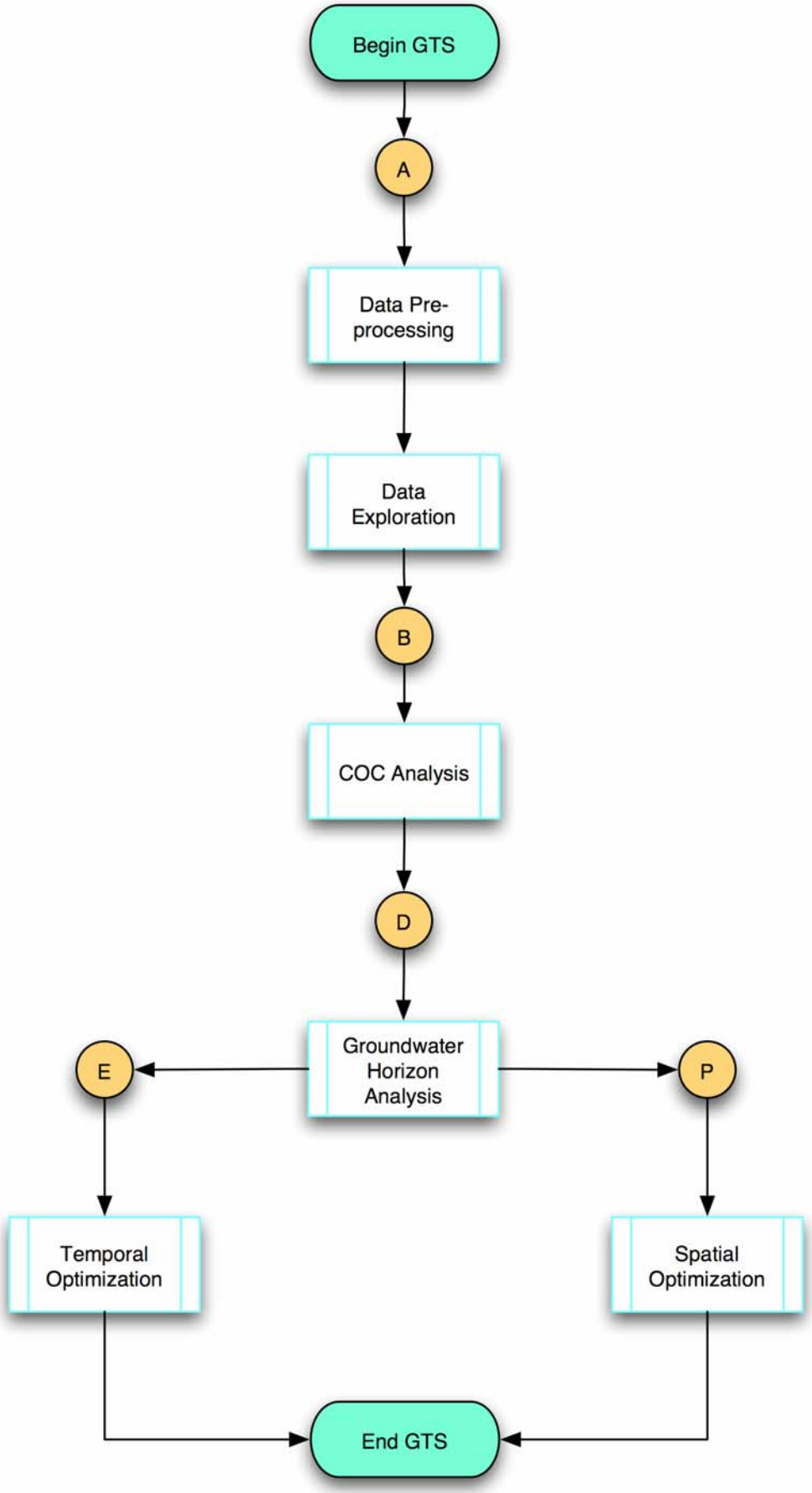
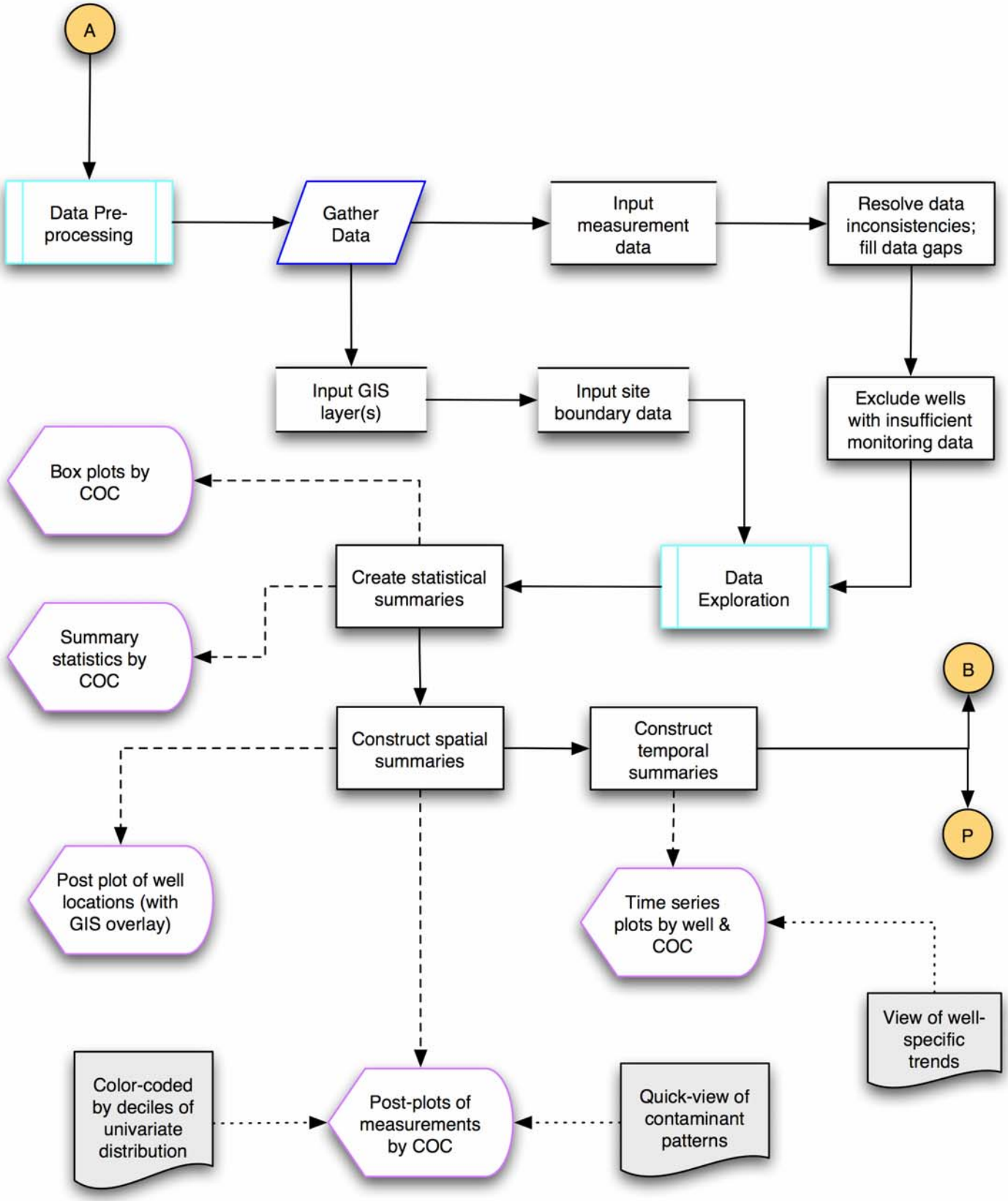
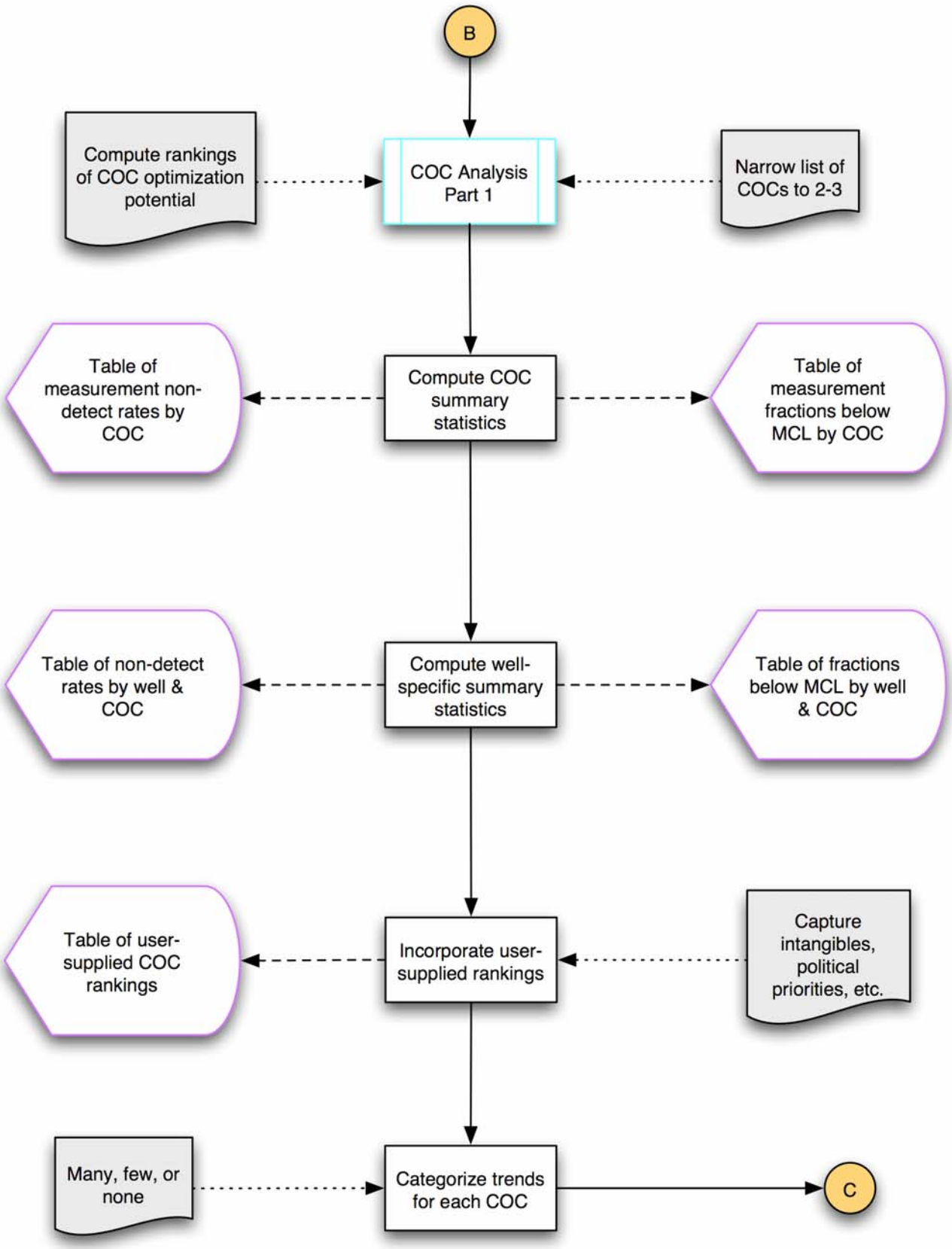
