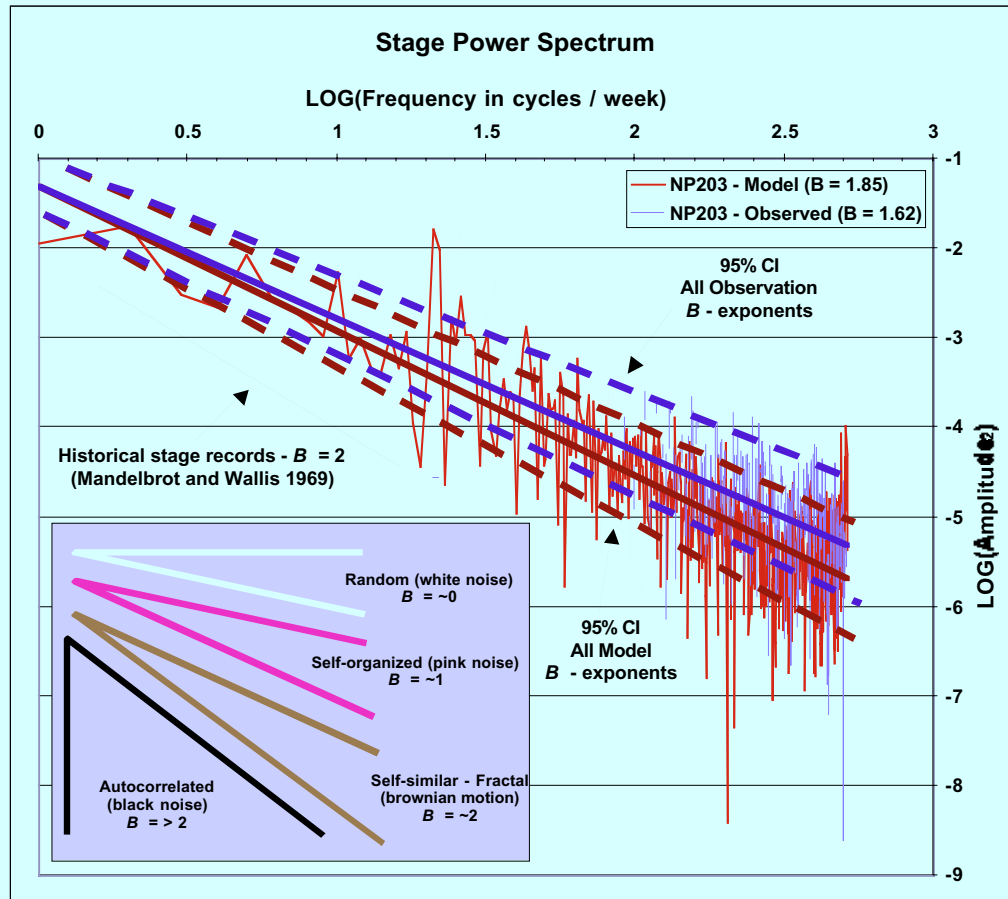
10 5 0 10
Kilometers

**TP Sampling
Locations**

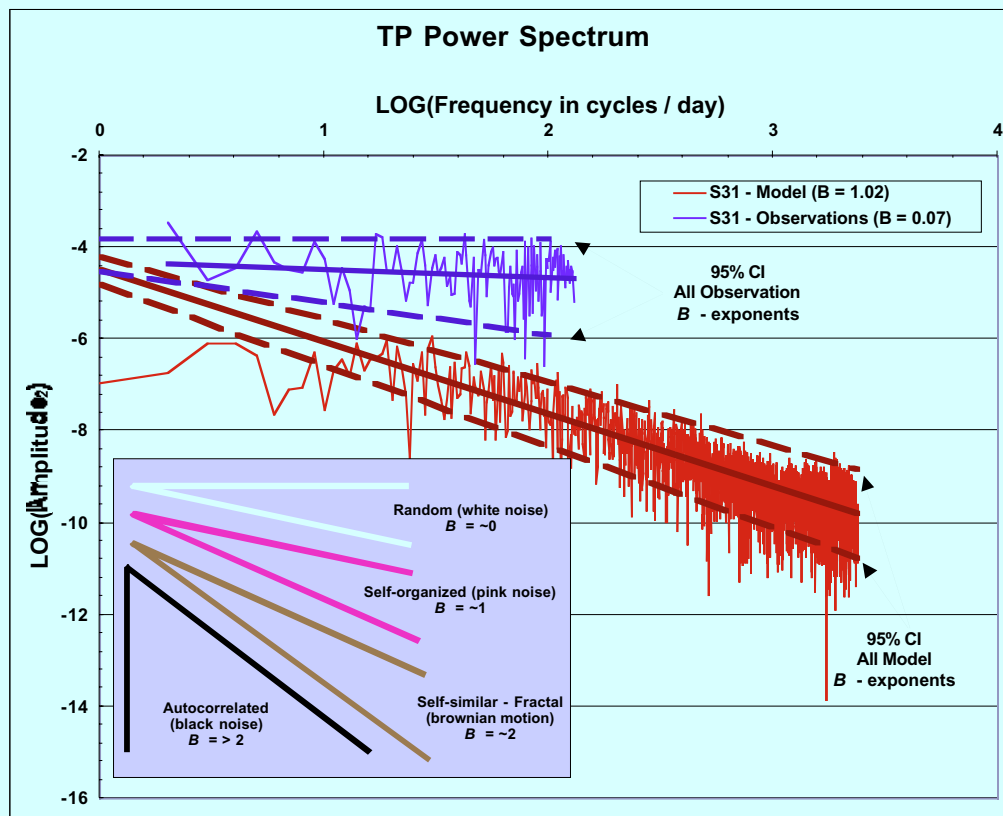
