APPENDIX B

Equations Used for the Probabilistic Analysis of Metal Loadings

CONTENTS

ABBREVIATIONS AND ACRONYMS v

B.1.0 INTRODUCTION 1

B.2.0 DETERMINISTIC FORMULATION 3
- B.2.1 FUTURE OR POST-REMEDIATION LOADING, $F_X(t)$ 3
- B.2.2 REMEDIATION FACTOR, R(t) 5
  - B.2.2.1 Remediation Factor at Time Zero (t=0), R_0 6
  - B.2.2.2 Source Depletion (Decay) Factor and Source Depletion (Decay) Rate 15
- B.2.3 LOAD RATIO, Lr 24
  - B.2.3.1 Approximate Load Ratio Scaling Factors 26
  - B.2.3.2 Gross Allocation Loads 28
  - B.2.3.3 Background Loads and Concentrations 28
- B.2.4 POST-REMEDIATION CONCENTRATIONS, C"(t) 29

B.3.0 PROBABILISTIC FORMULATION 31
- B.3.1 PRE-REMEDIATION LOADING, L 31
- B.3.2 POST-REMEDIATION LOADING, F(t) 34
- B.3.3 REMEDIATION FACTOR, R(t) 37
  - B.3.3.1 Remediation Factor at Time Zero (t=0), R_0 39
  - B.3.3.2 Source Depletion Natural Recovery Rate and Factor 45
  - B.3.3.3 Approximations for E[R(t)] and CV[R(t)] 48
- B.3.4 LOAD RATIO, Lr 50
  - B.3.4.1 Loading Capacity, C_L 52
  - B.3.4.2 Probability True Load Ratio Does Not Exceed a Multiple of the Loading Capacity or AWQC 55
  - B.3.4.3 Load Ratio Estimate Required to Achieve a Target True Load Ratio With Specified Probability 55
  - B.3.4.4 Remediation Factor Required to Achieve a Target True Load Ratio With Specified Probability 56
- B.3.5 COMPOSITE REMEDIATION AND SCALING FACTORS FOR COMBINED SOURCES 62
  - B.3.5.1 Combined Loads: Pinehurst to Harrison 64
  - B.3.5.2 Combined Loads That Include the BHSS 65
- B.3.6 LOADING DUE TO MULTIPLE SOURCES AND "WHAT-IF" SCENARIO ANALYSES 70
  - B.3.6.1 Special Cases 73

CONTENTS (Continued)

B.3.6.2 Remediation Factor Estimated from Pre- and Post-remediation Loadings 74
B.3.6.3 What-if Scenario Analyses 76
B.3.7 TIME INTEGRALS 79

B.4.0 CONCLUSIONS 81

B.5.0 REFERENCES 83

FIGURES

B-1 Loading Reduction Over Time 85
B-2 Loading Reduction Over Time with Effect from Uncertainty in B 86

ABBREVIATIONS AND ACRONYMS

AWQC	Ambient Water Quality Criteria
BHSS	Bunker Hill Superfund Site
CDF	cumulative density function
CDR	Coeur d'Alene River
cfs	cubic foot per second
cy	cubic yard
EPA	U.S. Environmental Protection Agency
Eq	equation
FP	floodplain sediment
FS	feasibility study
Ft^3	cubic foot
LB	lower bound
lb/d	pound per day
MS	Microsoft
M/T	metal mass per unit time
M/V	metal mass per volume of water
µg/L	microgram per liter
mg/kg	milligram per kilogram
NFCDR	North Fork Coeur d'Alene River
PAT	probabilistic analysis tool
PDF	probability density function
RAA	remedial action area
RI	remedial investigation
RI/FS	remedial investigation/feasibility study
RLP	relative loading potential
SFCDR	South Fork Coeur d'Alene River
TMDL	total maximum daily load
ton/cy	tons per cubic yard
UB	upper bound
USGS	United States Geological Survey
V/T	volume per unit time
yr	year