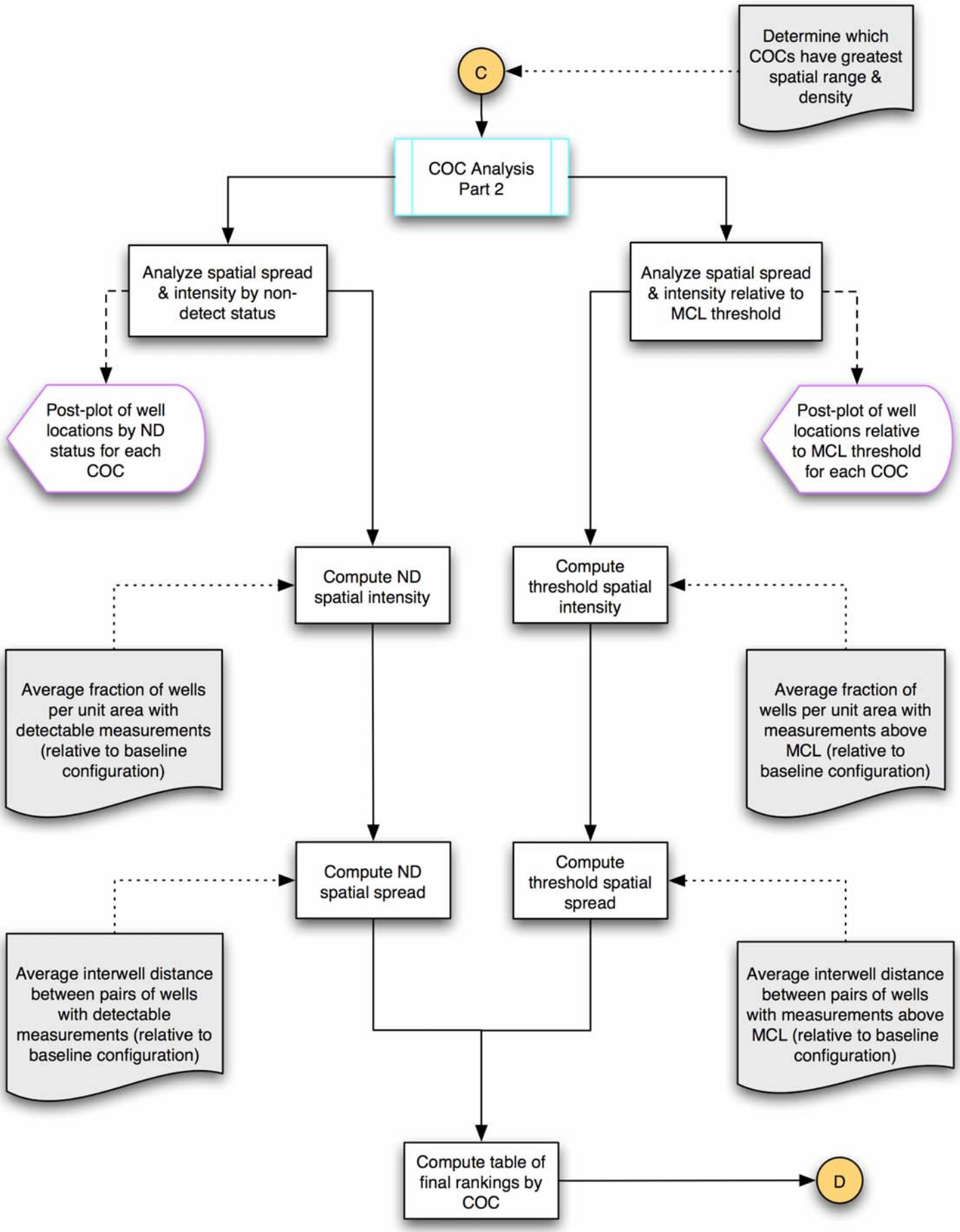
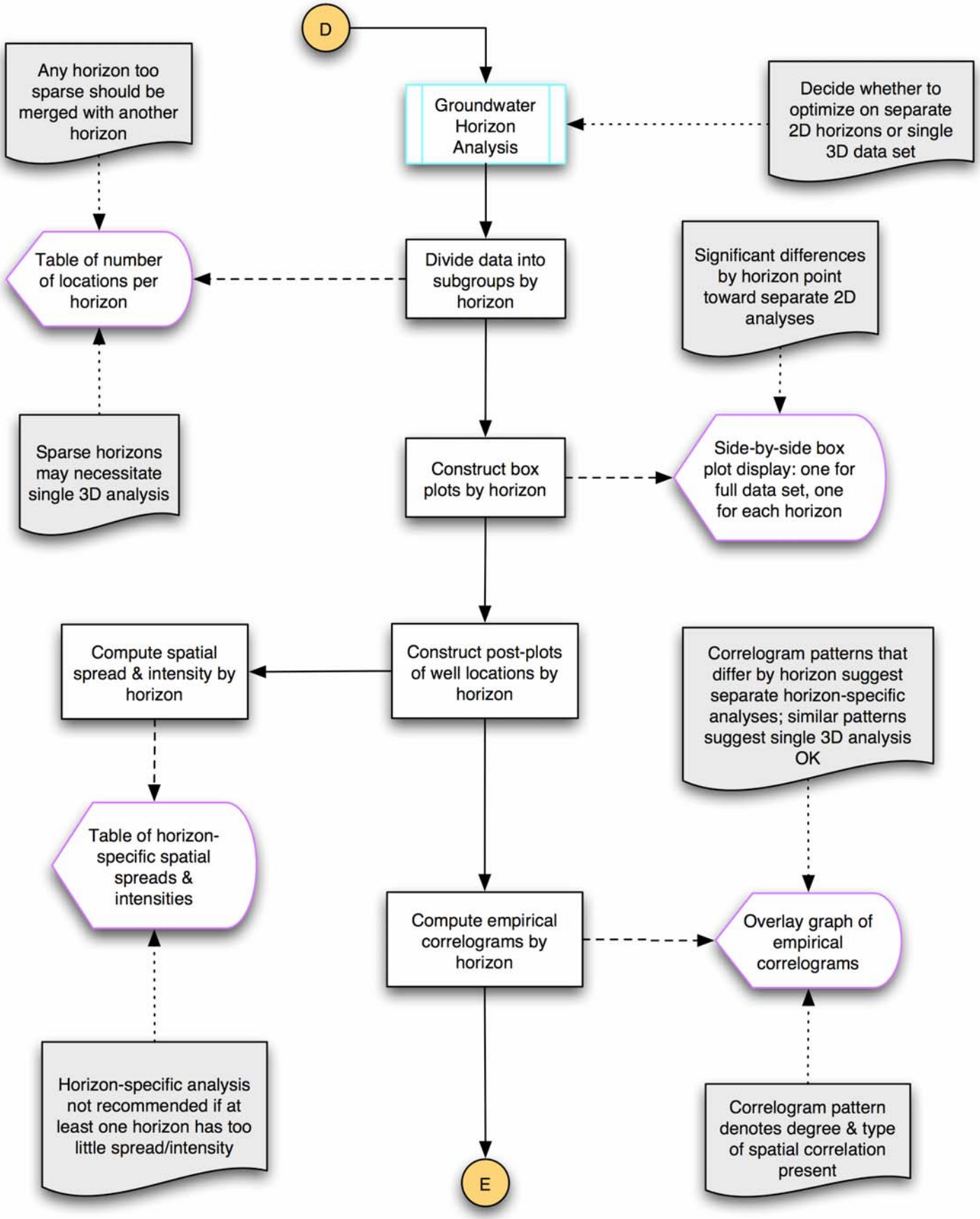
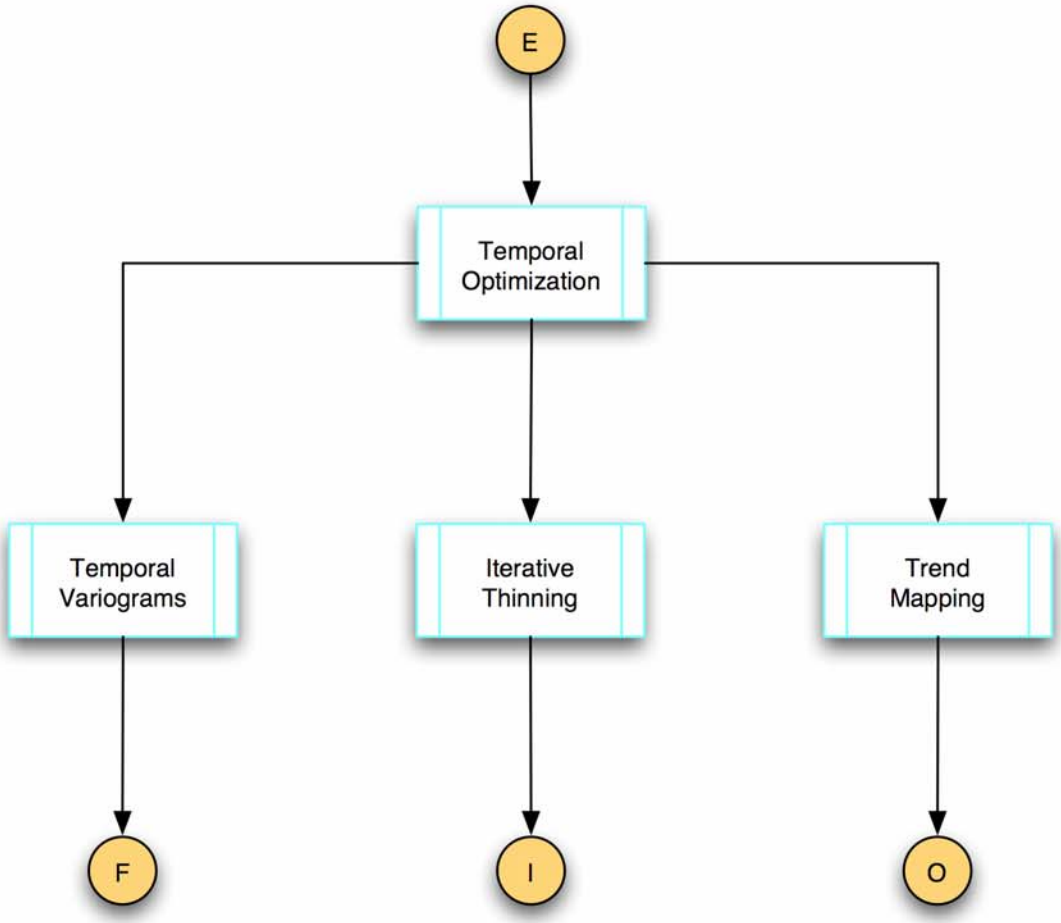

