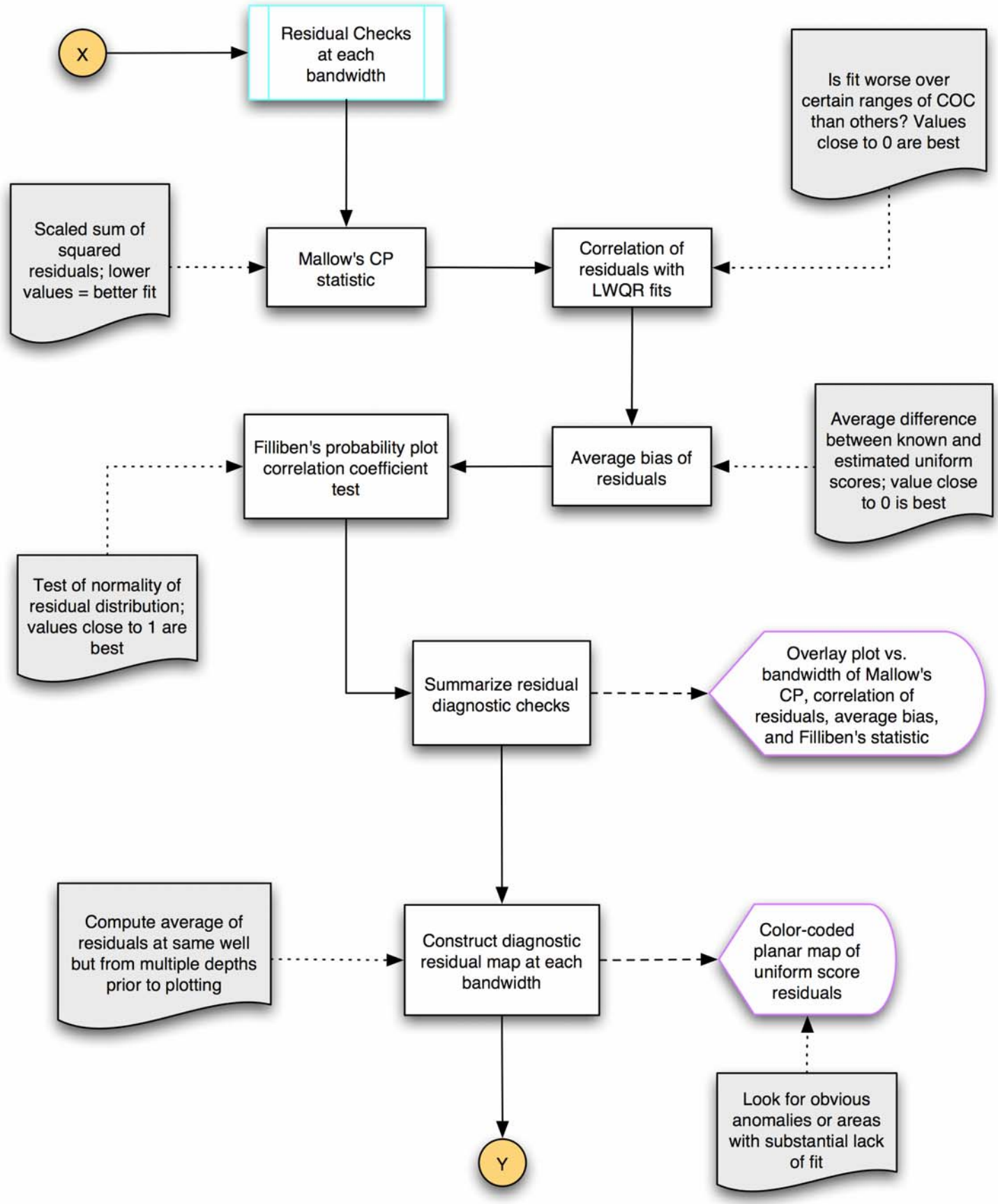
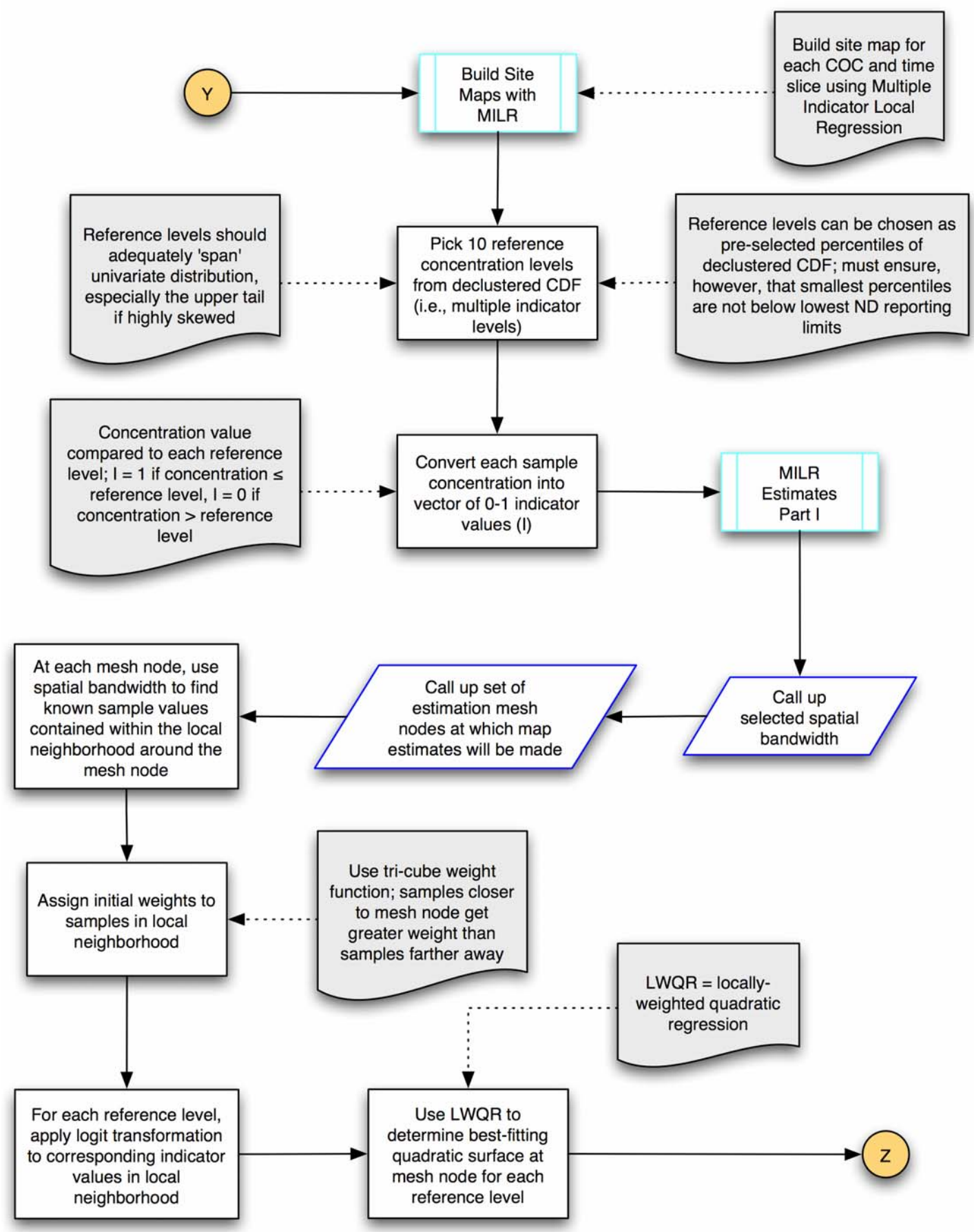