B.1.0 INTRODUCTION

As introduced in the main body of this technical memorandum, this appendix documents and details the mathematical equations and supporting hypotheses, assumptions, and approximations used in the probabilistic analysis of metal loadings in the Coeur d'Alene River Basin. As included in this appendix, the probabilistic analysis consists of two major parts:

- The analysis of the remedial alternatives developed in the draft feasibility study (FS) report (USEPA 2000b)
- The "what-if" scenario analysis of hypothetical loadings, as discussed in this memorandum

The analysis provides quantitative estimates of potential post-remediation metal loadings, concentrations, and associated measures at selected stream monitoring locations in the basin. Although the analysis was initially developed to help evaluate the FS alternatives, it can also be used as a tool to help select a remedy. Estimates are made for various times, t, following remedial action and include the following quantities:

- Remediation factors, R(t), the proportion of pre-remediation zinc loading or concentration following remediation
- Zinc loadings, F(t)
- Loading ratios, Lr(t), the ratio of zinc loading to total maximum daily load (TMDL) loading capacity and zinc concentration to Ambient Water Quality Criteria (AWQC)
- Probability of meeting load ratios, including Lr=1, which means meeting AWQC for dissolved zinc

The two parts of the analysis have been implemented in two complementary MS Excel 97 spreadsheets. The spreadsheets are titled:

- "Probabilistic Analysis Tool 1" (PAT1)
- "Probabilistic Analysis Tool 2" (PAT2)

PAT1 analyzes the six FS alternatives based on the source-by-source remedial actions documented in the draft FS (USEPA 2000b). PAT1 is presently limited to making post-remediation estimates for the South Fork Coeur d'Alene River (SFCDR) at Pinehurst (USGS station SF271) and the Coeur d'Alene River (CDR) at Harrison (USGS station LC60).

PAT2 is a "what-if" scenario analysis that allows users to input hypothetical or estimated remediation factors for specific areas of the basin, including major tributaries and reaches of the SFCDR (upper basin) and CDR (lower basin). As output, PAT2 makes post-remediation estimates for the mouth of Canyon Creek (USGS station CC288), Ninemile Creek (USGS station NM305), Pine Creek (USGS station PC305), the SFCDR at Pinehurst (USGS station SF271), and the CDR at Harrison (USGS station LC60).

The remainder of this appendix documents the analysis implemented in PAT1 and PAT2. The basic deterministic relationships are presented in Section B.2. The probabilistic formulation of the deterministic relationships is documented in Section B.3. The probabilistic formulation provides the basis for quantifying the uncertainty associated with the estimates. Appendix A discusses uncertainties in more general terms and provides the mathematical basis for the specific equations documented here and used in the analysis. Appendix C presents the principal input values used in the analysis. Principal results are presented and discussed in Section 4 of the main text of the technical memorandum.

B.2.0 DETERMINISTIC FORMULATION

This section summarizes the deterministic relationships for estimating metal mass loading, concentrations, and associated measures for a given flowing water body. Although the relationships are used in the analysis for surface water, they would also generally apply to groundwater. The deterministic relationships are the fundamental basis for the probabilistic formulation, as documented in Section B.3.

The deterministic relationships do not explicitly account for uncertainty in the various parameters and quantities on which they are based. They result in single-valued "point estimates," which would be considered adequate for analysis in the absence of significant uncertainty, including natural variability. However, because the post-remediation performance of any alternative, including no action, is unavoidably uncertain to some extent, the deterministic relationships have been formulated probabilistically to explicitly deal with that uncertainty. The probabilistic formulation documented in Section B.3 is based on the deterministic relationships documented in Section B.2.

Consistent units for the equations are tacitly assumed in all equations presented in this document. Metal mass loadings in streams (or adits) are in dimensions of metal mass per unit time (M/T), and typically expressed in units of pounds of metal per day (lb/d). Concentrations are in dimensions of metal mass per volume of water (M/V), and typically expressed in units of micrograms per liter (µg/L). Stream (or adit) discharges are in dimensions of water volume per unit time (V/T), and typically expressed in units of cubic feet per second (cfs). Loadings in units of lb/d calculated as the product of concentrations in units of µg/L and discharge in units of cfs includes a constant units conversion factor of 0.0058 lb/d per ft^3 µg/sec L. Solid source materials are typically expressed in volume units of cubic yards (cy), mass units of tons, unit weights (or densities) of tons per cubic yard (ton/cy), and metal concentration units of milligrams of metal per kilogram of solid (mg/kg).

