

Comparison of Methods for Tracking Marsh Phosphorus
Concentrations in Loxahatchee National Wildlife Refuge
Under the Consent Decree

i.e.

4PT vs. LTL

prepared for
U.S. Department of the Interior

by

William W. Walker, Jr, Ph.D.
Environmental Engineer

TOC Meeting, December 20, 2006

? December 2006 ?



4-Part Test vs. Long-Term Levels

TOC decides "whichever is lower"

Expected Long-Term Geometric Means:

<= 10 ppb for 4PT

<= 7 ppb for LTL

Analysis:

7 is less than 10

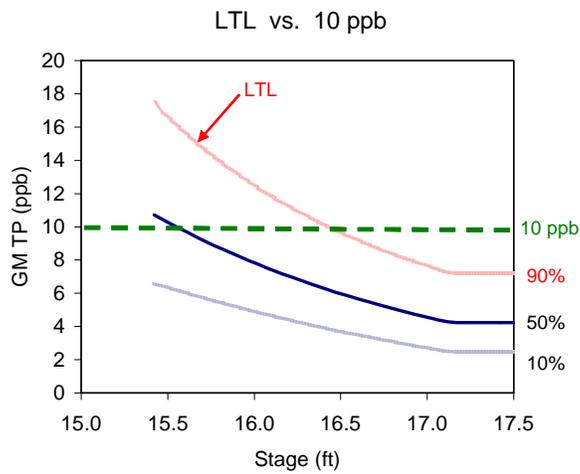
Conclusion:

LTL's are lower

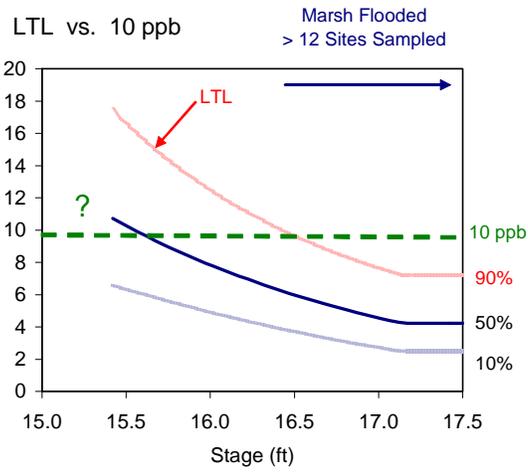
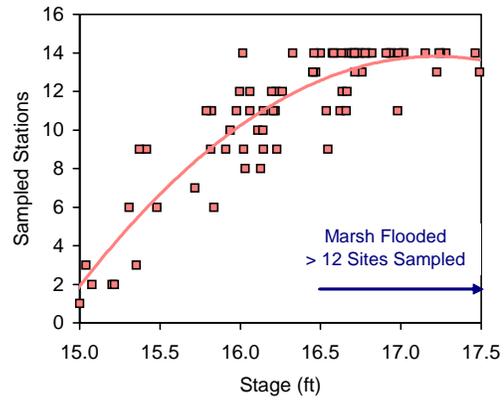
Factor	4PT	LTL
Basis	P Criterion	1978-1979 Data
Long-Term Geo Mean	<= 10 ppb	<= 7 ppb
Moderating Provision	Yes	No
Interior Sites	> 14	14
Exterior Sites	> 0	0
Spatially Averaged	Yes	Yes
Individual Sites	Yes	No
Stage-Dependent	No	Yes
Limit Percentile	95%	90%
Test Interval	Water Year	12 Consecutive Monthly Samples
Minimum Duration	1 - 5 years	~ 1 year
Sample Water Depth	> 10 cm	> 10 cm
Marsh Stage Constraint		> 15.4 ft
Sample Frequency	> 6 samples/yr	
Round-off Convention	1 ppb	0.1 ppb
Historical Failure Rate 14 Interior Stations, 1999-2006	0%	~25% excursions
LTGM at < 10% Failure Rate	< 9.5 ppb	< 7.0 ppb