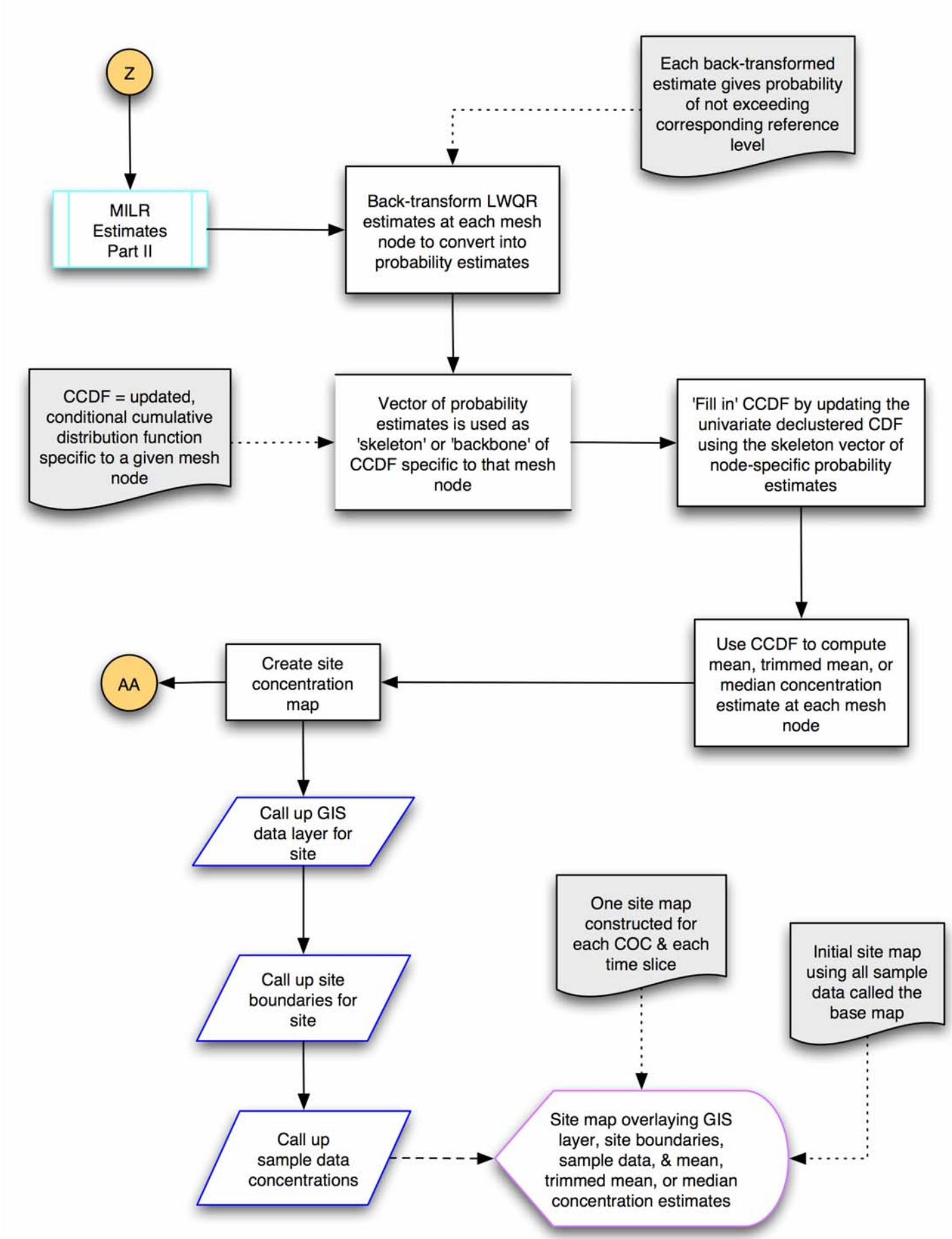
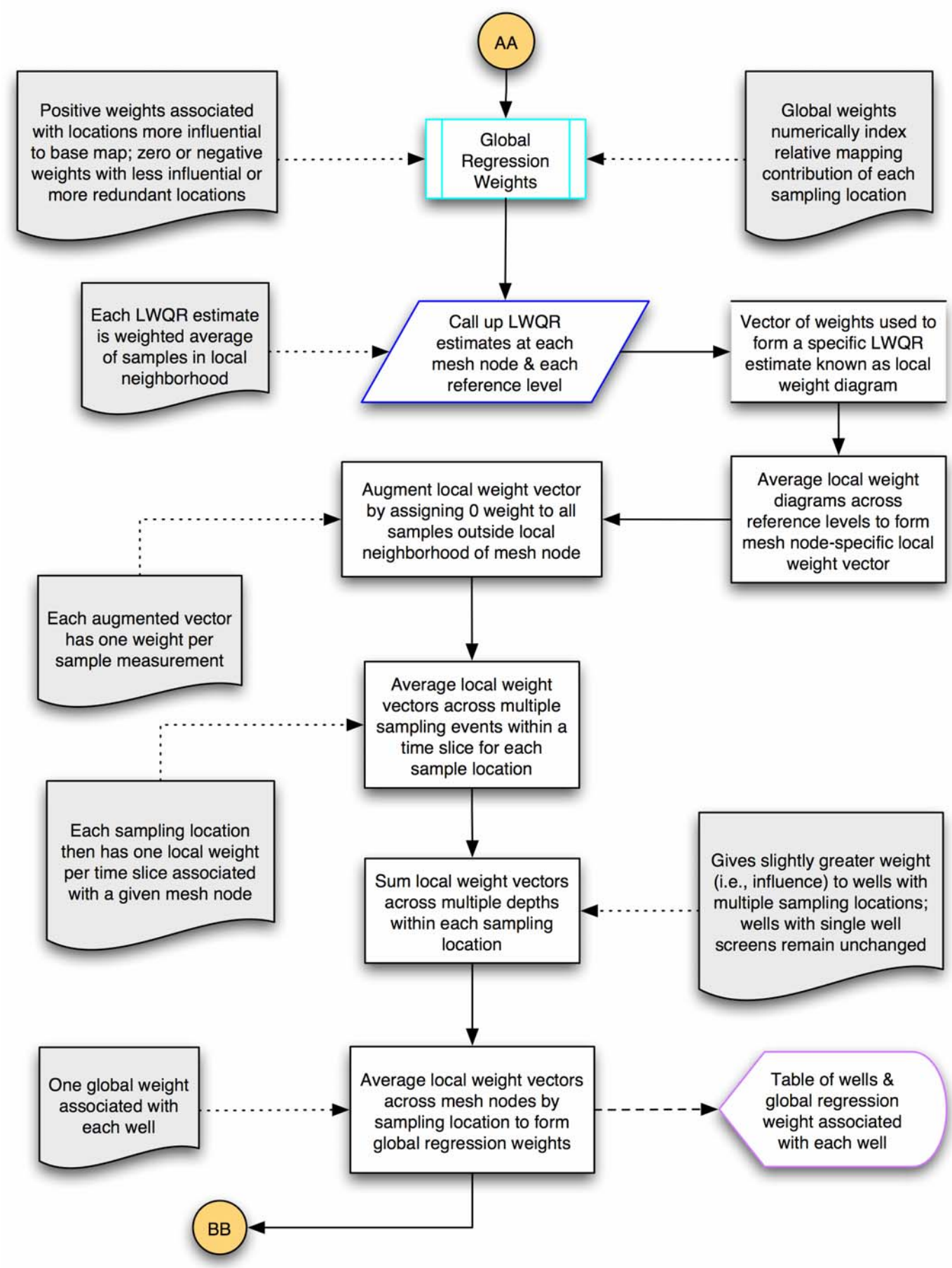