Begin GTS
A
Data Pre-processing
Data Exploration
B
COC Analysis
D
Groundwater Horizon Analysis
E
P
Temporal Optimization
Spatial Optimization
End GTS

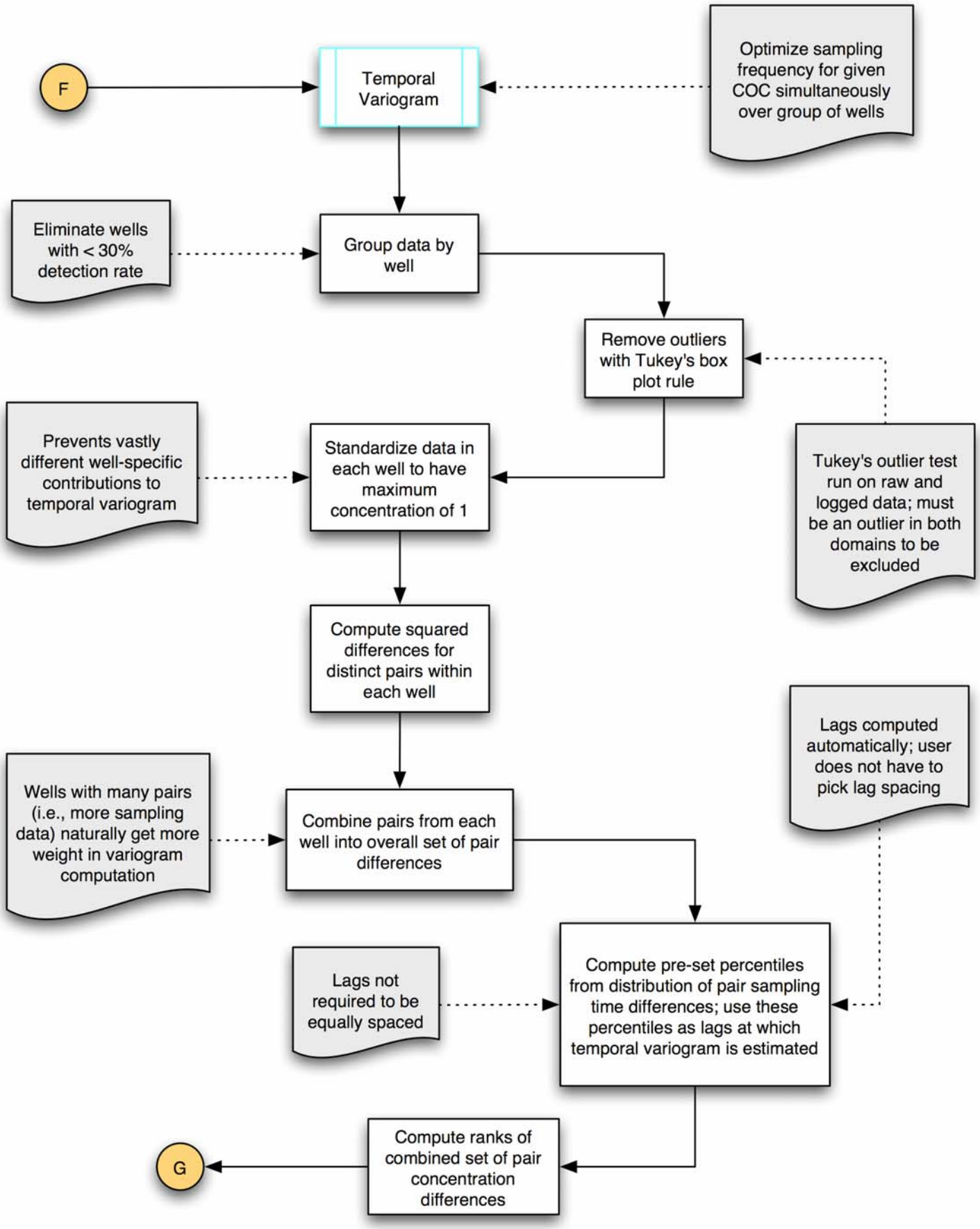

A
Data Pre-processing
Gather Data
Input measurement data
Resolve data inconsistencies; fill data gaps
Input GIS layer(s)
Input site boundary data
Exclude wells with insufficient monitoring data
Box plots by COC
Create statistical summaries
Data Exploration
Summary statistics by COC
B
Construct spatial summaries
Construct temporal summaries
P
Post plot of well locations (with GIS overlay)
Time series plots by well & COC
View of well-specific trends
Color-coded by deciles of univariate distribution
Post-plots of measurements by COC
Quick-view of contaminant patterns