4-Part Test

Test Number	Description
1	5 year geometric mean less than or equal to 10 ppb
2	Annual geometric mean averaged across all stations is less than or equal to 11 ppb
3	Annual geometric mean averaged across all stations is less than or equal to 10 ppb for three of five years
4	Annual geometric mean at individual stations is less than or equal to 15 ppb



Number of Sampled Sites vs. Stage



Statistical Power of Tests

- “Power” = probability of failing test when objective has not been achieved
- Objectives for long-term geometric mean:
 - ≤ 7 ppb at interior sites for LTL
 - ≤ 10 ppb at exterior & interior sites for 4PT
- Comparisons of factors influencing power:
 - Confidence levels
 - Round-off procedures
 - Application to historical data

Specified Confidence Levels for Deriving Compliance Limits

- LTL set at 90th percentile (or “confidence level”)
- 4PT limits set at 95th percentile
- Specifying higher percentiles causes higher limits & lower statistical power, other factors being equal

Alternative Confidence Levels for 4PT 1-Year, Single Station < 15 ppb

Assumed Confidence Level	Data Source (FDEP, 2003)	
	WCA-2A	WCA-2A + WCA-1
95% (4PT)	15.1 ppb	14.6 ppb
90% (LTL)	13.7 ppb	13.4 ppb

Equation (FDEP, 2003, Table 5-4):
 Limit = $10 \text{ Exp} (t_{n,\alpha} \sigma)$, 1-tailed t, σ = std dev of natural log yearly GM

Effect of 1 ppb Round-Off Convention on Power of 4PT

Data Rounded to 0.1 ppb	? > 10 ppb	Data Rounded to 1.0 ppb	? > 10 ppb
9.5	Pass	10	Pass
9.6	Pass	10	Pass
9.7	Pass	10	Pass
9.8	Pass	10	Pass
9.9	Pass	10	Pass
10.0	Pass	10	Pass
10.1	Fail	10	Pass
10.2	Fail	10	Pass
10.3	Fail	10	Pass
10.4	Fail	10	Pass
10.5	Fail	11	Fail
10.6	Fail	11	Fail
10.7	Fail	11	Fail
10.8	Fail	11	Fail
10.9	Fail	11	Fail
11.0	Fail	11	Fail
Failure Rate	63%		38%

Data Rounded to Nearest 0.01 ppb in Deriving & Testing the 4PT

FDEP, 2003 ECR, Table 5-5 (partial)

Table 5-5. Example application of phosphorus criterion measurement methodology using WCA-2A gradient transect data

Year	Annual Geometric Mean TP concentrations (ppb)						
	Reference Sites				1st Impacted Sites		
	E5	F5	U1	U2	U3	E4	F4
1994	8.80	9.76	7.85	7.98	6.81	19.26	16.94
1995	5.95	7.69	5.25	5.63	5.37	13.08	17.47
1996	7.75	9.95	8.70	8.23	8.44	14.56	16.88
1997	8.50	10.63	9.79	8.04	8.35	13.91	17.45
1998	7.94	10.12	7.43	9.42	9.61	14.77	13.75
1999	7.55	10.43	6.99	8.32	6.72	12.41	17.57
2000	4.58	9.45	5.18	8.64	6.15	13.53	16.74
2001	7.65	10.00	7.79	8.13	7.46	13.29	17.33
Step 1. Is annual geometric mean 10 ppb or less.					If "Yes" then in criterion achieved. If "No" then go to Step 2.		
1994	Yes	Yes	Yes	Yes	Yes	No	No
1995	Yes	Yes	Yes	Yes	Yes	No	No
1996	Yes	Yes	Yes	Yes	Yes	No	No
1997	Yes	No	Yes	Yes	Yes	No	No
1998	Yes	No	Yes	Yes	Yes	No	No
1999	Yes	No	Yes	Yes	Yes	No	No
2000	Yes	Yes	Yes	Yes	Yes	No	No
2001	Yes	Yes	Yes	Yes	Yes	No	No

Comments on Dr. Goforth's Analysis

June 2006, Amended Nov 2006

- Applied 4PT to hypothetical time series with marsh TP = LTL in each month (highly unlikely)
- Numerous assumptions
- Small differences between the tests
- Incorrect averaging method for yearly geometric means
 - Used arithmetic mean of 12 monthly values
 - FDEP protocol uses the geometric mean
- Ignores 4PT data screening protocol; excludes site-years with < 6 measurements, generally from sites with shallow depths & higher TP concentrations
- Does consider variations in spatial coverage vs. stage