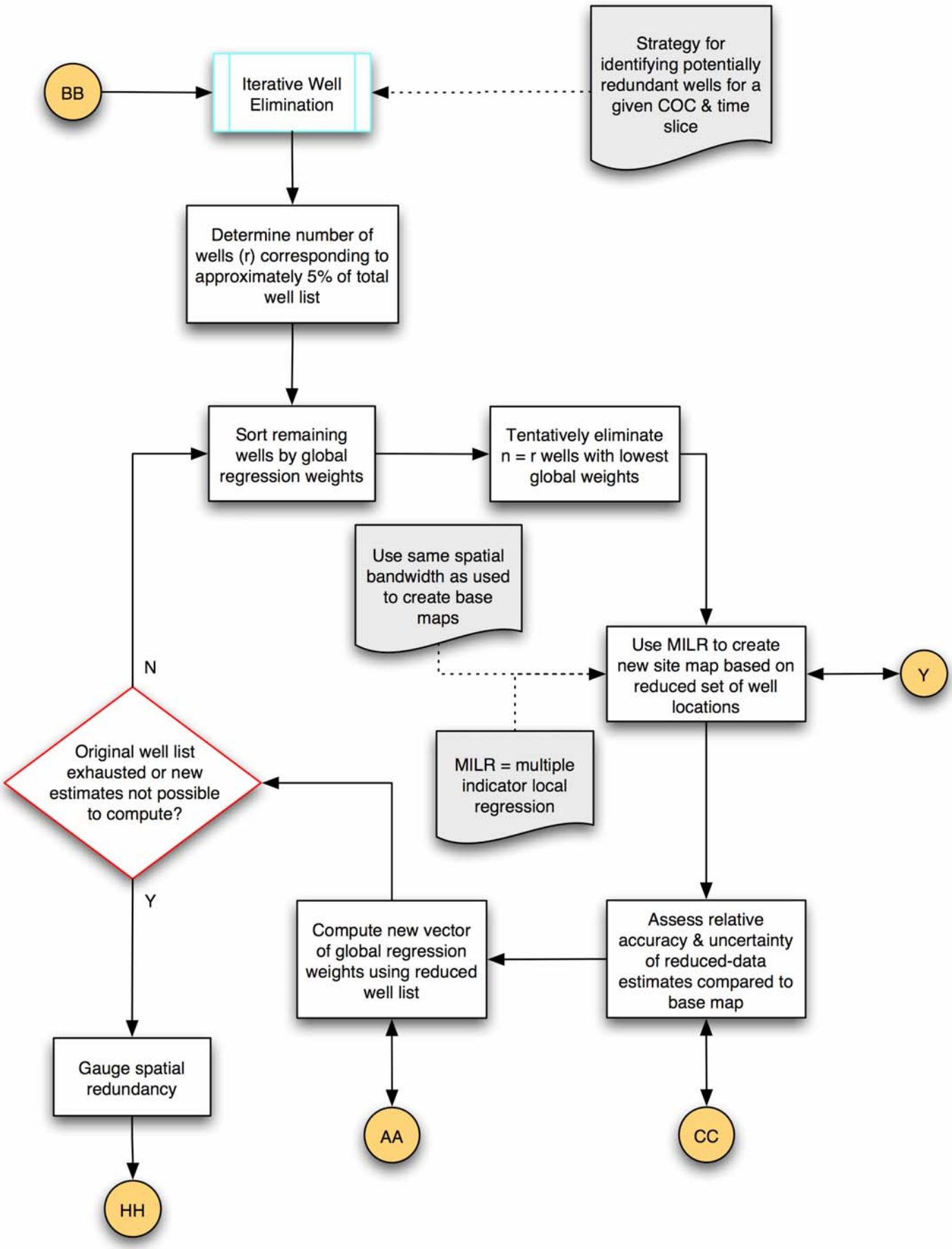
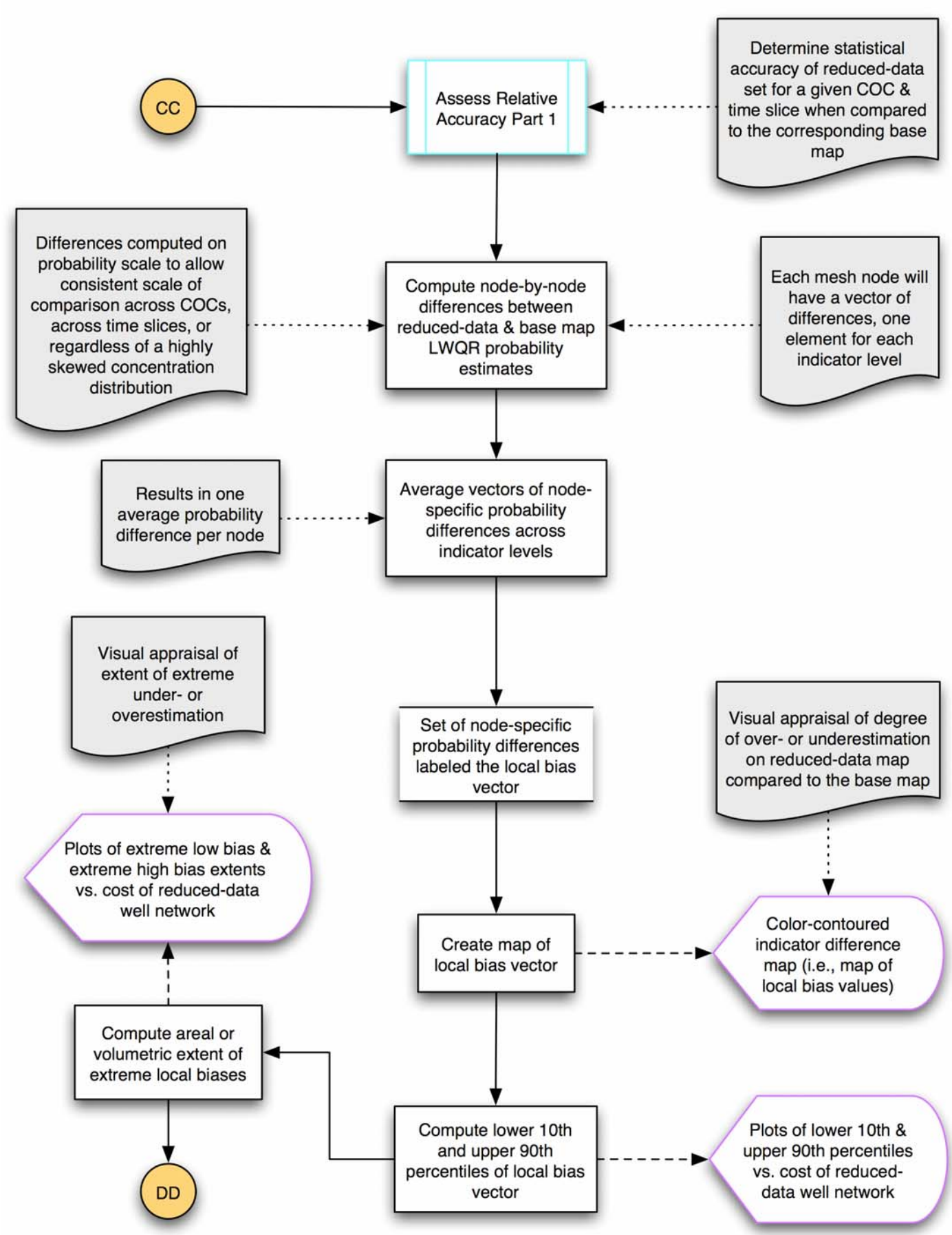