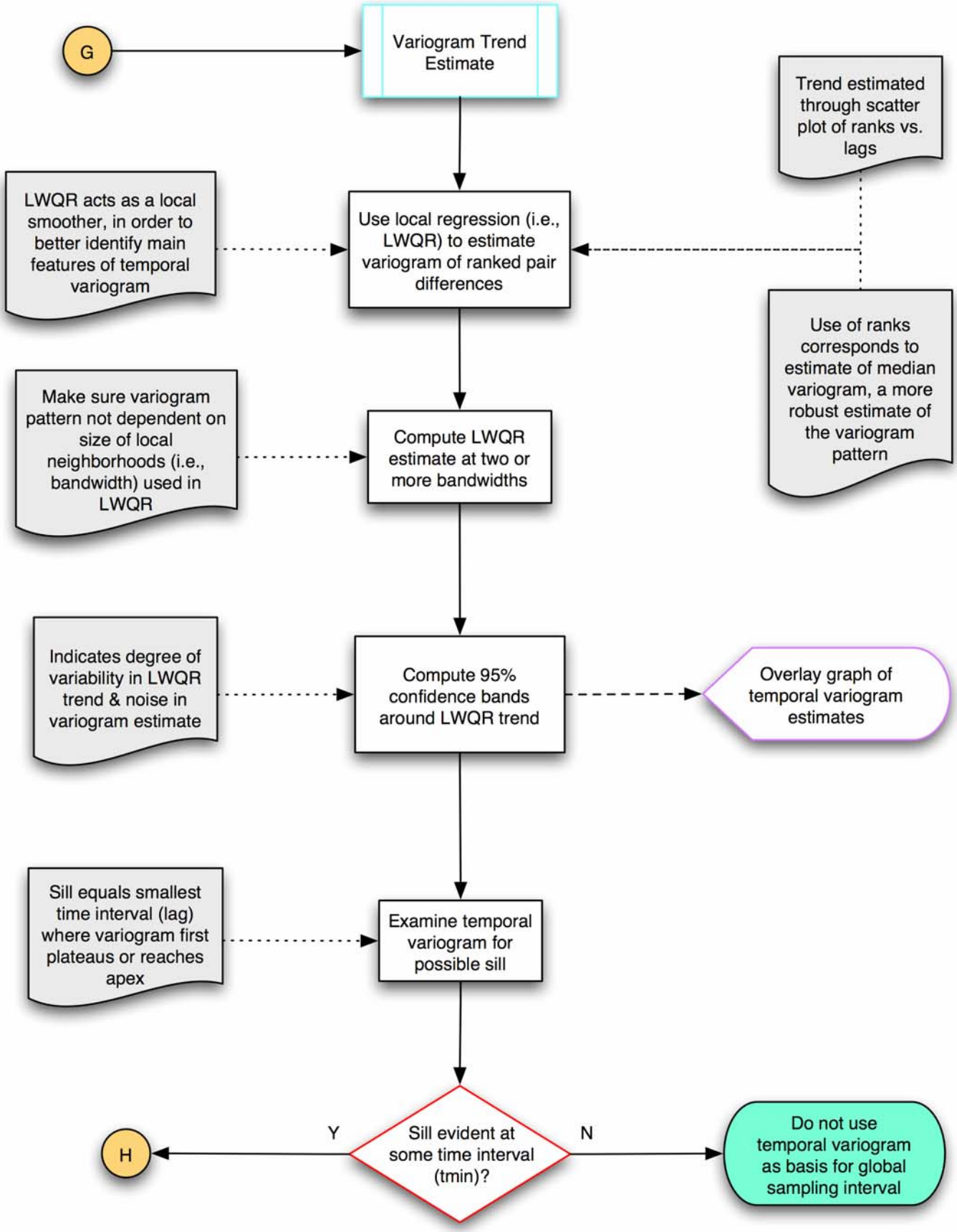

B
Compute rankings of COC optimization potential
COC Analysis Part 1
Narrow list of COCs to 2-3
Table of measurement non-detect rates by COC
Compute COC summary statistics
Table of measurement fractions below MCL by COC
Table of non-detect rates by well & COC
Compute well-specific summary statistics
Table of fractions below MCL by well & COC
Table of user-supplied COC rankings
Incorporate user-supplied rankings
Capture intangibles, political priorities, etc.
Many, few, or none
Categorize trends for each COC
C

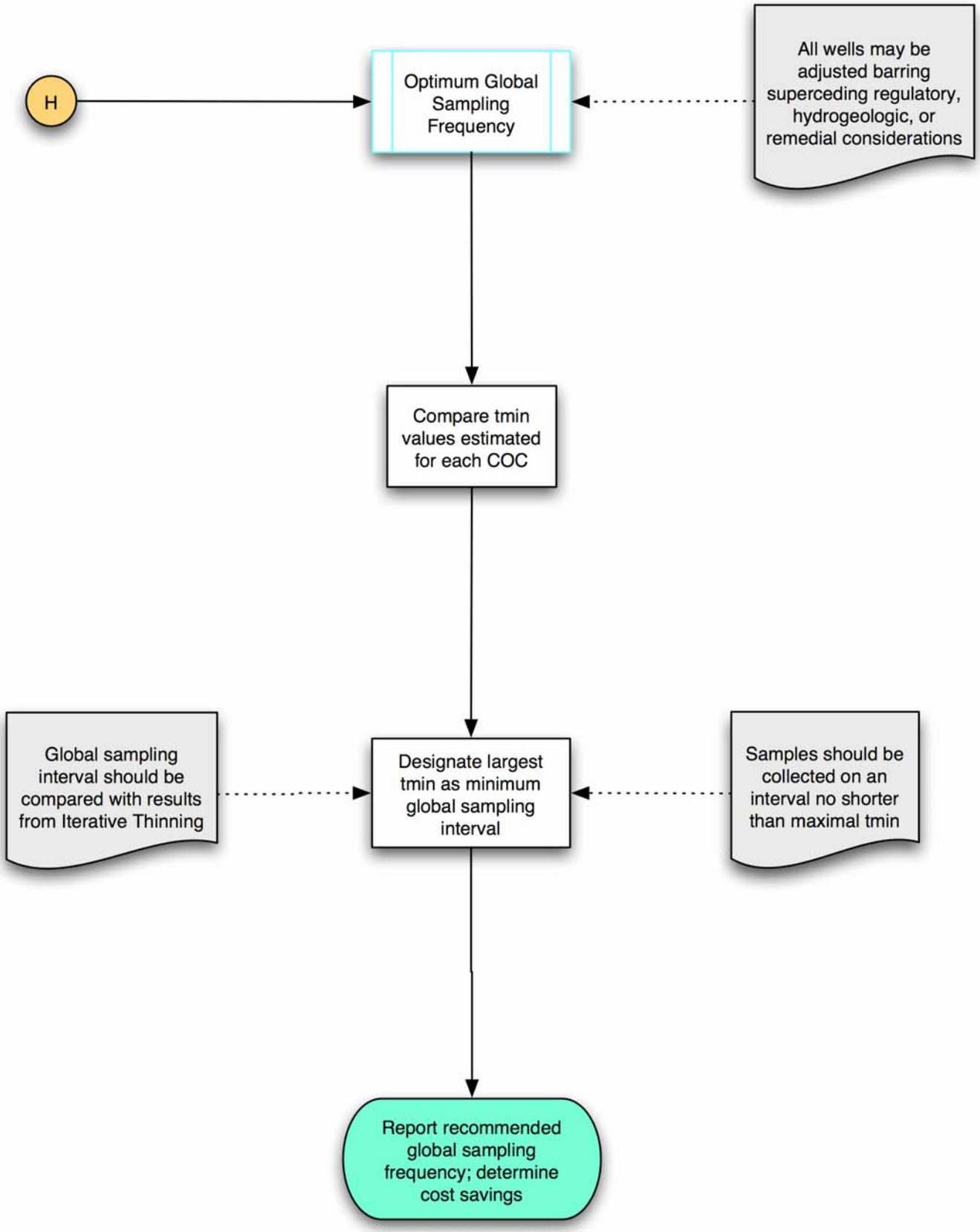

Determine which COCs have greatest spatial range & density
C
COC Analysis Part 2
Analyze spatial spread & intensity by non-detect status
Analyze spatial spread & intensity relative to MCL threshold
Post-plot of well locations by ND status for each COC
Post-plot of well locations relative to MCL threshold for each COC
Compute ND spatial intensity
Compute threshold spatial intensity
Average fraction of wells per unit area with detectable measurements (relative to baseline configuration)
Average fraction of wells per unit area with measurements above MCL (relative to baseline configuration)
Compute ND spatial spread
Compute threshold spatial spread
Average interwell distance between pairs of wells with detectable measurements (relative to baseline configuration)
Average interwell distance between pairs of wells with measurements above MCL (relative to baseline configuration)
Compute table of final rankings by COC
D