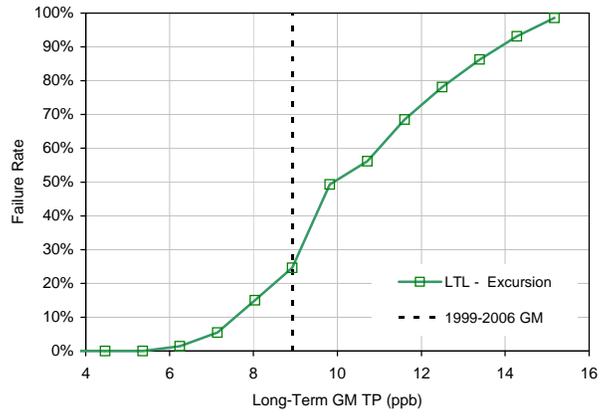
Evaluating Power of Tests Applied to Historical Interior Marsh Data

- Compile data from 14 stations, May 1995 - April 2006
- Screen & summarize data using protocols for each test
- Compute failure rates for each test, May 1999 – April 2006
- Test sensitivity to variations in LTGM by rescaling data over range of 4 to 15 ppb (vs. 8.9 ppb for 1999-2006) and re-computing failure rates
- Assumes spatial & temporal variability around the LTGM is independent of the LTGM, when expressed on a percentage basis

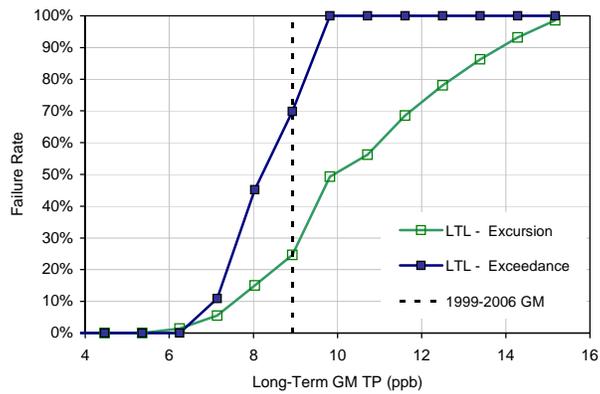
Factors Considered in This Analysis

Factor	4PT	LTL
Basis	P Criterion	1978-1979 Data
Long-Term Geo Mean	<= 10 ppb	<= 7 ppb
Moderating Provision	Yes	No
Interior Sites	> 14	14
Exterior Sites	> 0	0
Spatially Averaged	Yes	Yes
Individual Sites	Yes	No
Stage-Dependent	No	Yes
Limit Percentile	95%	90%
Test Interval	Water Year	12 Consecutive Monthly Samples
Minimum Duration	1 - 5 years	~ 1 year
Sample Water Depth	> 10 cm	> 10 cm
Marsh Stage Constraint		> 15.4 ft
Sample Frequency	> 6 samples/yr	
Round-off Convention	1 ppb	0.1 ppb
Historical Failure Rate 14 Interior Stations, 1999-2006	0%	~25% excursions
LTGM at < 10% Failure Rate	< 9.5 ppb	< 7.0 ppb