- Marsh
- Canal



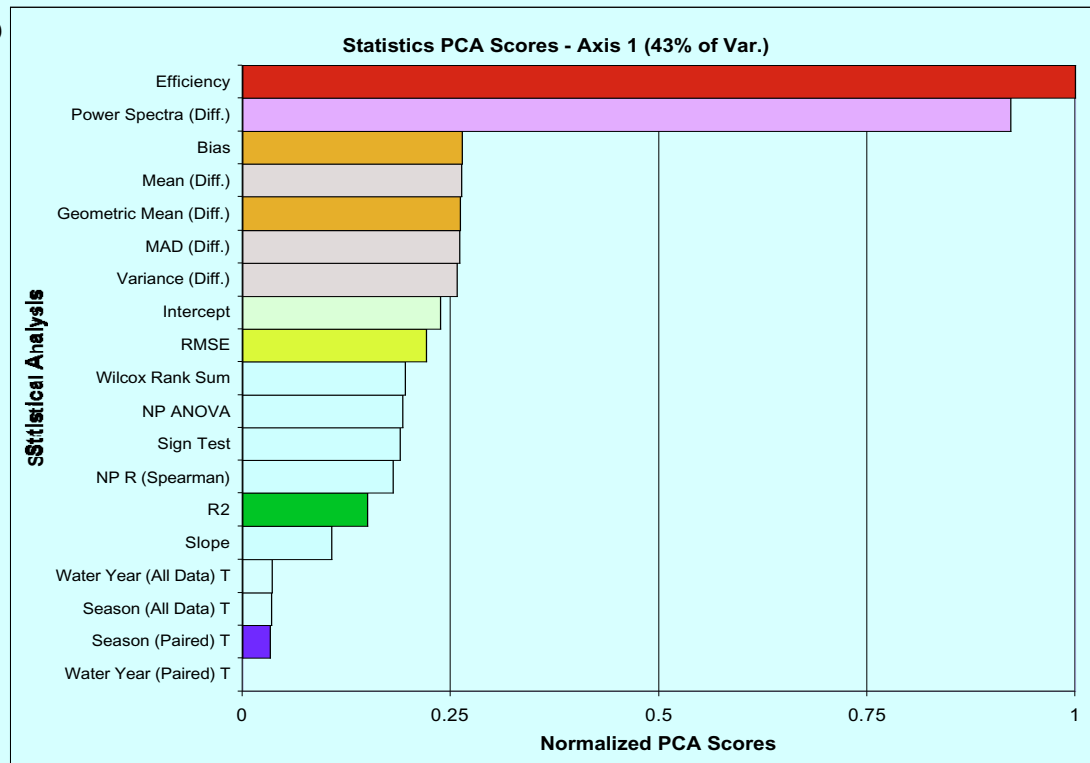
.1)



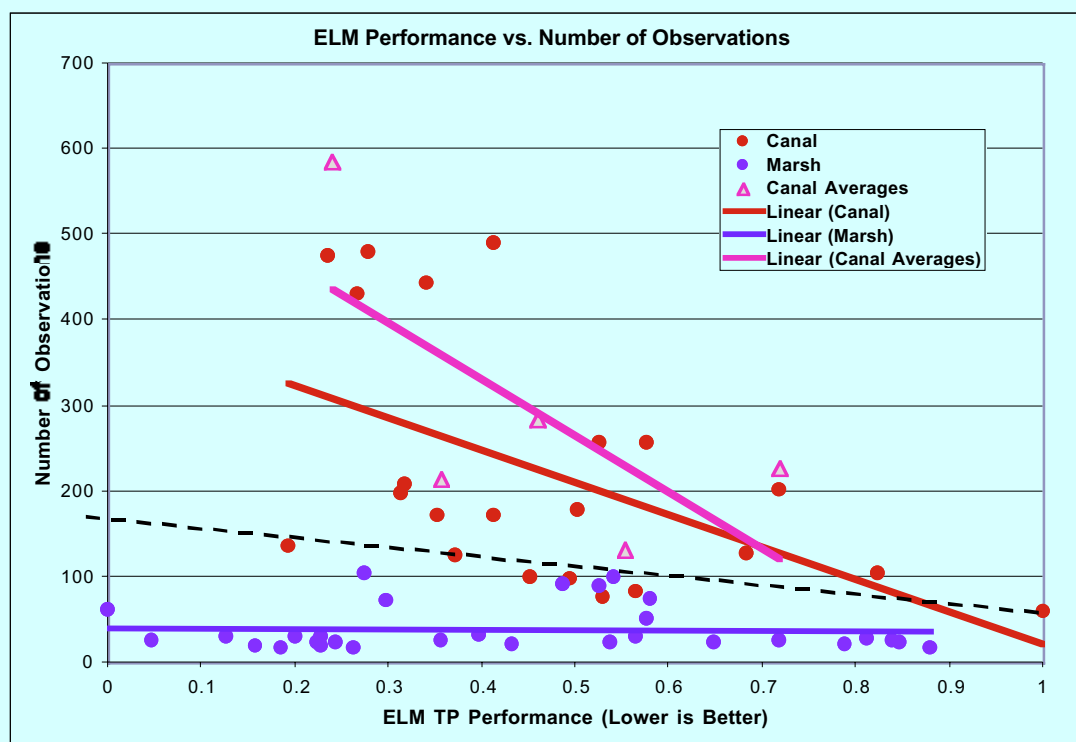
.2)