B.2.1 FUTURE OR POST-REMEDIATION LOADING, $F_X(t)$

Post-remediation loadings are estimated from pre-remediation (historical) conditions using a remediation factor, as discussed in Section 2. Both pre-remediation and post-remediation loadings, and the remediation factor, are estimated at or between given stream monitoring locations, symbolized by "x." Pre-remediation loadings are symbolized L_x; the remediation factor and post-remediation loading are symbolized R_x and F_x, respectively. Both R_x and F_x are functions of time "t" after remedy implementation, including any lag time. The formal relationship is as follows:

[1] $F_x(t) = R_x(t) \cdot L_x$
$= R_x(t) \cdot Q_x \cdot C_x$

where:

L_x	=	Pre-remediation (historical) metal loading at location x
	=	$Q_x \cdot C_x$, the product of pre-remediation discharge (Q) and concentration (C) measured at location x
$R_x(t)$	=	Remediation factor, ratio of post-remediation to pre-remediation loading
	=	Portion of L not eliminated by remediation, at location x and time t after remedial action
t	=	Time after remedial action is complete, measured in one-water-year increments

Because metal mass loading is the product of metal concentration and discharge, at any time and location, post-remediation loading, $F_x(t)$, is also the product of the post-remediation metal concentration, C_x"(t), and discharge $Q_x(t)$; that is: $F_x(t)=C_x$"$(t) \cdot Q_x(t)$. This relationship is implicit in Eq 1, and the definition of the remediation factor R as the ratio of post-remediation loading, $F_x(t)$, to pre-remediation loading, L_x. Therefore, and as will be further discussed in Section B.2.4, remediation factor R is also the ratio of post-remediation concentration C_x"(t) to pre-remediation concentration C_x.

Pre-remediation loading, L_x, is estimated using the statistical analysis of historical data documented in the remedial investigation (RI) report (USEPA 2000a, Section 5 of Parts 2 through 6), and summarized in the draft FS (USEPA 2000b, Section 1.5.4). Temporal loading variations reflecting the natural variability (including daily, seasonal, and year-to-year) inherent in the historical data are implicit in the estimates of L_x. The potential loading variability within a given water year is thus included in L_x. This variability is assumed constant over time after remediation (except for correlation effects with R(t), as discussed in Section B.3). Therefore, L_x is *not* a function of time after remediation ($t \geq 0$). The same arguments and conclusions hold for C_x and Q_x.

In contrast to L_x, remediation factor $R_x(t)$ is a function of time, starting at time t=0, and averaged over one-water-year increments to account for daily and seasonal variations. $R_x(t)$ thus represent one-year averages over each water year (one water year is the minimum time increment for time t). In practical terms, $R_x(t)$ measures the expected yearly average reduction in L_x over time due to remediation. The total variability in $F_x(t)$, as the product of L_x and $R_x(t)$, thus includes the natural temporal variability inherent in the historical data used to estimate L_x and the uncertainty in the *prediction* of $R_x(t)$.

Remediation factors are estimated as discussed in the following section. To help reduce complexity in the symbology, location x is implicit in all the following equations.