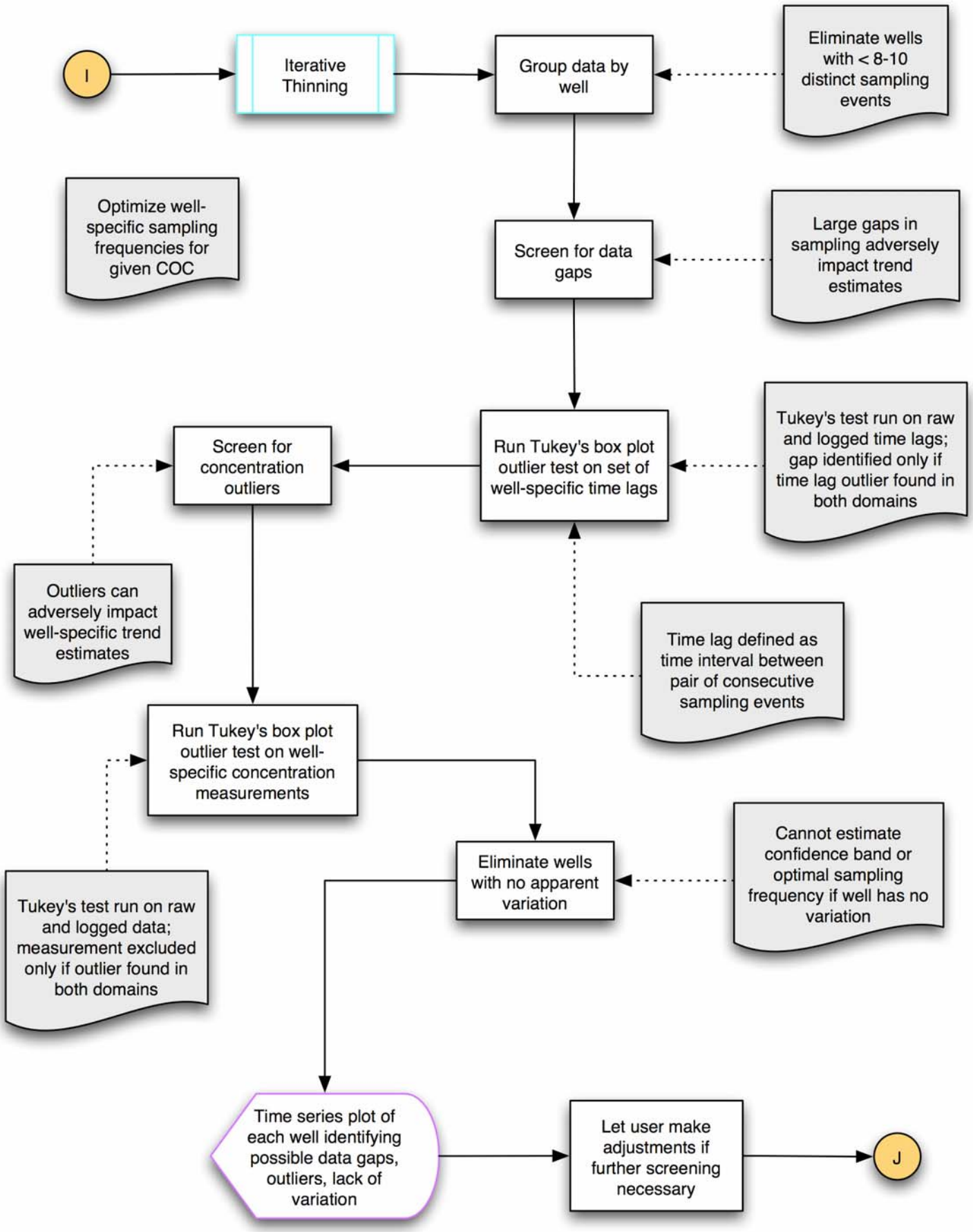

D
Any horizon too sparse should be merged with another horizon
Groundwater Horizon Analysis
Decide whether to optimize on separate 2D horizons or single 3D data set
Table of number of locations per horizon
Divide data into subgroups by horizon
Significant differences by horizon point toward separate 2D analyses
Sparse horizons may necessitate single 3D analysis
Construct box plots by horizon
Side-by-side box plot display: one for full data set, one for each horizon
Compute spatial spread & intensity by horizon
Construct post-plots of well locations by horizon
Correlogram patterns that differ by horizon suggest separate horizon-specific analyses; similar patterns suggest single 3D analysis OK
Table of horizon-specific spatial spreads & intensities
Compute empirical correlograms by horizon
Overlay graph of empirical correlograms
Horizon-specific analysis not recommended if at least one horizon has too little spread/intensity
Correlogram pattern denotes degree & type of spatial correlation present
E

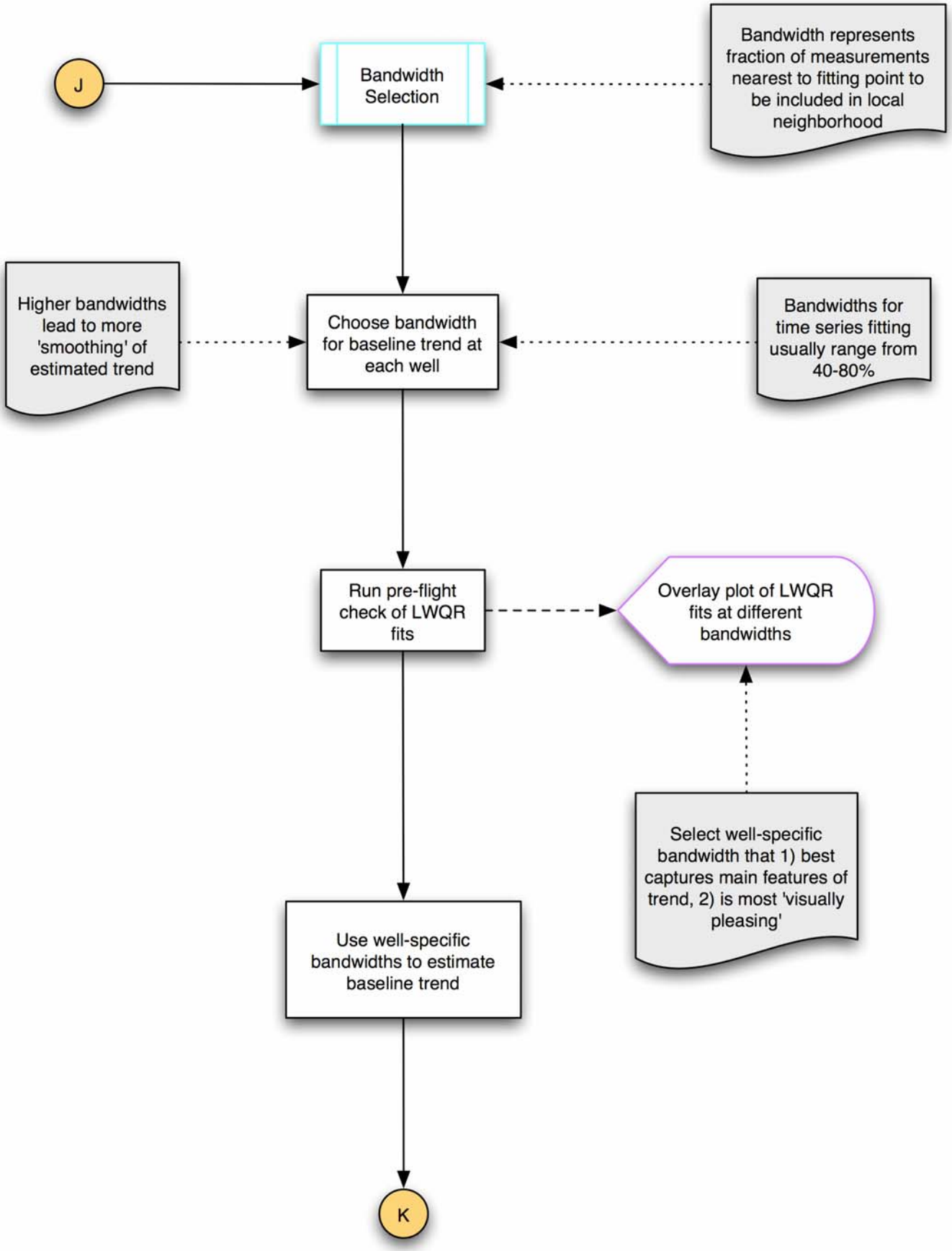

E
Temporal Optimization
Temporal Variograms
Iterative Thinning
Trend Mapping
F
I
O