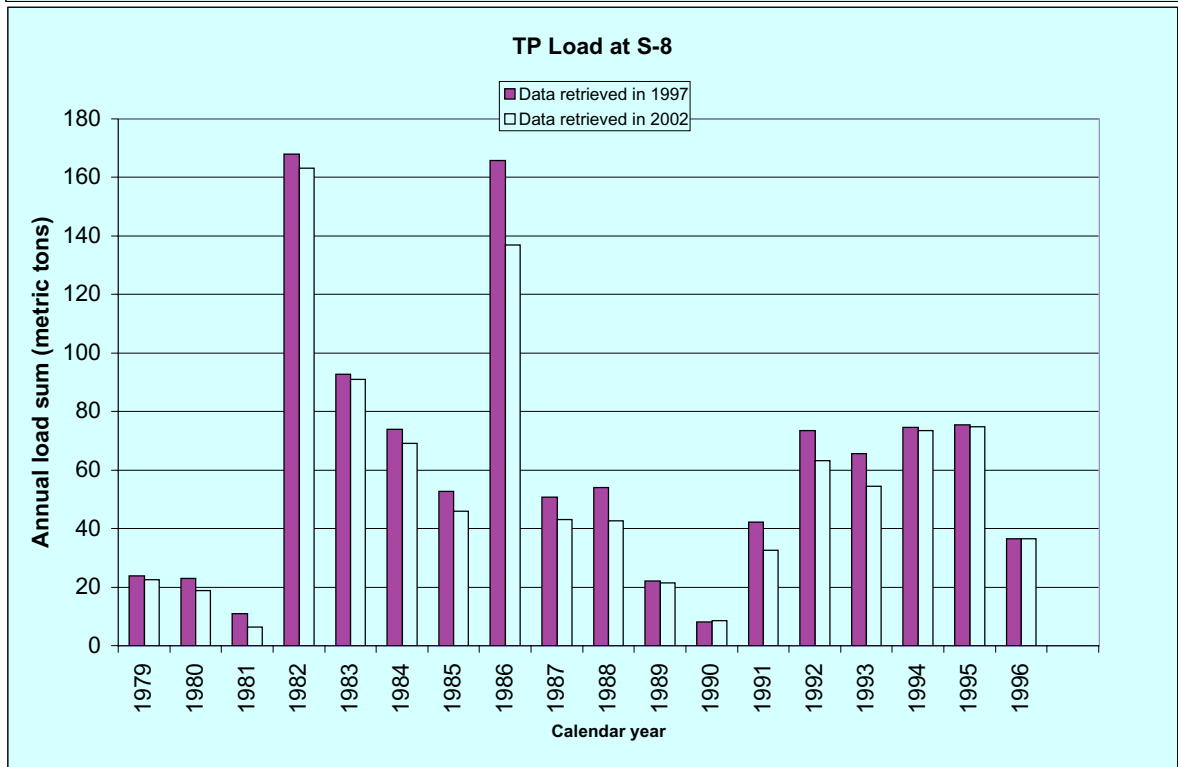
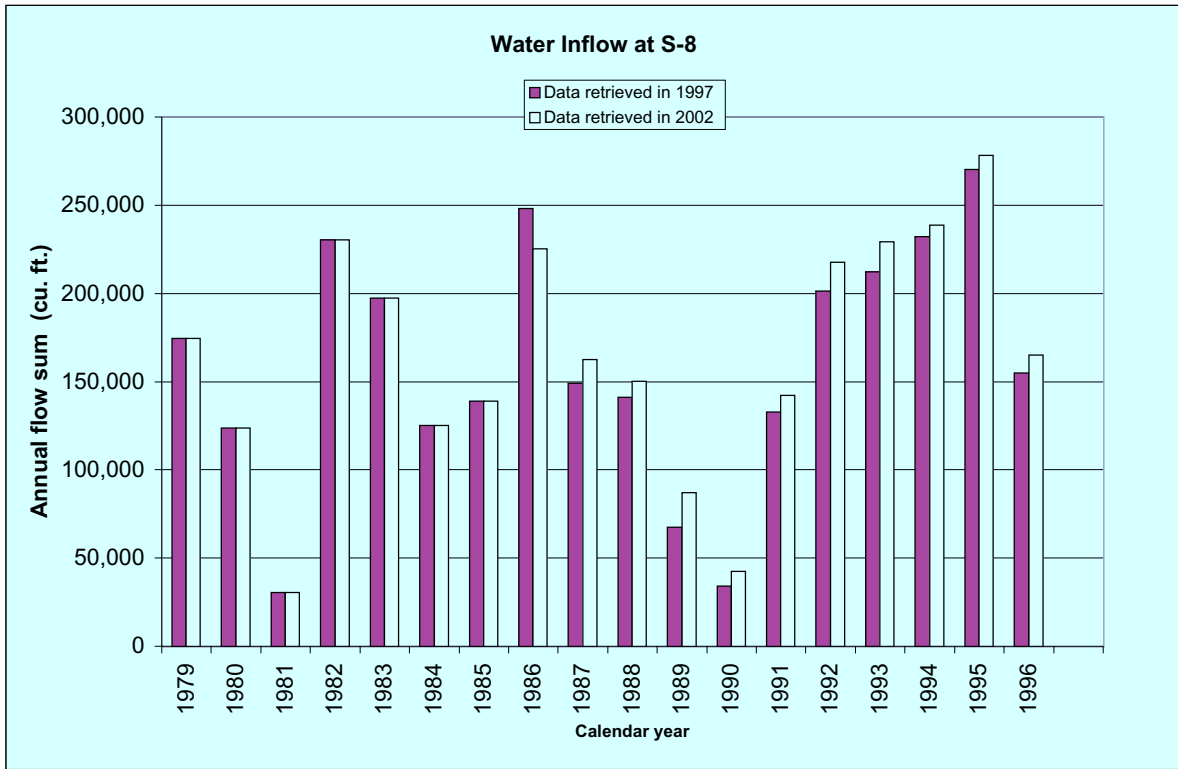


.1)

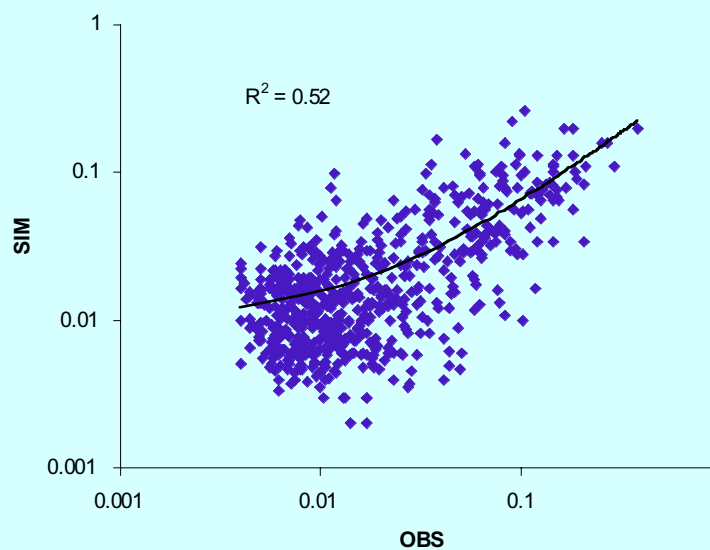


.2)