X
Residual Checks at each bandwidth
Is fit worse over certain ranges of COC than others? Values close to 0 are best
Scaled sum of squared residuals; lower values = better fit
Mallow's CP statistic
Correlation of residuals with LWQR fits
Average difference between known and estimated uniform scores; value close to 0 is best
Filliben's probability plot correlation coefficient test
Average bias of residuals
Test of normality of residual distribution; values close to 1 are best
Summarize residual diagnostic checks
Overlay plot vs. bandwidth of Mallow's CP, correlation of residuals, average bias, and Filliben's statistic
Compute average of residuals at same well but from multiple depths prior to plotting
Construct diagnostic residual map at each bandwidth
Color-coded planar map of uniform score residuals
Look for obvious anomalies or areas with substantial lack of fit
Y

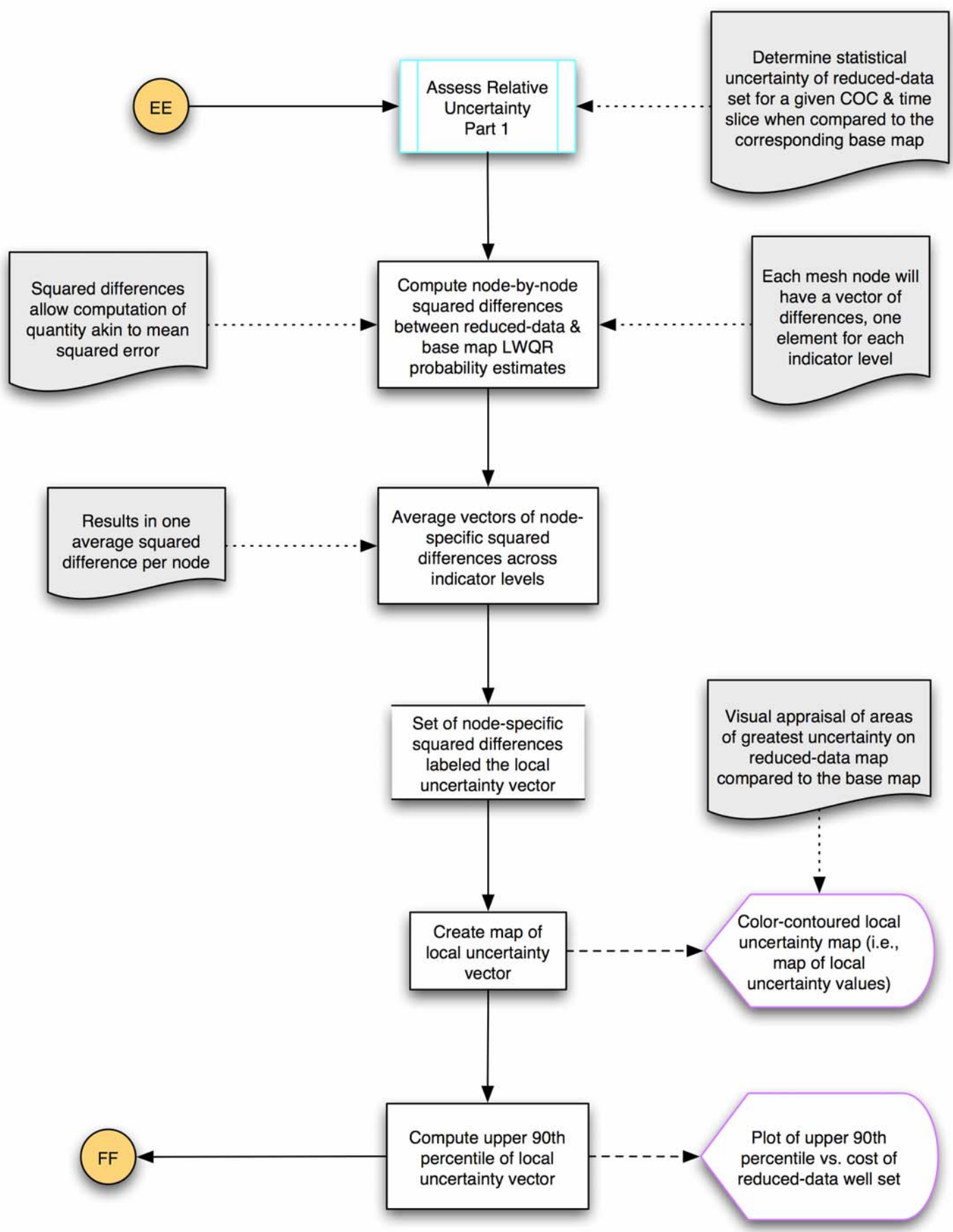

Y
Build Site Maps with MILR
Build site map for each COC and time slice using Multiple Indicator Local Regression
Reference levels should adequately 'span' univariate distribution, especially the upper tail if highly skewed
Pick 10 reference concentration levels from declustered CDF (i.e., multiple indicator levels)
Reference levels can be chosen as pre-selected percentiles of declustered CDF; must ensure, however, that smallest percentiles are not below lowest ND reporting limits
Concentration value compared to each reference level; I = 1 if concentration ≤ reference level, I = 0 if concentration > reference level
Convert each sample concentration into vector of 0-1 indicator values (I)
MILR Estimates Part I
At each mesh node, use spatial bandwidth to find known sample values contained within the local neighborhood around the mesh node
Call up set of estimation mesh nodes at which map estimates will be made
Call up selected spatial bandwidth
Assign initial weights to samples in local neighborhood
Use tri-cube weight function; samples closer to mesh node get greater weight than samples farther away
LWQR = locally-weighted quadratic regression
For each reference level, apply logit transformation to corresponding indicator values in local neighborhood
Use LWQR to determine best-fitting quadratic surface at mesh node for each reference level
Z