B.2.2 REMEDIATION FACTOR, R(t)

The remediation factor, R(t), is itself composed of two factors: an initial factor, R_0, that applies immediately after remedial action (and "lag period") at time zero, t=0, and a time-dependent source depletion factor, $e^{-\$t}$, that decreases over time. The formal relationship is as follows, where R(t), R_0, and **$** are all for location x:

[1] $R(t) = R_0 e^{-\$t}$

where:

R_0	=	R after remedial action (and "lag period") at time zero, t=0
$e^{-\$t}$	=	Source depletion factor
$	=	Source decay rate, the natural recovery rate due to source depletion
	=	$\ln 2/t_{1/2}$, where $t_{1/2}$ is the source depletion "half-life"

Substituting $R(t)=R_0e^{-\$t}$ into F(t)=R(t)L yields the following expressions for F(t), where F_0 is F(t) at time t=0. That is:

[2] $F(t) = R_0 L e^{-\$ t}$
$= F_0 e^{-\$ t}$

where:

$F_0 = R_0 L$

Estimates of R assume that metal loading in a stream is proportional to the total mass of metal in upstream source material. The proportionality is estimated based on the total relative loading volume of source material. These estimates are based on pre-remediation conditions (time t<0), conditions immediately after remedy implementation (t=0), and thereafter (t>0), as documented in the following Section B.2.2.1. The source depletion factor $e^{-\$ t}$ is developed in detail in Section B.2.2.2.

B.2.2.1 Remediation Factor at Time Zero (t=0), R_0

The remediation factor at time t=0, R_0, is the ratio of post-remediation loading F_0 to pre-remediation loading L; that is: $R_0=F_0/L$. R_0 can be conceptually defined as follows for a total of S individual sources of pre-remediation loading, L_s (s=1,S).[1] Post-remediation loading at each source is the product L_sR_s where R_s represents the net effect of remedial action in terms of source load reduction at t=0. Depending on the source-specific remedial action, R_s can range from unity (no reduction in loading—generally representing no action) to zero (100 percent reduction in loading, or completely effective remedial action). R_0 is then the ratio of the sum of L_sR_s to the sum of L_s where the sums are over all sources; that is: $R_0=\mathbf{E}_{s=1,S}L_sR_s / \mathbf{E}_{s=1,S}L_s$. This theoretically correct conceptual definition of R_0 was approximated as documented in the following development.

In practical terms, R_0 was estimated as the ratio of post-remediation relative loading volume to pre-remediation relative loading volume. Relative loading volumes are based on the source types discussed in Section 2.4.1 of the main text and the concept of relative loading potentials (RLPs). RLPs are discussed in Section 2 and further developed in the following discussion. The relative loading volumes are estimated from RLPs of source types compared to a reference RLP of one for the most severely impacted floodplain sediments. Formally, the relationship is as follows:

$$[1] \quad R_0 = \frac{\text{Post-remediation Relative Loading Volume (RLV")}}{\text{Pre-remediation Relative Loading Volume (RLV')}}$$

$$= \frac{\mathbf{E}_{j=1,K}\, RLP_j * \{\mathbf{E}_{i=1,Nj}\, V_{i,j}R_{i,j}\}}{\mathbf{E}_{j=1,K}\, RLP_j * \{\mathbf{E}_{i=1,Nj}\, V_{i,j}\}} = \frac{\mathbf{E}(RLP*\mathbf{E}VR)}{\mathbf{E}(RLP*V)}$$

where:

index "j" = Source (waste) types: impacted flood plain sediments (FP), tailings, waste rock, adit discharges, and so on for a total of K source types (j=1,K)

[1] Note that s=1,S is shorthand notation for s=1,2,…,S; where s=1 is the first source, s=2 is the second source, and so on, with s=S being the last source. The order of the sources is not important. Similar shorthand notation is used throughout this document to represent individual elements in a sum or product.