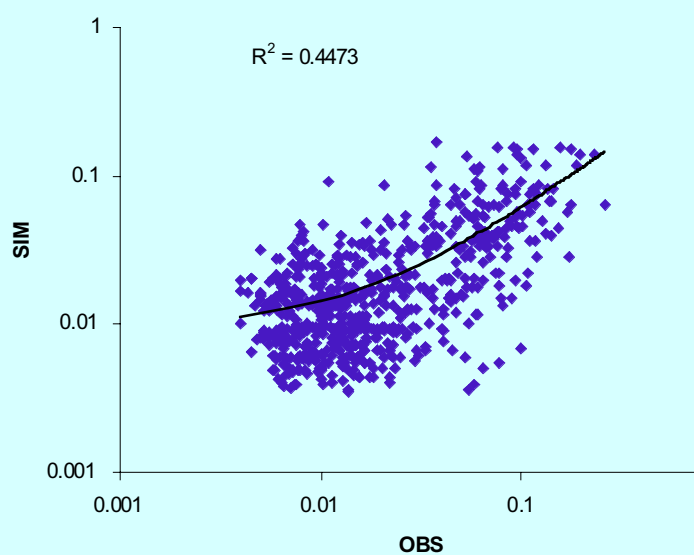




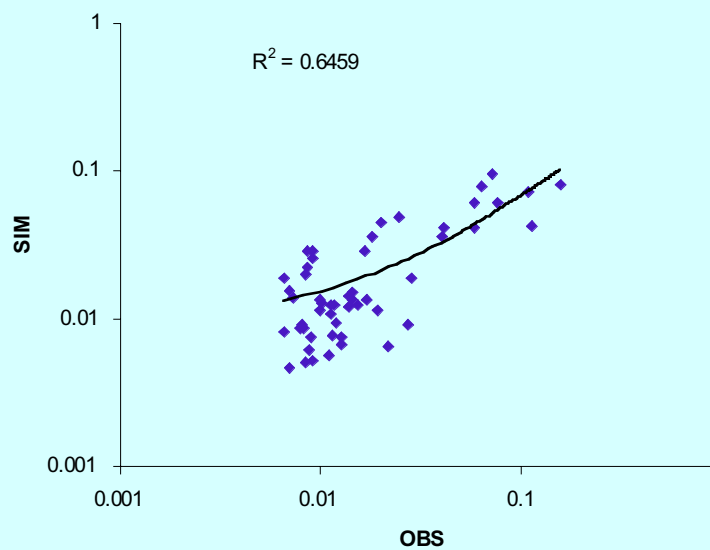
Comparison of Simulated and Observed Values:
weighted seasonal means



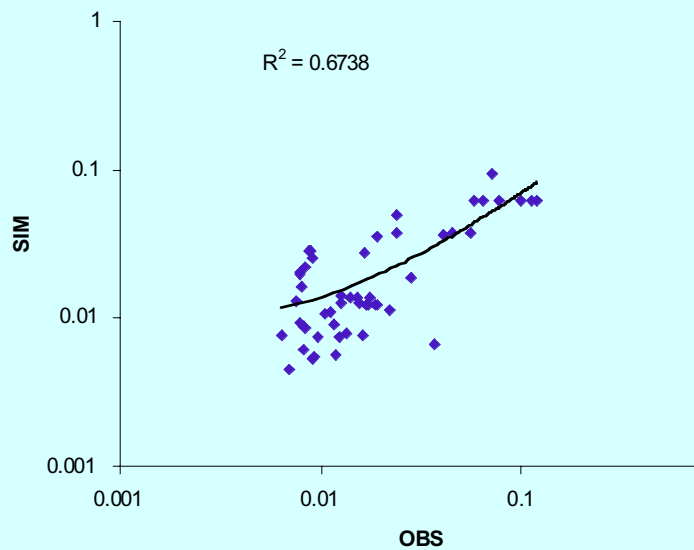
Comparison of Simulated and Observed Values:
seasonal means

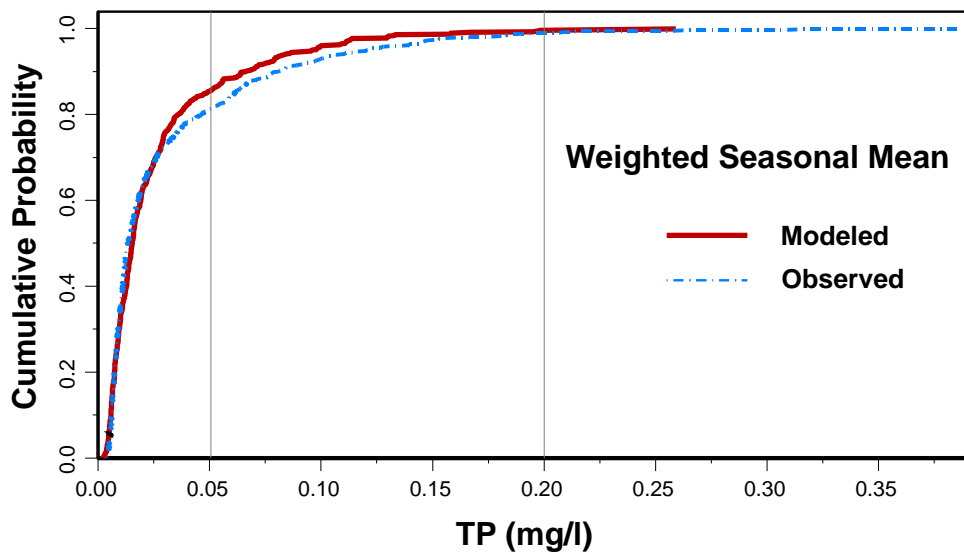


Comparison of Simulated and Observed Values:
weighted seasonal means averaged by sites