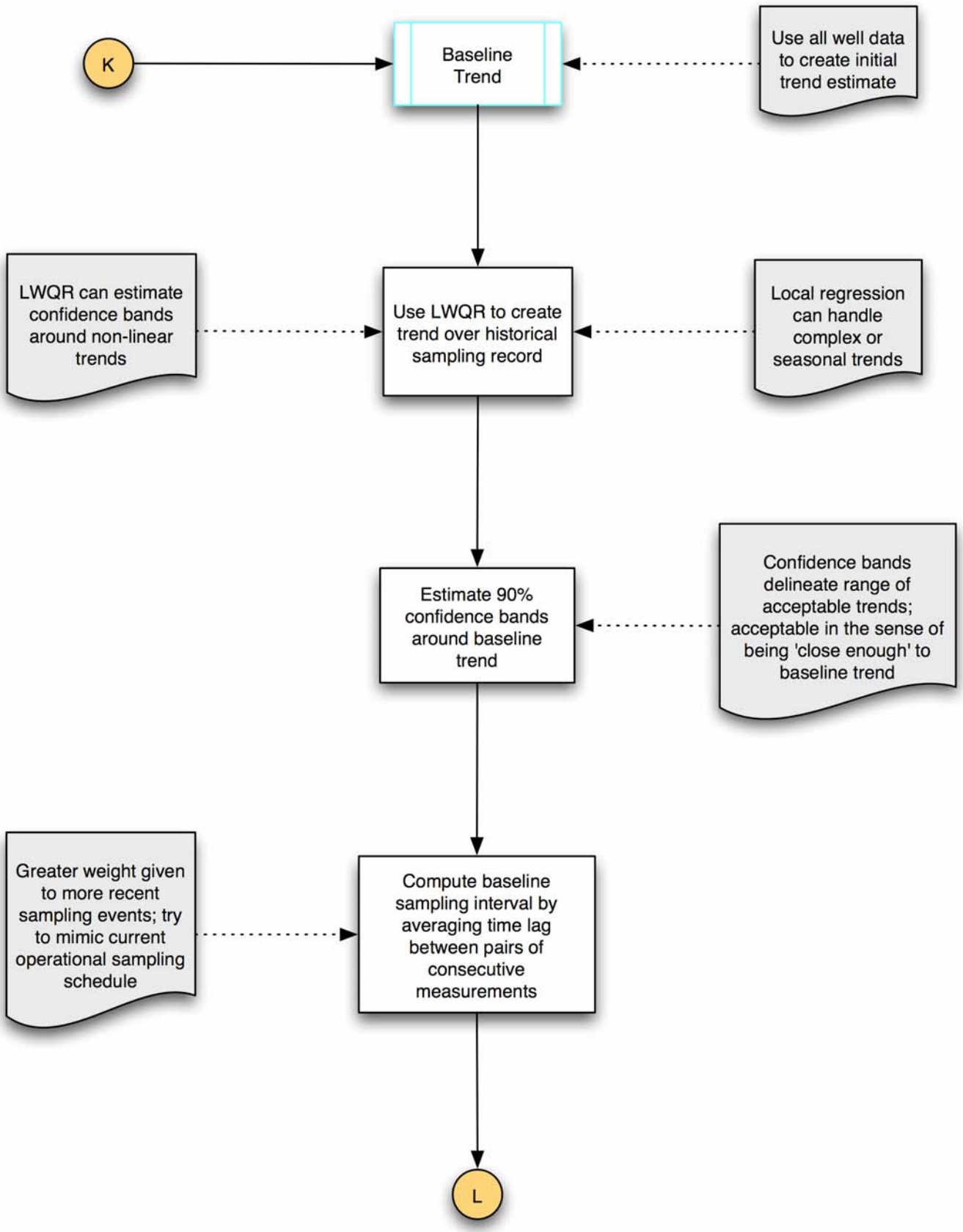

F
Temporal Variogram
Optimize sampling frequency for given COC simultaneously over group of wells
Eliminate wells with < 30% detection rate
Group data by well
Remove outliers with Tukey's box plot rule
Tukey's outlier test run on raw and logged data; must be an outlier in both domains to be excluded
Prevents vastly different well-specific contributions to temporal variogram
Standardize data in each well to have maximum concentration of 1
Compute squared differences for distinct pairs within each well
Lags computed automatically; user does not have to pick lag spacing
Wells with many pairs (i.e., more sampling data) naturally get more weight in variogram computation
Combine pairs from each well into overall set of pair differences
Lags not required to be equally spaced
Compute pre-set percentiles from distribution of pair sampling time differences; use these percentiles as lags at which temporal variogram is estimated
Compute ranks of combined set of pair concentration differences
G

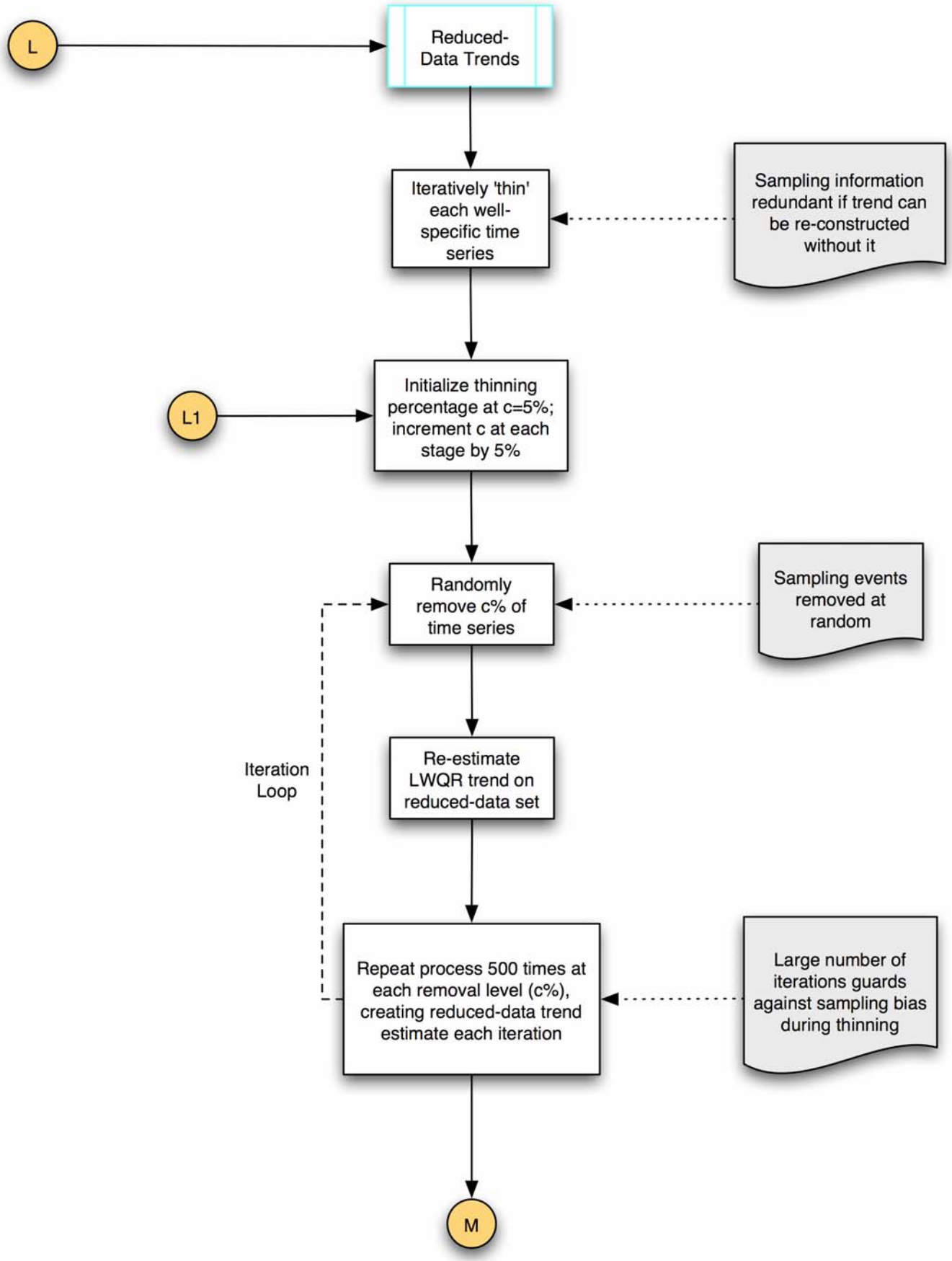

G
Variogram Trend Estimate
Trend estimated through scatter plot of ranks vs. lags
LWQR acts as a local smoother, in order to better identify main features of temporal variogram
Use local regression (i.e., LWQR) to estimate variogram of ranked pair differences
Use of ranks corresponds to estimate of median variogram, a more robust estimate of the variogram pattern
Make sure variogram pattern not dependent on size of local neighborhoods (i.e., bandwidth) used in LWQR
Compute LWQR estimate at two or more bandwidths
Indicates degree of variability in LWQR trend & noise in variogram estimate
Compute 95% confidence bands around LWQR trend
Overlay graph of temporal variogram estimates
Sill equals smallest time interval (lag) where variogram first plateaus or reaches apex
Examine temporal variogram for possible sill
Y
Sill evident at some time interval (tmin)?
N
H
Do not use temporal variogram as basis for global sampling interval