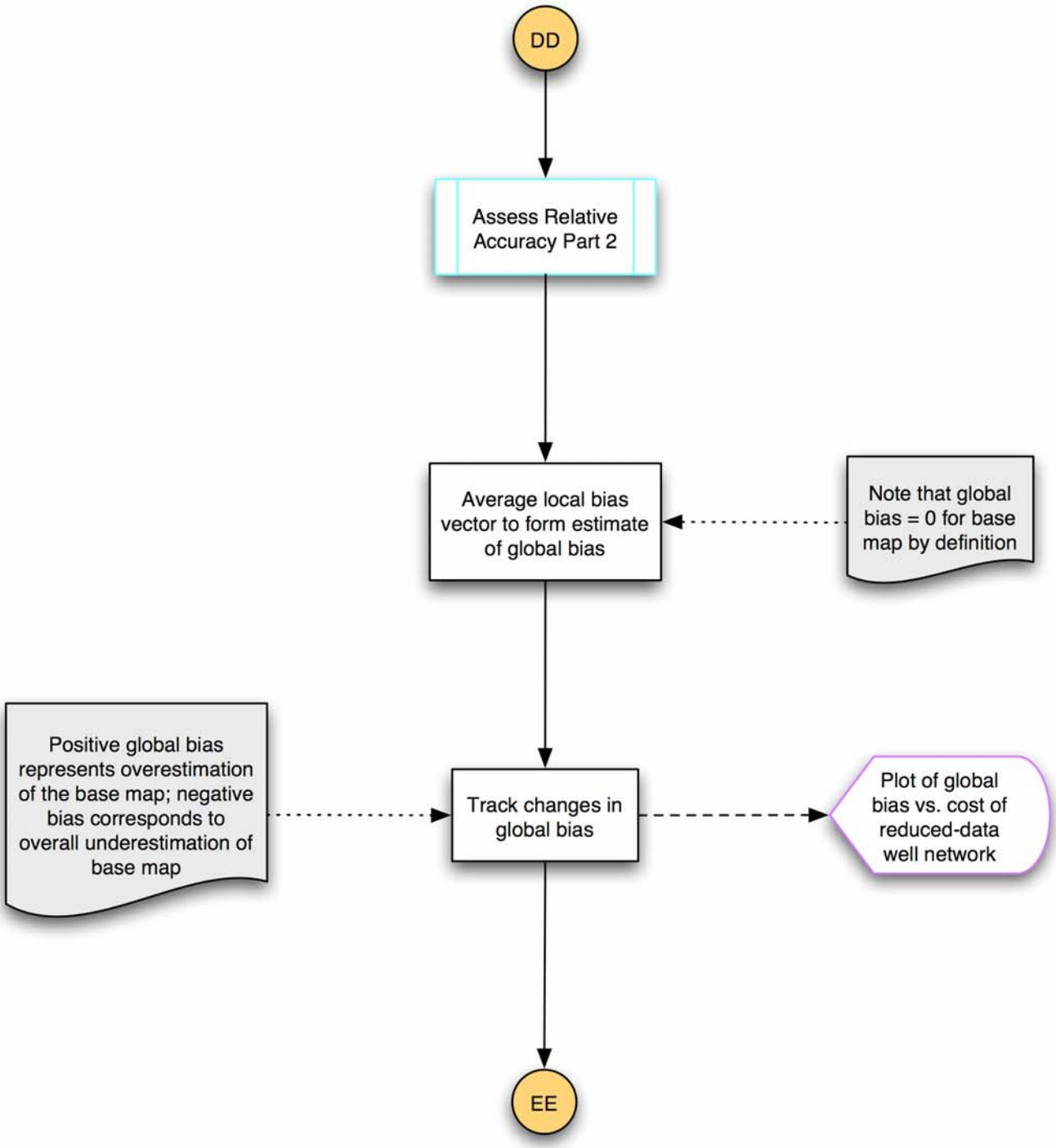

Z
MILR Estimates Part II
Back-transform LWQR estimates at each mesh node to convert into probability estimates
Each back-transformed estimate gives probability of not exceeding corresponding reference level
CCDF = updated, conditional cumulative distribution function specific to a given mesh node
Vector of probability estimates is used as 'skeleton' or 'backbone' of CCDF specific to that mesh node
'Fill in' CCDF by updating the univariate declustered CDF using the skeleton vector of node-specific probability estimates
Use CCDF to compute mean, trimmed mean, or median concentration estimate at each mesh node
Create site concentration map
AA
Call up GIS data layer for site
Call up site boundaries for site
Call up sample data concentrations
One site map constructed for each COC & each time slice
Initial site map using all sample data called the base map
Site map overlaying GIS layer, site boundaries, sample data, & mean, trimmed mean, or median concentration estimates

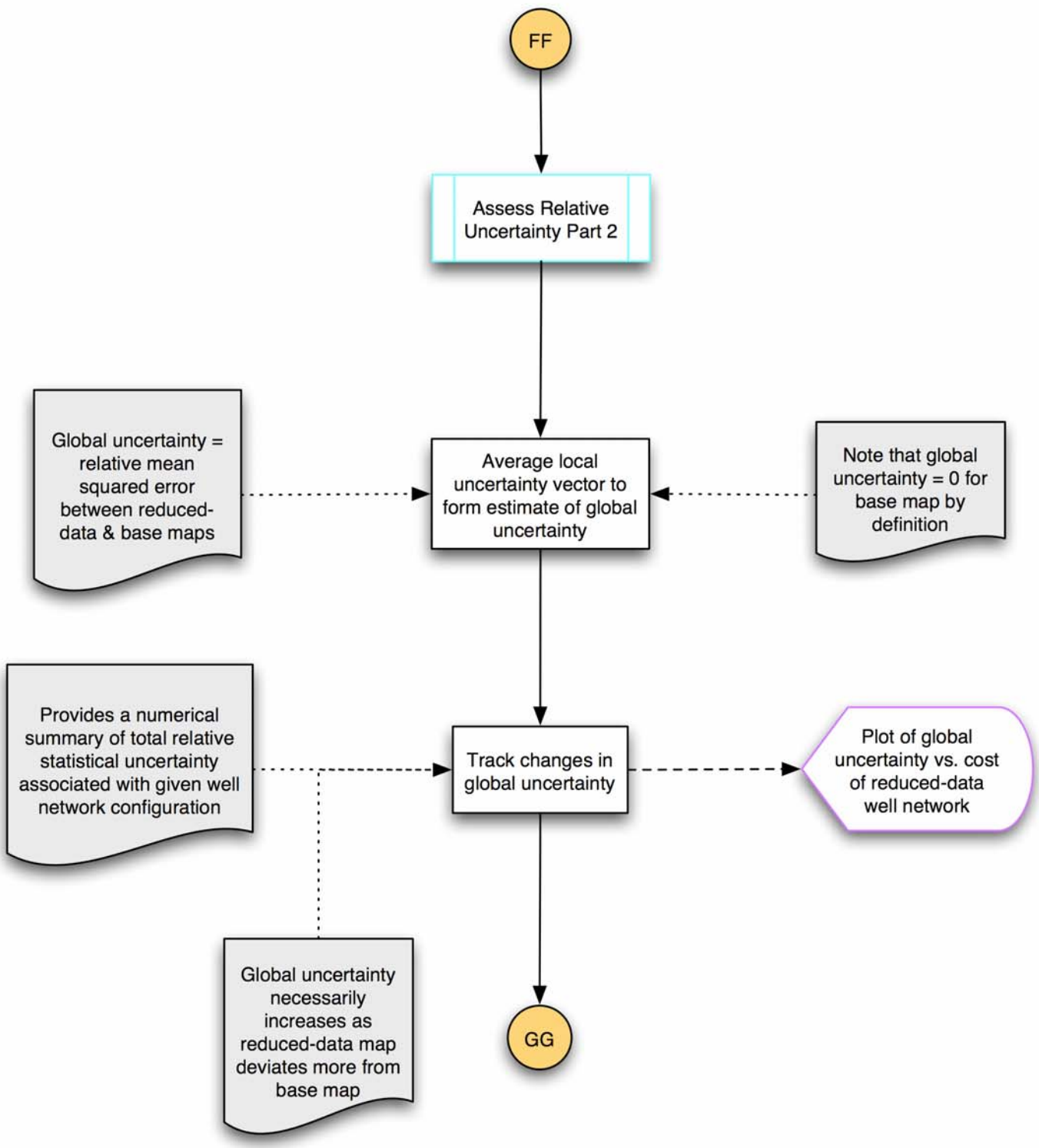

AA
Positive weights associated with locations more influential to base map; zero or negative weights with less influential or more redundant locations
Global Regression Weights
Global weights numerically index relative mapping contribution of each sampling location
Each LWQR estimate is weighted average of samples in local neighborhood
Call up LWQR estimates at each mesh node & each reference level
Vector of weights used to form a specific LWQR estimate known as local weight diagram
Average local weight diagrams across reference levels to form mesh node-specific local weight vector
Augment local weight vector by assigning 0 weight to all samples outside local neighborhood of mesh node
Each augmented vector has one weight per sample measurement
Average local weight vectors across multiple sampling events within a time slice for each sample location
Each sampling location then has one local weight per time slice associated with a given mesh node
Sum local weight vectors across multiple depths within each sampling location
Gives slightly greater weight (i.e., influence) to wells with multiple sampling locations; wells with single well screens remain unchanged
One global weight associated with each well
Average local weight vectors across mesh nodes by sampling location to form global regression weights
Table of wells & global regression weight associated with each well
BB