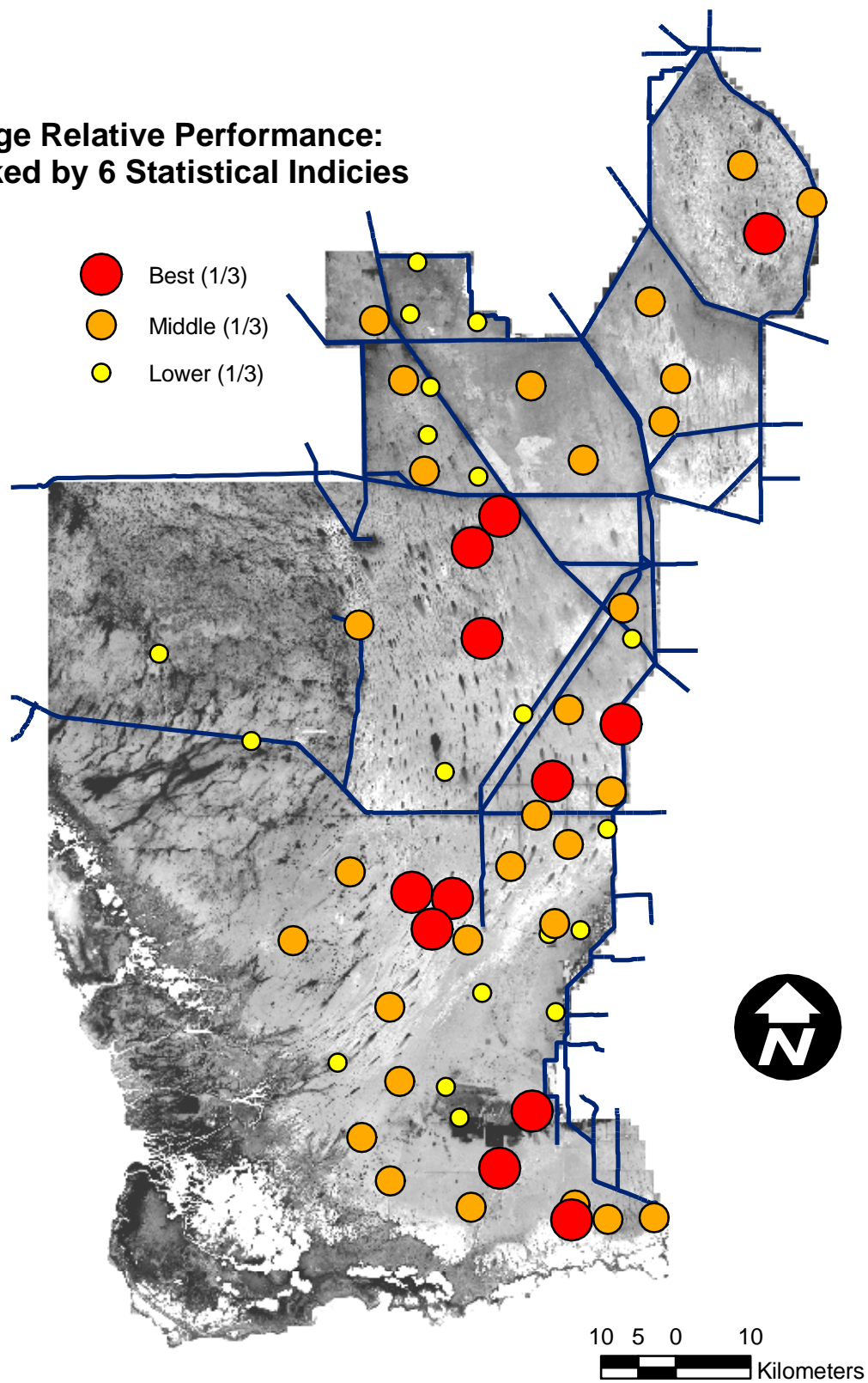


Comparison of Simulated and Observed Values:
seasonal means averaged by sites

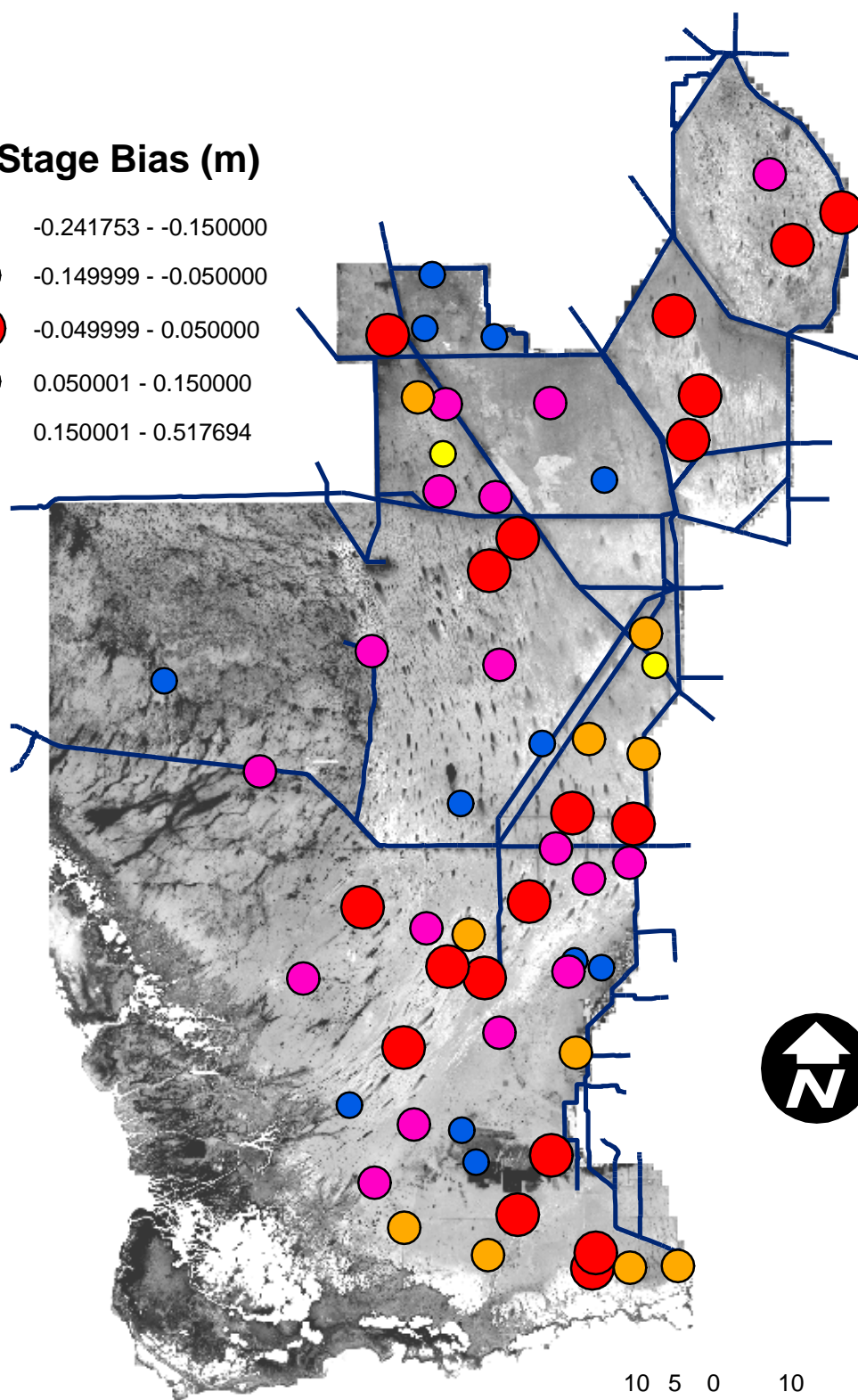
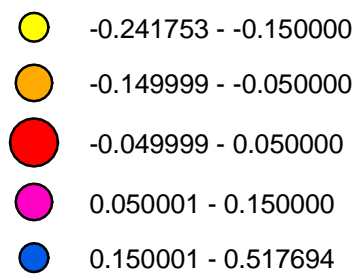