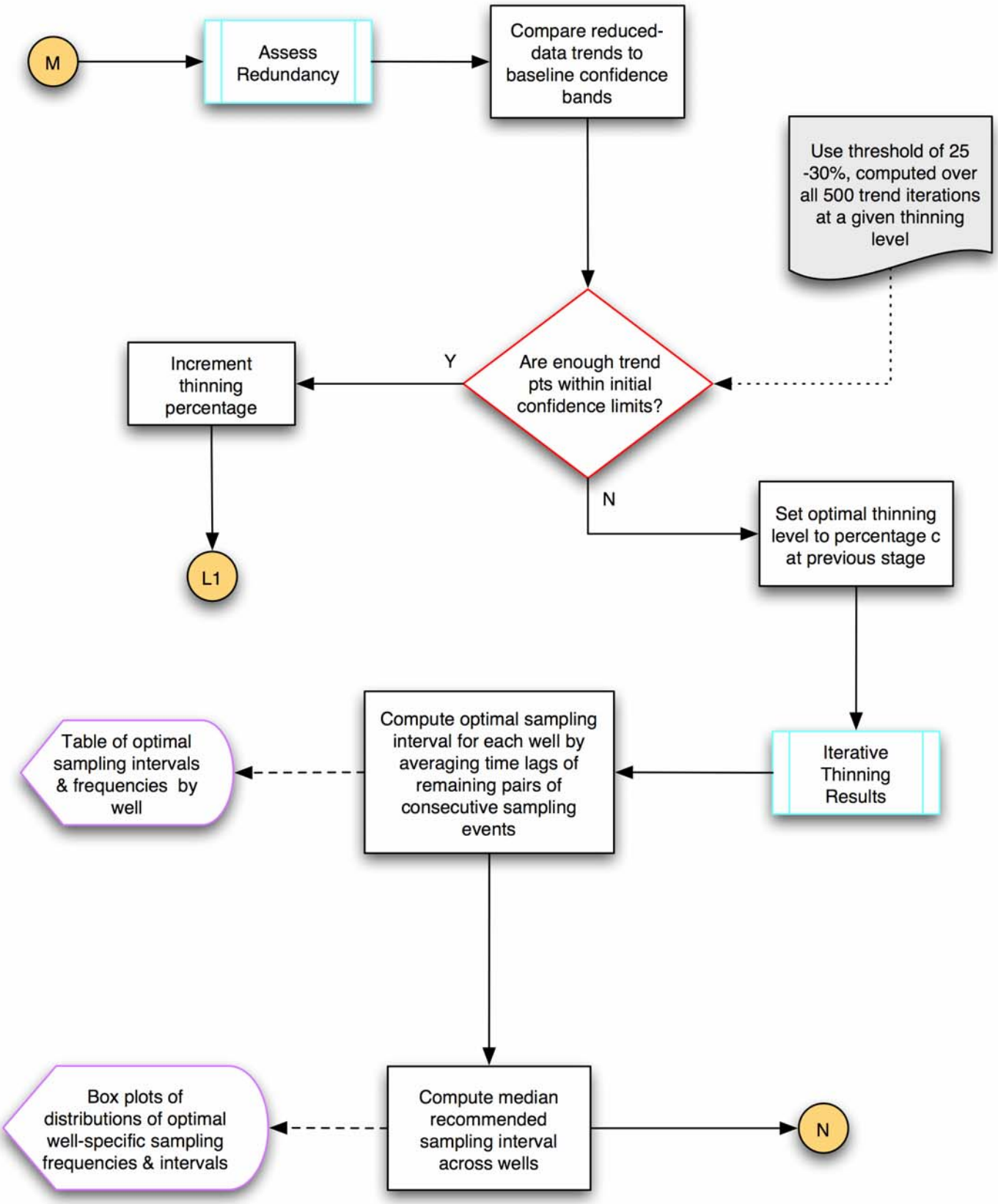

H
Optimum Global Sampling Frequency
All wells may be adjusted barring superceding regulatory, hydrogeologic, or remedial considerations
Compare tmin values estimated for each COC
Global sampling interval should be compared with results from Iterative Thinning
Designate largest tmin as minimum global sampling interval
Samples should be collected on an interval no shorter than maximal tmin
Report recommended global sampling frequency; determine cost savings

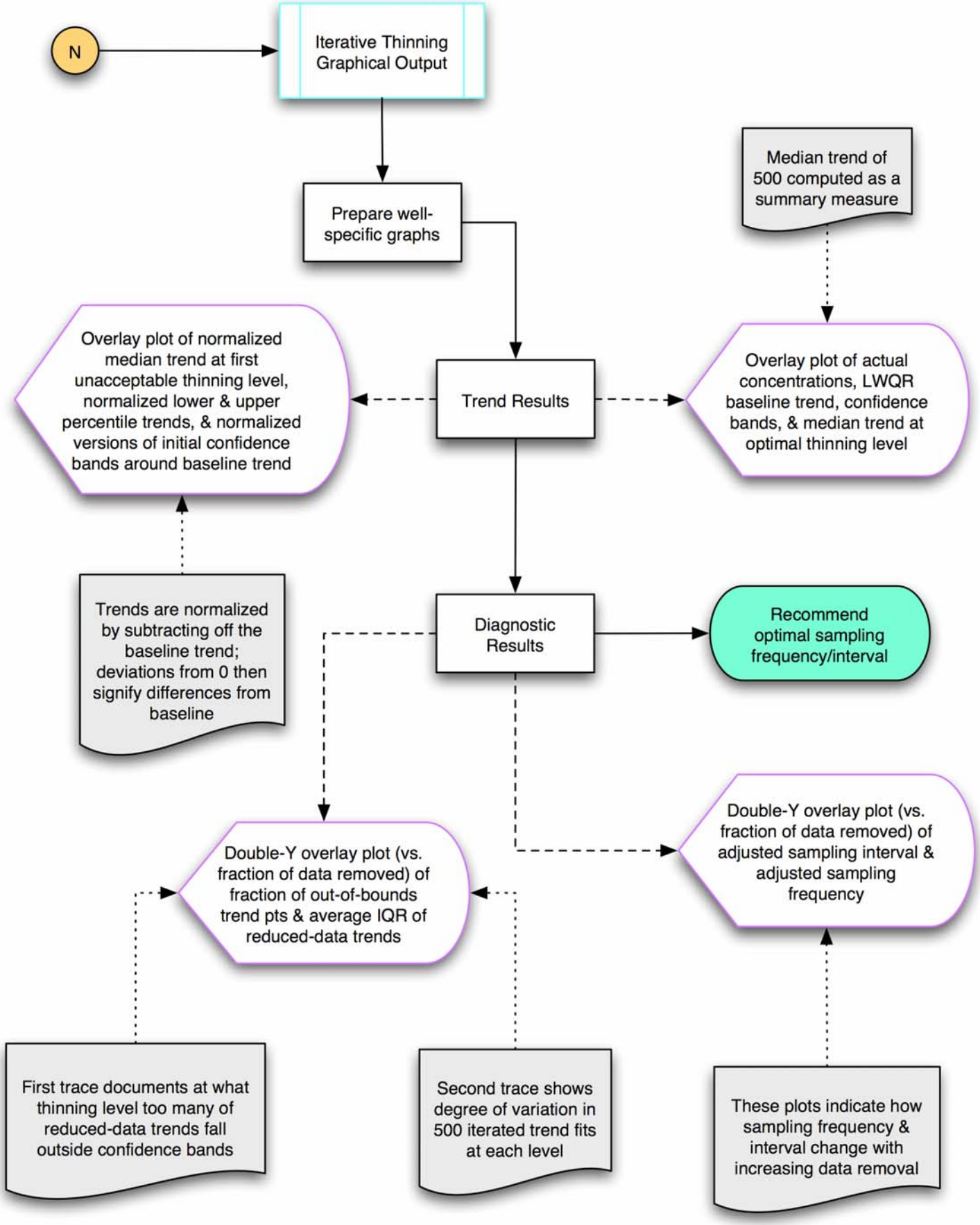

I
Iterative Thinning
Group data by well
Eliminate wells with < 8-10 distinct sampling events
Optimize well-specific sampling frequencies for given COC
Screen for data gaps
Large gaps in sampling adversely impact trend estimates
Run Tukey's box plot outlier test on set of well-specific time lags
Tukey's test run on raw and logged time lags; gap identified only if time lag outlier found in both domains
Screen for concentration outliers
Outliers can adversely impact well-specific trend estimates
Time lag defined as time interval between pair of consecutive sampling events
Run Tukey's box plot outlier test on well-specific concentration measurements
Tukey's test run on raw and logged data; measurement excluded only if outlier found in both domains
Eliminate wells with no apparent variation
Cannot estimate confidence band or optimal sampling frequency if well has no variation
Time series plot of each well identifying possible data gaps, outliers, lack of variation
Let user make adjustments if further screening necessary
J