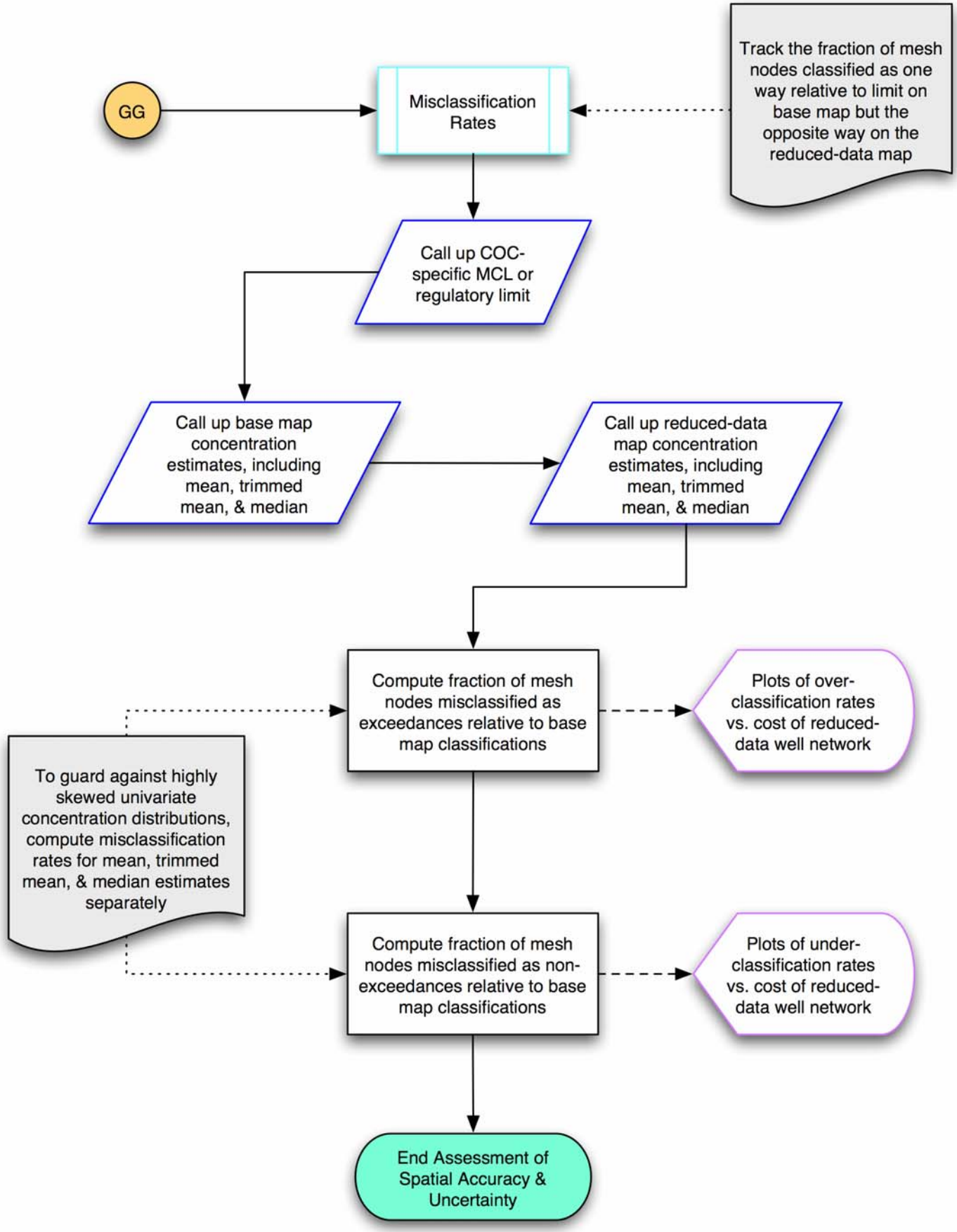

BB
Iterative Well Elimination
Strategy for identifying potentially redundant wells for a given COC & time slice
Determine number of wells (r) corresponding to approximately 5% of total well list
Sort remaining wells by global regression weights
Tentatively eliminate n = r wells with lowest global weights
Use same spatial bandwidth as used to create base maps
Use MILR to create new site map based on reduced set of well locations
Y
N
Original well list exhausted or new estimates not possible to compute?
MILR = multiple indicator local regression
Y
Compute new vector of global regression weights using reduced well list
Assess relative accuracy & uncertainty of reduced-data estimates compared to base map
Gauge spatial redundancy
AA
CC
HH