**Stage Relative Performance:
Ranked by 6 Statistical Indices**



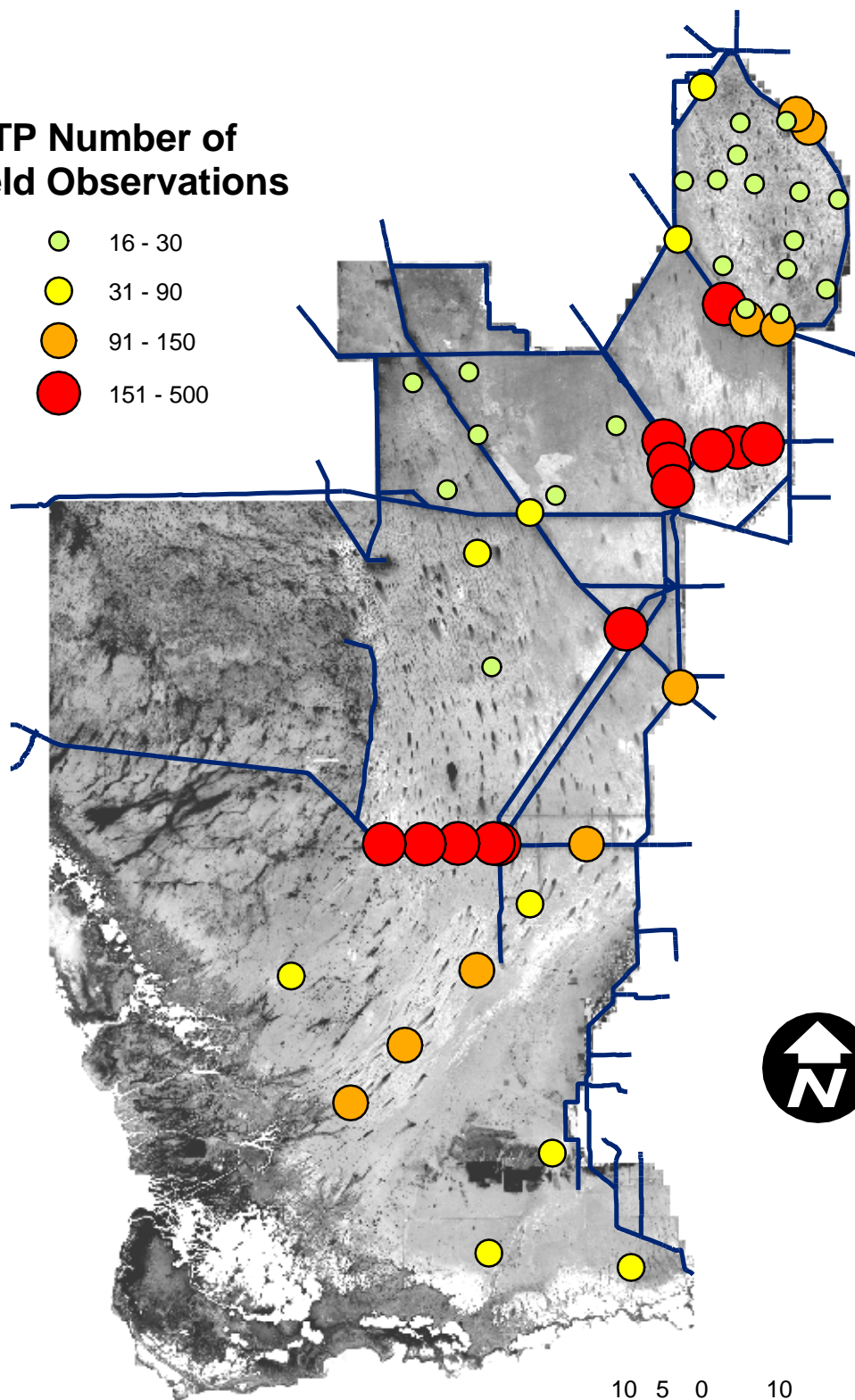
Stage Bias (m)



10 5 0 10
Kilometers

TP Number of Field Observations

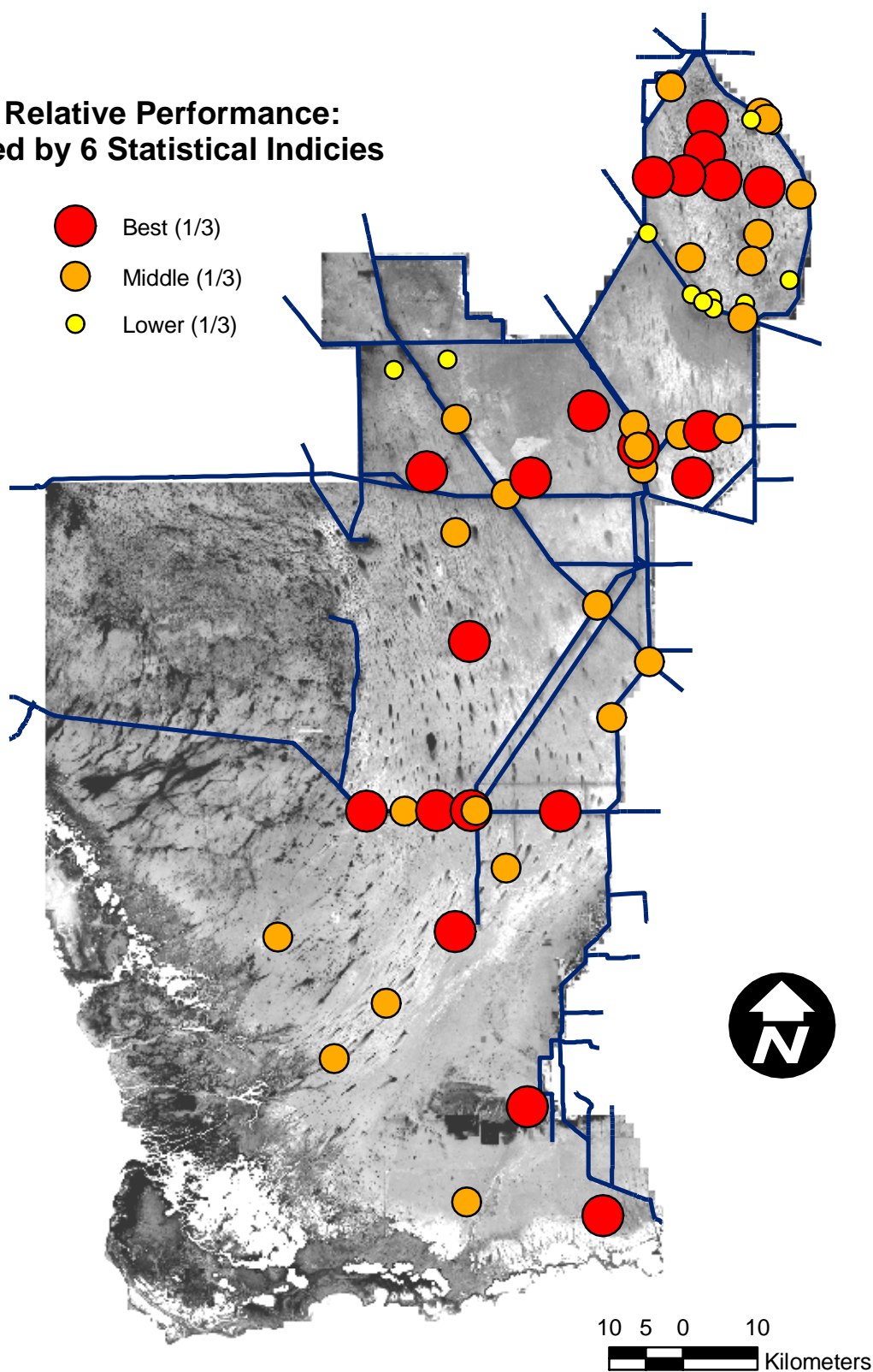
- 16 - 30
- 31 - 90
- 91 - 150
- 151 - 500