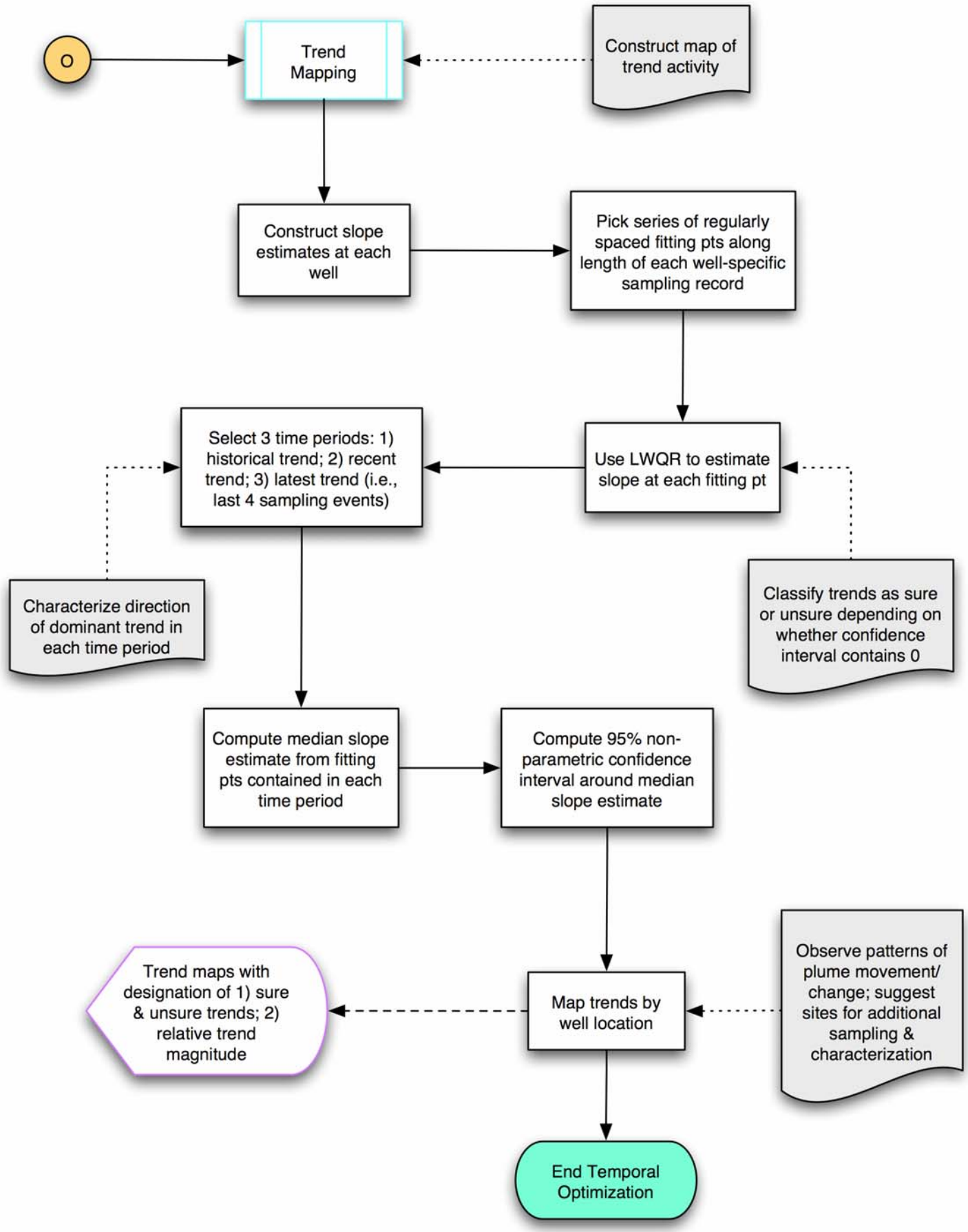

J
Bandwidth Selection
Bandwidth represents fraction of measurements nearest to fitting point to be included in local neighborhood
Higher bandwidths lead to more 'smoothing' of estimated trend
Choose bandwidth for baseline trend at each well
Bandwidths for time series fitting usually range from 40-80%
Run pre-flight check of LWQR fits
Overlay plot of LWQR fits at different bandwidths
Select well-specific bandwidth that 1) best captures main features of trend, 2) is most 'visually pleasing'
Use well-specific bandwidths to estimate baseline trend
K


K
Baseline Trend
Use all well data to create initial trend estimate
LWQR can estimate confidence bands around non-linear trends
Use LWQR to create trend over historical sampling record
Local regression can handle complex or seasonal trends
Estimate 90% confidence bands around baseline trend
Confidence bands delineate range of acceptable trends; acceptable in the sense of being 'close enough' to baseline trend
Greater weight given to more recent sampling events; try to mimic current operational sampling schedule
Compute baseline sampling interval by averaging time lag between pairs of consecutive measurements
L


L
Reduced-Data Trends
Iteratively 'thin' each well-specific time series
Sampling information redundant if trend can be re-constructed without it
L1
Initialize thinning percentage at c=5%; increment c at each stage by 5%
Randomly remove c% of time series
Sampling events removed at random
Iteration Loop
Re-estimate LWQR trend on reduced-data set
Repeat process 500 times at each removal level (c%), creating reduced-data trend estimate each iteration
Large number of iterations guards against sampling bias during thinning
M


M
Assess Redundancy
Compare reduced-data trends to baseline confidence bands
Use threshold of 25 -30%, computed over all 500 trend iterations at a given thinning level
Are enough trend pts within initial confidence limits?
Y
Increment thinning percentage
L1
N
Set optimal thinning level to percentage c at previous stage
Iterative Thinning Results
Compute optimal sampling interval for each well by averaging time lags of remaining pairs of consecutive sampling events
Table of optimal sampling intervals & frequencies by well
Compute median recommended sampling interval across wells
Box plots of distributions of optimal well-specific sampling frequencies & intervals
N


N
Iterative Thinning Graphical Output
Prepare well-specific graphs
Median trend of 500 computed as a summary measure
Overlay plot of normalized median trend at first unacceptable thinning level, normalized lower & upper percentile trends, & normalized versions of initial confidence bands around baseline trend
Trend Results
Overlay plot of actual concentrations, LWQR baseline trend, confidence bands, & median trend at optimal thinning level
Trends are normalized by subtracting off the baseline trend; deviations from 0 then signify differences from baseline
Diagnostic Results
Recommend optimal sampling frequency/interval
Double-Y overlay plot (vs. fraction of data removed) of fraction of out-of-bounds trend pts & average IQR of reduced-data trends
Double-Y overlay plot (vs. fraction of data removed) of adjusted sampling interval & adjusted sampling frequency
First trace documents at what thinning level too many of reduced-data trends fall outside confidence bands
Second trace shows degree of variation in 500 iterated trend fits at each level
These plots indicate how sampling frequency & interval change with increasing data removal


Trend Mapping
Construct map of trend activity
Construct slope estimates at each well
Pick series of regularly spaced fitting pts along length of each well-specific sampling record
Select 3 time periods: 1) historical trend; 2) recent trend; 3) latest trend (i.e., last 4 sampling events)
Use LWQR to estimate slope at each fitting pt
Characterize direction of dominant trend in each time period
Classify trends as sure or unsure depending on whether confidence interval contains 0
Compute median slope estimate from fitting pts contained in each time period
Compute 95% non-parametric confidence interval around median slope estimate
Trend maps with designation of 1) sure & unsure trends; 2) relative trend magnitude
Map trends by well location
Observe patterns of plume movement/ change; suggest sites for additional sampling & characterization
End Temporal Optimization