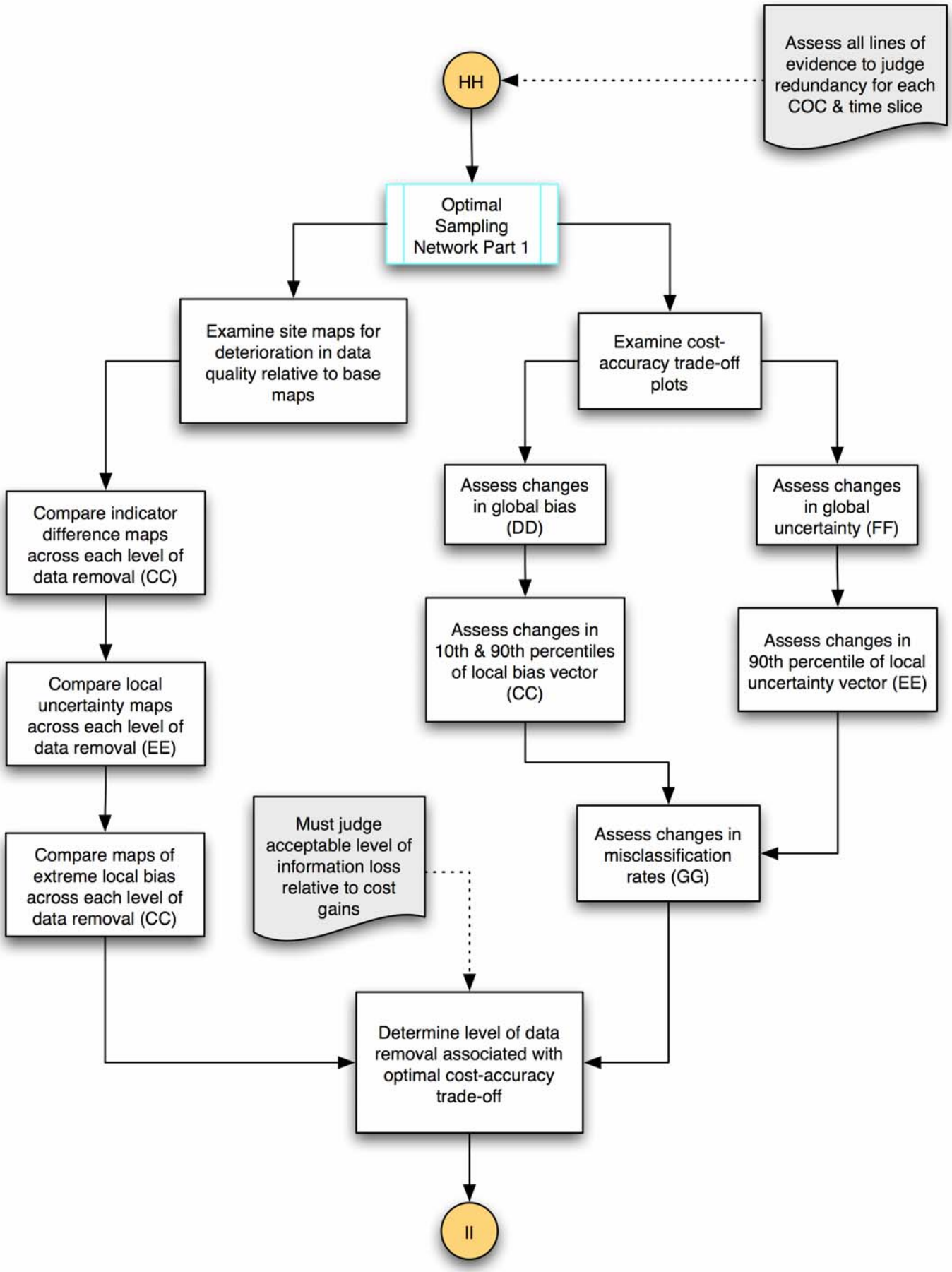

CC
Assess Relative Accuracy Part 1
Determine statistical accuracy of reduced-data set for a given COC & time slice when compared to the corresponding base map
Differences computed on probability scale to allow consistent scale of comparison across COCs, across time slices, or regardless of a highly skewed concentration distribution
Compute node-by-node differences between reduced-data & base map LWQR probability estimates
Each mesh node will have a vector of differences, one element for each indicator level
Results in one average probability difference per node
Average vectors of node-specific probability differences across indicator levels
Visual appraisal of extent of extreme under- or overestimation
Set of node-specific probability differences labeled the local bias vector
Visual appraisal of degree of over- or underestimation on reduced-data map compared to the base map
Plots of extreme low bias & extreme high bias extents vs. cost of reduced-data well network
Create map of local bias vector
Color-contoured indicator difference map (i.e., map of local bias values)
Compute areal or volumetric extent of extreme local biases
Compute lower 10th and upper 90th percentiles of local bias vector
Plots of lower 10th & upper 90th percentiles vs. cost of reduced-data well network
DD

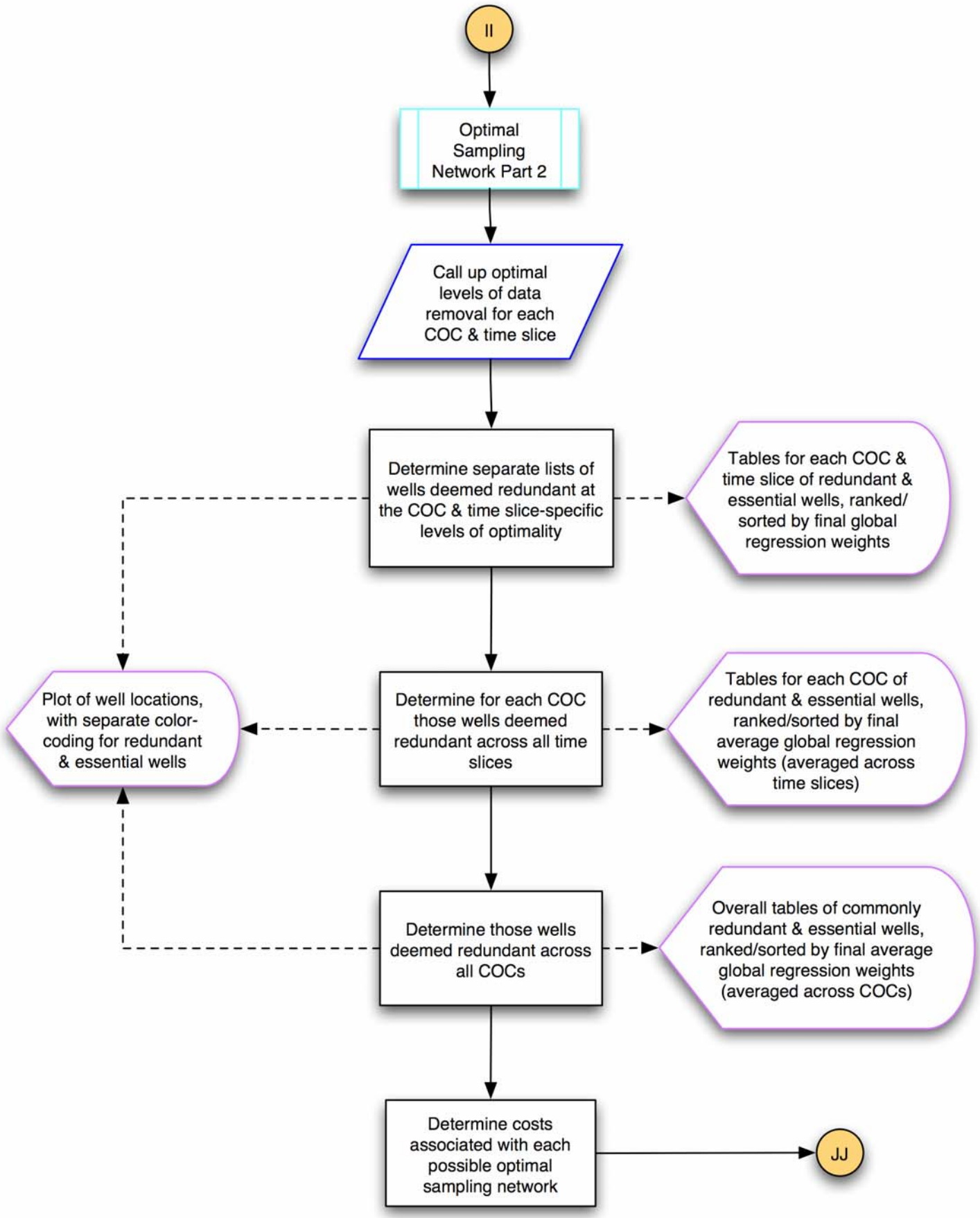

EE
Assess Relative Uncertainty Part 1
Determine statistical uncertainty of reduced-data set for a given COC & time slice when compared to the corresponding base map
Squared differences allow computation of quantity akin to mean squared error
Compute node-by-node squared differences between reduced-data & base map LWQR probability estimates
Each mesh node will have a vector of differences, one element for each indicator level
Results in one average squared difference per node
Average vectors of node-specific squared differences across indicator levels
Set of node-specific squared differences labeled the local uncertainty vector
Visual appraisal of areas of greatest uncertainty on reduced-data map compared to the base map
Create map of local uncertainty vector
Color-contoured local uncertainty map (i.e., map of local uncertainty values)
FF
Compute upper 90th percentile of local uncertainty vector
Plot of upper 90th percentile vs. cost of reduced-data well set