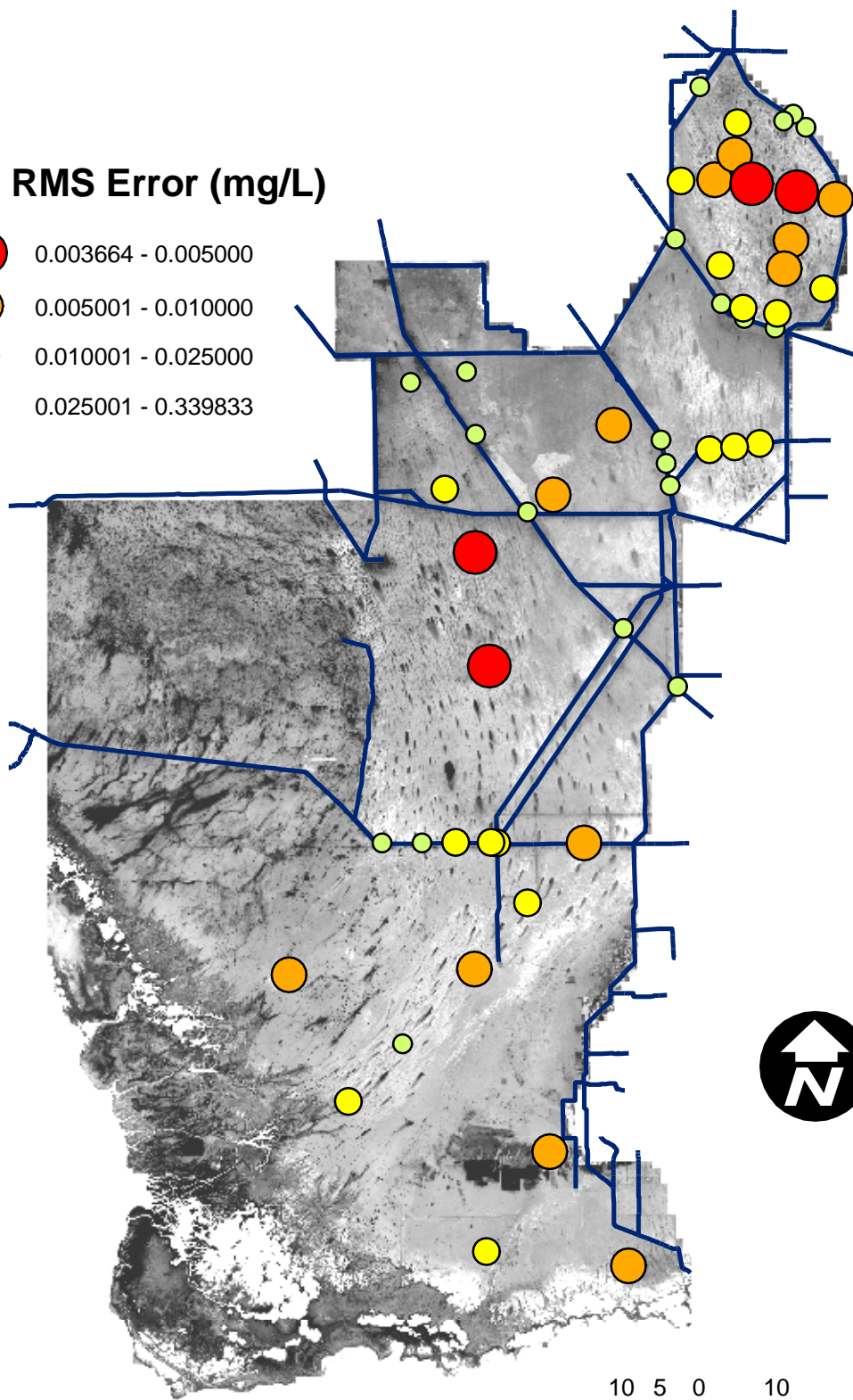
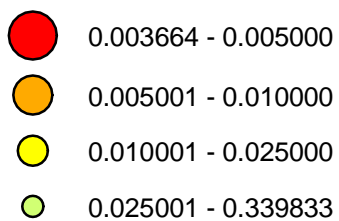
10 5 0 10
Kilometers