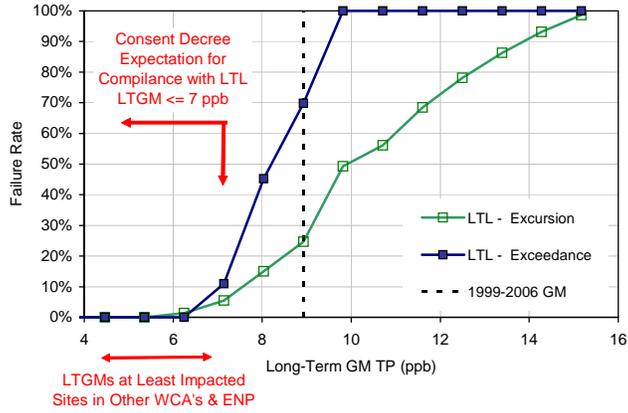
LTL Excursion Frequencies



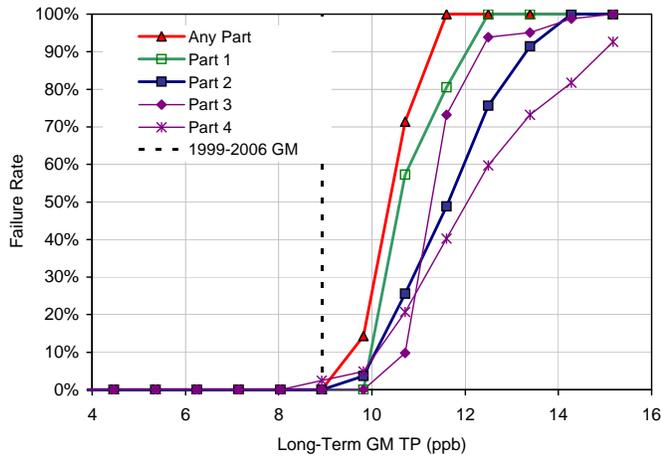
LTL Excursion & Exceedance Frequencies



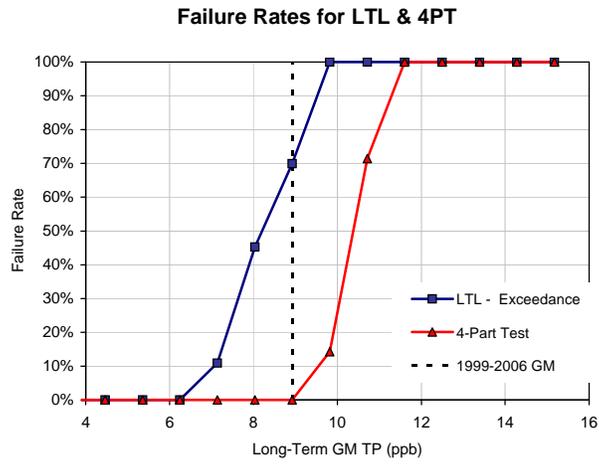
LTL Excursion & Exceedance Frequencies



Failure Rates for 4-Part Test



Test Number	Description
1	5 year geometric mean less than or equal to 10 ppb
2	Annual geometric mean averaged across all stations is less than or equal to 11 ppb
3	Annual geometric mean averaged across all stations is less than or equal to 10 ppb for three of five years
4	Annual geometric mean at individual stations is less than or equal to 15 ppb



Comparison of Test Failure Rates Applied to Interior Marsh Data

<u>LTGM (ppb)</u>	<u>Failure Rates</u>	
	<u>LTL</u>	<u>4PT</u>
< 6	0%	0%
6 - 10	0 - 100%	0 - 25%
10 - 12	100%	25 - 100%
> 12	100%	100%

Conclusions: Statistical Power of LTL vs. 4PT

- LTL has lower objective for LTGM at interior sites
< 7 ppb for LTL vs. < 10 ppb for 4PT
- LTL < 10 ppb at stage > 16.5 ft, when marsh is flooded
- LTL > 10 ppb at stage < 16.5 ft, when marsh is partially flooded and 4PT rejects some data because of low sampling frequency
- Setting limits at 95th percentile decreases power of 4PT (vs. 90th percentile for LTL)
- Rounding off to 1 ppb decreases power of 4PT (vs. 0.1 ppb for LTL)

Conclusions: Statistical Power of LTL vs. 4PT

- When applied to historical interior marsh data:
 - Tests have equal power when LTGM < 6 ppb or > 12 ppb
 - LTL more powerful than 4PT when LTGM is 6 – 12 ppb
 - LTL failure rate < 10% when LTGM < 7.0 ppb
 - 4PT failure rate is < 10% when LTGM < 9.5 ppb
- Achieving compliance with LTL will require lower TP concs in interior marsh, as compared with the 4PT
- Net effects of other differences in the tests (exterior marsh sampling, moderating provision) unknown