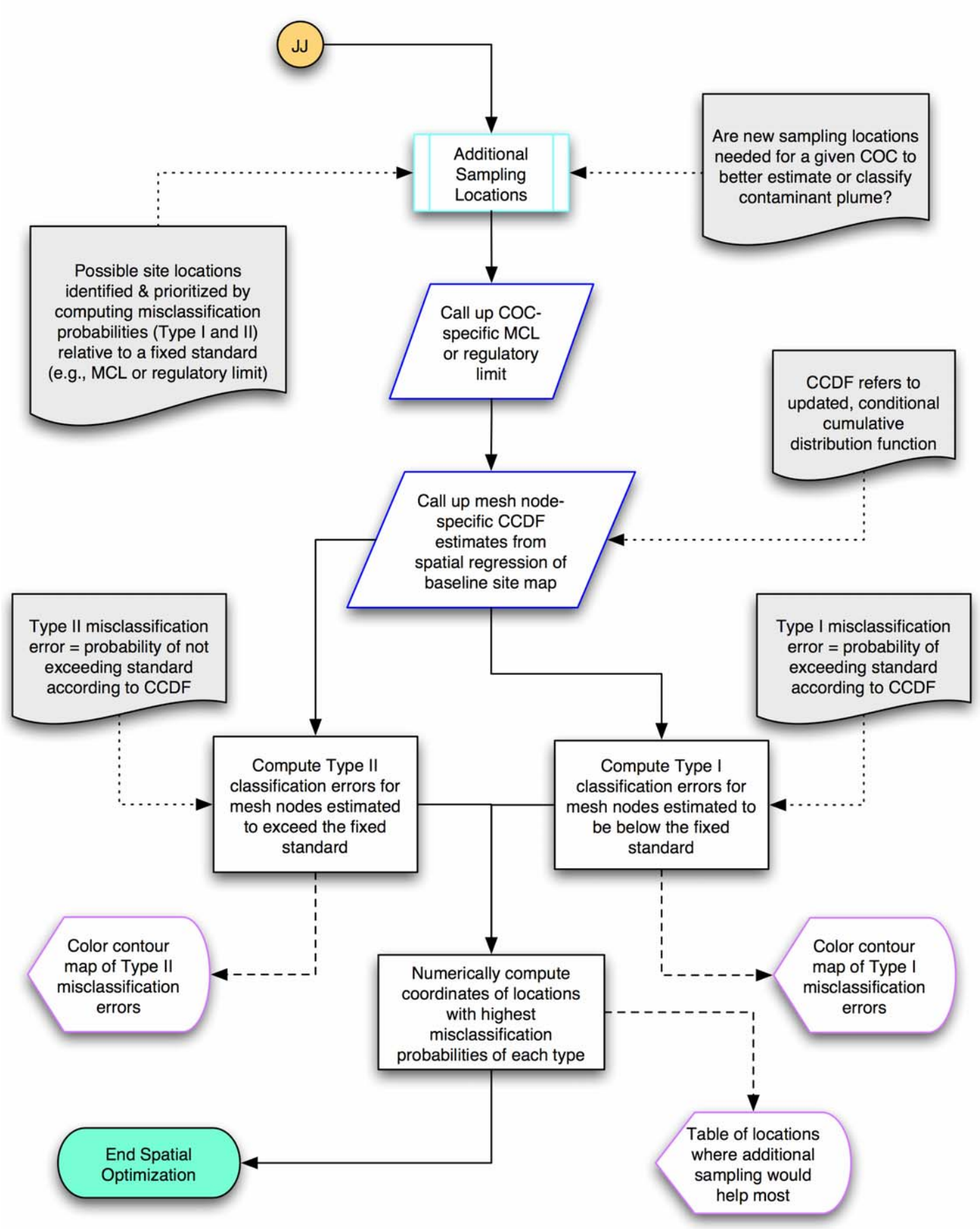

DD
Assess Relative Accuracy Part 2
Average local bias vector to form estimate of global bias
Note that global bias = 0 for base map by definition
Positive global bias represents overestimation of the base map; negative bias corresponds to overall underestimation of base map
Track changes in global bias
Plot of global bias vs. cost of reduced-data well network
EE


FF
Assess Relative Uncertainty Part 2
Global uncertainty = relative mean squared error between reduced-data & base maps
Average local uncertainty vector to form estimate of global uncertainty
Note that global uncertainty = 0 for base map by definition
Provides a numerical summary of total relative statistical uncertainty associated with given well network configuration
Track changes in global uncertainty
Plot of global uncertainty vs. cost of reduced-data well network
Global uncertainty necessarily increases as reduced-data map deviates more from base map
GG


GG
Misclassification Rates
Track the fraction of mesh nodes classified as one way relative to limit on base map but the opposite way on the reduced-data map
Call up COC-specific MCL or regulatory limit
Call up base map concentration estimates, including mean, trimmed mean, & median
Call up reduced-data map concentration estimates, including mean, trimmed mean, & median
Compute fraction of mesh nodes misclassified as exceedances relative to base map classifications
Plots of over-classification rates vs. cost of reduced-data well network
To guard against highly skewed univariate concentration distributions, compute misclassification rates for mean, trimmed mean, & median estimates separately
Compute fraction of mesh nodes misclassified as non-exceedances relative to base map classifications
Plots of under-classification rates vs. cost of reduced-data well network
End Assessment of Spatial Accuracy & Uncertainty


HH
Assess all lines of evidence to judge redundancy for each COC & time slice
Optimal Sampling Network Part 1
Examine site maps for deterioration in data quality relative to base maps
Examine cost-accuracy trade-off plots
Compare indicator difference maps across each level of data removal (CC)
Assess changes in global bias (DD)
Assess changes in global uncertainty (FF)
Compare local uncertainty maps across each level of data removal (EE)
Assess changes in 10th & 90th percentiles of local bias vector (CC)
Assess changes in 90th percentile of local uncertainty vector (EE)
Compare maps of extreme local bias across each level of data removal (CC)
Must judge acceptable level of information loss relative to cost gains
Assess changes in misclassification rates (GG)
Determine level of data removal associated with optimal cost-accuracy trade-off
II


II
Optimal Sampling Network Part 2
Call up optimal levels of data removal for each COC & time slice
Determine separate lists of wells deemed redundant at the COC & time slice-specific levels of optimality
Tables for each COC & time slice of redundant & essential wells, ranked/ sorted by final global regression weights
Plot of well locations, with separate color-coding for redundant & essential wells
Determine for each COC those wells deemed redundant across all time slices
Tables for each COC of redundant & essential wells, ranked/sorted by final average global regression weights (averaged across time slices)
Determine those wells deemed redundant across all COCs
Overall tables of commonly redundant & essential wells, ranked/sorted by final average global regression weights (averaged across COCs)
Determine costs associated with each possible optimal sampling network
JJ


JJ
Additional Sampling Locations
Are new sampling locations needed for a given COC to better estimate or classify contaminant plume?
Possible site locations identified & prioritized by computing misclassification probabilities (Type I and II) relative to a fixed standard (e.g., MCL or regulatory limit)
Call up COC-specific MCL or regulatory limit
CCDF refers to updated, conditional cumulative distribution function
Call up mesh node-specific CCDF estimates from spatial regression of baseline site map
Type II misclassification error = probability of not exceeding standard according to CCDF
Type I misclassification error = probability of exceeding standard according to CCDF
Compute Type II classification errors for mesh nodes estimated to exceed the fixed standard
Compute Type I classification errors for mesh nodes estimated to be below the fixed standard
Color contour map of Type II misclassification errors
Numerically compute coordinates of locations with highest misclassification probabilities of each type
Color contour map of Type I misclassification errors
End Spatial Optimization
Table of locations where additional sampling would help most