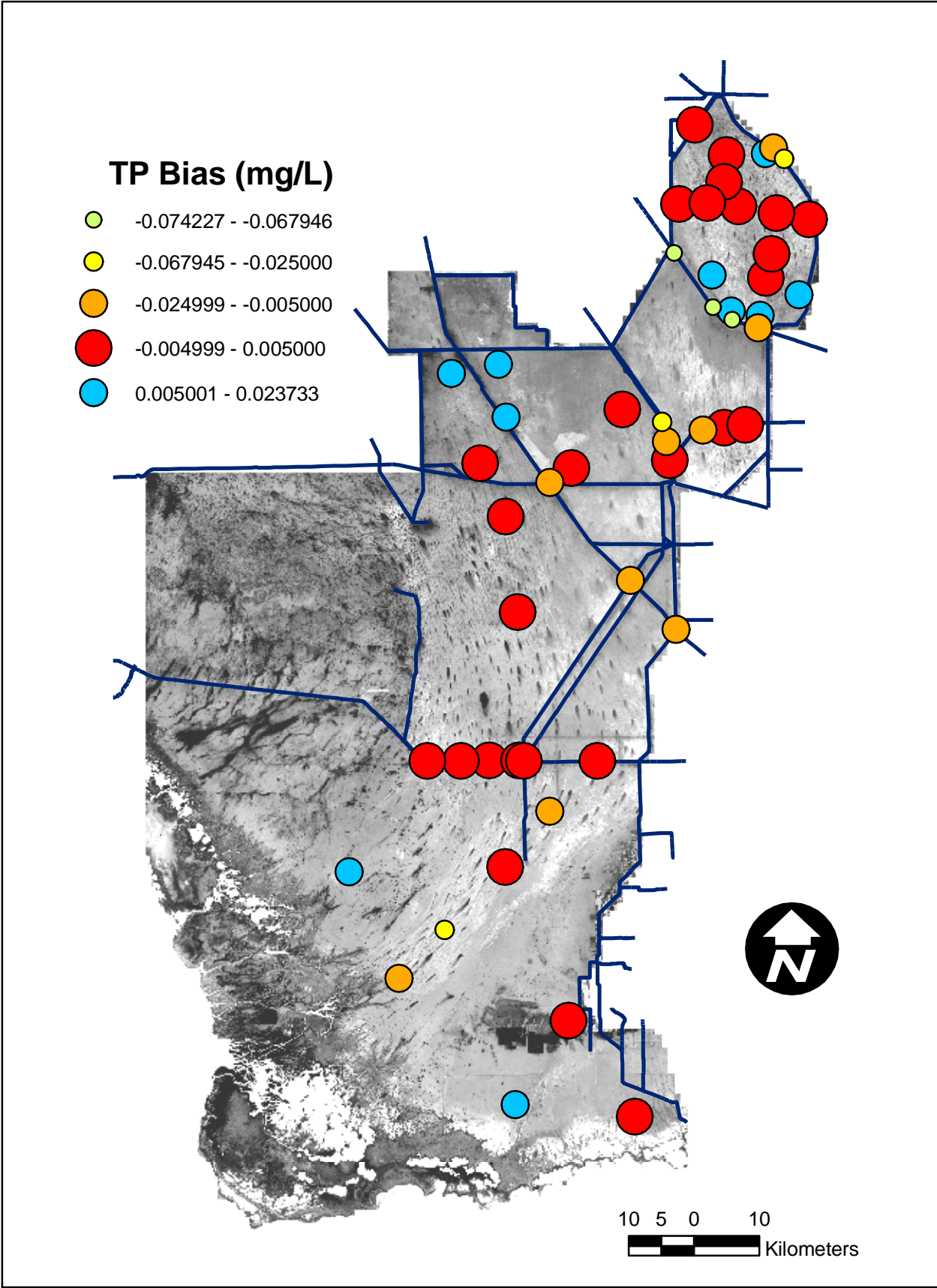
**TP Relative Performance:
Ranked by 6 Statistical Indices**

- Best (1/3)
- Middle (1/3)
- Lower (1/3)



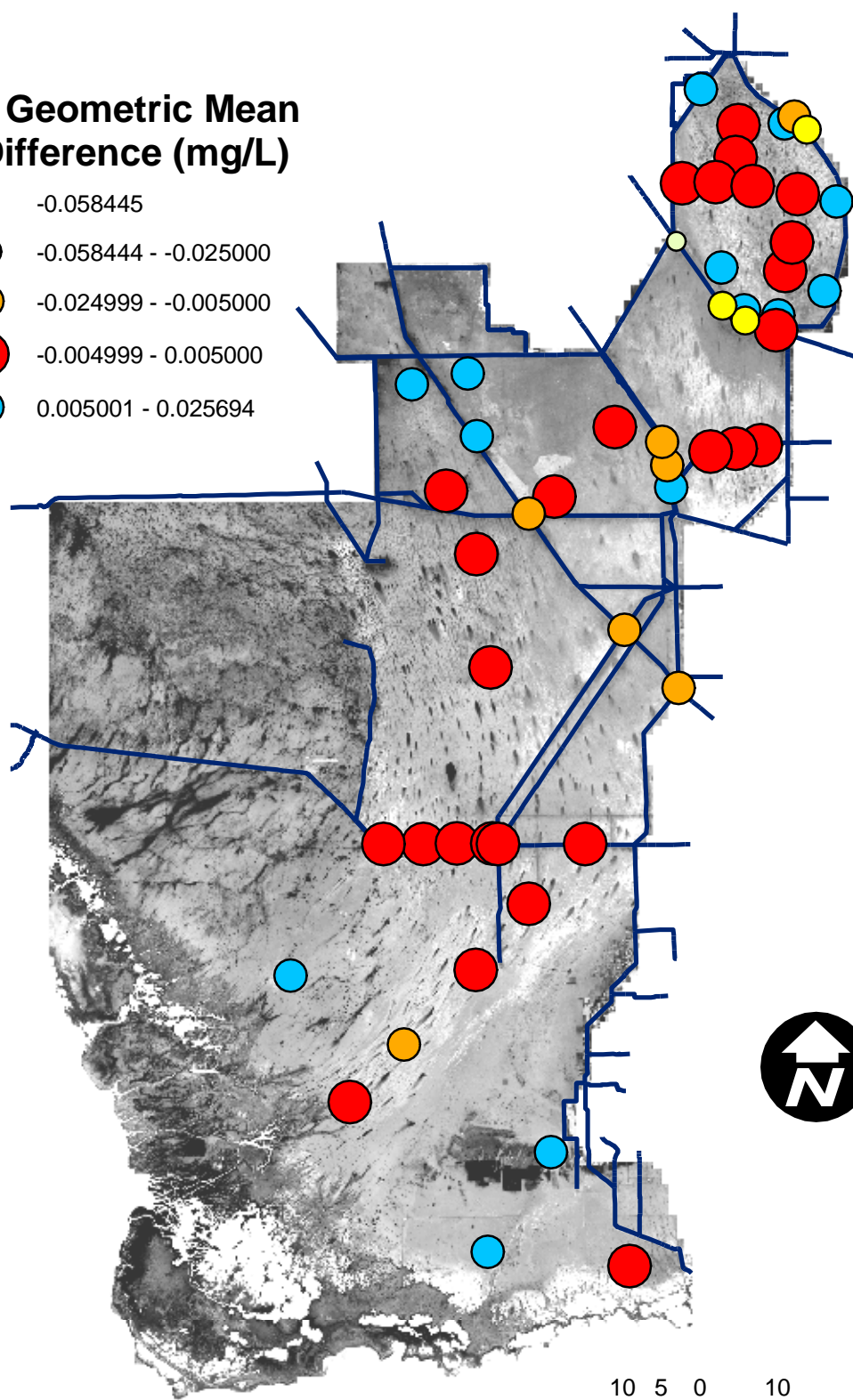
TP RMS Error (mg/L)





TP Geometric Mean Difference (mg/L)

- 0.058445
- 0.058444 - -0.025000
- 0.024999 - -0.005000
- 0.004999 - 0.005000
- 0.005001 - 0.025694



10 5 0 10
Kilometers