index "i"	=	Remedial actions for each source type; e.g., removal/repository, regrading/capping, hydraulic isolation, etc., including "no action," for a total of N_j remedial actions for source type "j" (j=1,Nj)
RLP_j	=	Relative Loading Potential for source type "j" (loading relative to impacted floodplain sediments having an RLP=1)
$V_{i,j}$	=	Remediated Volume of source type "j" being remediated by Remedial Action "i" including no action
$R_{i,j}$	=	Remediation Factor for Remedial Action "i" and source type "j"
	=	Post-remediation loading divided by pre-remediation loading from remediated volume $V_{i,j}$
	=	Complement of effectiveness $E_{i,j}$; $E_{i,j} + R_{i,j} = 1$
RLV"	=	Σ(RLP*ΣVR) summed over all source types, remedial actions, and remediated volumes (including no action), as defined by $\Sigma_{j=1,K} RLP_j * \{\Sigma_{i=1,Nj} V_{i,j} R_{i,j}\}$
RLV'	=	Σ(RLP*V) summed over all source types and total source volumes, as defined by $\Sigma_{j=1,K} RLP_j * \{\Sigma_{i=1,Nj} V_{i,j}\}$ and equal to $\Sigma_{j=1,K} RLP_j * V_j$, where $\Sigma_{i=1,Nj} V_{i,j} = V_j$, the total source type volume

The pre-remediation relative loading volume, RLV'=Σ(RLP*V), is the pre-remediation equivalent total source volume of all source types in terms of severely impacted floodplain sediment having an RLP of 1.0. The pre-remediation relative loading volume, RLV', is therefore also referred to as the "equivalent total source volume." Similarly, the post-remediation relative loading volume, RLV"=Σ(RLP*ΣVR), is the post-remediation equivalent total source volume of all source types in terms of severely impacted floodplain sediment having an RLP of 1.0. The RLP concept is further developed after the next section, which presents sensitivity relationships based on Eq 1.

Sensitivity Relationships

First-order sensitivity relationships, based on Eq 1, that quantify the change in R_0 (ΔR_0) for small changes in RLP_j, $V_{i,j}$, and $R_{i,j}$ (ΔRLP_j, $\Delta V_{i,j}$, and $\Delta R_{i,j}$)[2] are as follows, starting with $\Delta R_0/\Delta RLP_j$. The change in remediation factor R_0 with changes in the (value of the) relative loading potential of source type j, RLP_j is approximated as follows:

[2] Δ means "small change" in the variable.

[2a] $$\partial R_0/\partial RLP_j = \partial\{\Sigma_{j=1,K}\, RLP_j\{\Sigma_{i=1,Nj}\, V_{i,j}R_{i,j}\}[\Sigma_{j=1,K}\, RLP_j\{\Sigma_{i=1,Nj}\, V_{i,j}\}]^{-1}\}/\partial RLP_j$$

$$= \partial\{\Sigma_{j=1,K}\, RLP_j\{\Sigma_{i=1,Nj}\, V_{i,j}R_{i,j}\}\}/\partial RLP_j\}\ [\Sigma_{j=1,K}\, RLP_j\{\Sigma_{i=1,Nj}\, V_{i,j}\}]^{-1} + \partial\{[\Sigma_{j=1,K}\, RLP_j\{\Sigma_{i=1,Nj}\, V_{i,j}\}]^{-1}\}/\partial RLP_j\ [\Sigma_{j=1,K}\, RLP_j\{\Sigma_{i=1,Nj}\, V_{i,j}R_{i,j}\}]$$

$$= \{\Sigma_{i=1,Nj}\, V_{i,j}R_{i,j} - R_0\Sigma_{i=1,Nj}\, V_{i,j}\}[\Sigma_{j=1,K}\, RLP_j\{\Sigma_{i=1,Nj}\, V_{i,j}\}]^{-1}$$

$$= \{R_{Vj} - R_0V_j\} \,/\, \Sigma(RLP*V)$$

where R_{Vj} is the "remediated volume" of source type j, equal to $\Sigma_{i=1,Nj}V_{i,j}R_{i,j}$. The change in remediation factor R_0 with changes in remediated volume $V_{i,j}$ for a given RLP_j and remedial action $R_{i,j}$[3] is approximated as follows:

[2b] $$\partial R_0/\partial V_{i,j} = \partial\{\Sigma_{j=1,K}\, RLP_j\{\Sigma_{i=1,Nj}\, V_{i,j}R_{i,j}\}[\Sigma_{j=1,K}\, RLP_j\{\Sigma_{i=1,Nj}\, V_{i,j}\}]^{-1}\}/\partial V_{i,j}$$

$$= \partial\{\Sigma_{j=1,K}\, RLP_j\{\Sigma_{i=1,Nj}\, V_{i,j}R_{i,j}\}\}/\partial V_{i,j}\}\ [\Sigma_{j=1,K}\, RLP_j\{\Sigma_{i=1,Nj}\, V_{i,j}\}]^{-1} + \partial\{[\Sigma_{j=1,K}\, RLP_j\{\Sigma_{i=1,Nj}\, V_{i,j}\}]^{-1}\}/\partial V_{i,j}\ [\Sigma_{j=1,K}\, RLP_j\{\Sigma_{i=1,Nj}\, V_{i,j}R_{i,j}\}]$$

$$= RLP_j\,(R_{i,j} - R_0) \,/\, \Sigma(RLP*V)$$

Recall that loading reductions are assumed proportional to the volume remediated for a given source type. Eq 2b quantifies how overall remedial effectiveness, measured by R_0, changes with the volume remediated, $V_{i,j}$, and depend on the source type, RLP_j, and the difference between remedial action $R_{i,j}$ and R_0, $(R_{i,j} - R_0)$; that is:

$$\partial R_0 = \partial V_{i,j}\ \{RLP_j\,(R_{i,j} - R_0) \,/\, \Sigma(RLP*V)\}$$

The change in remediation factor R_0 with changes in the effectiveness of the remedial action, measured by $R_{i,j}$, on volume $V_{i,j}$ for a given RLP_j is approximated as follows:

[2c] $$\partial R_0/\partial R_{i,j} = \partial\{\Sigma_{j=1,K}\, RLP_j\{\Sigma_{i=1,Nj}\, V_{i,j}R_{i,j}\}[\Sigma_{j=1,K}\, RLP_j\{\Sigma_{i=1,Nj}\, V_{i,j}\}]^{-1}\}/\partial R_{i,j}$$

[3] Recall that $R_{i,j}$ is the remediation factor (the complement of the effectiveness) of the remedial action applied to volume $V_{i,j}$.

$$= \quad M\{\Sigma_{j=1,K}\, RLP_j\{\Sigma_{i=1,Nj}\, V_{i,j}R_{i,j}\}\}/MR_{i,j}\}\ [\Sigma_{j=1,K}\, RLP_j\{\Sigma_{i=1,Nj}\, V_{i,j}\}]^{-1}$$
$$+\ M\{[\Sigma_{j=1,K}\, RLP_j\{\Sigma_{i=1,Nj}\, V_{i,j}\}]^{-1}\}/MR_{i,j}\ [\Sigma_{j=1,K}\, RLP_j\{\Sigma_{i=1,Nj}\, V_{i,j}R_{i,j}\}]$$

$$= \quad RLP_j\, V_{i,j} \,/\, \Sigma(RLP*V)$$

Note:

[2d] $ME_{i,j}/MR_{i,j} = M(1 - R_{i,j})/MR_{i,j})$
$= -1$

The following section elaborates on RLPs.

Relative Loading Potentials, RLPs

RLPs (relative loading potentials) measure the relative load contribution from a given source type at a particular downstream monitoring location. RLPs provide a practical way to estimate current, pre-remediation loads from the various source types and are consistent with estimates of post-remediation loads. The RLP for each source type represents a temporal and volumetric average over site-specific sources that are upstream of the monitoring location. RLPs thus generally vary with location, consistent with the monitoring location. The following discussion provides a rigorous context for the concept of RLPs and their practical estimation.

To begin with, RLPs are provisionally defined for any of K source types using index "j" (j=1,K); that is:

[3] $RLP_j = \varepsilon_j\, L_j / V_j$

L_j is the pre-remediation (i.e., "current") load contribution from source type j at a given stream monitoring location (x); V_j is the total (upstream) volume of source type j that results in the loading L_j. Quantity ε_j is a proportionality factor, which will be discussed shortly. The sum of L_j for all K source types is L, the current, pre-remediation load at the stream monitoring location; that is: $\Sigma_{j=1,K}L_j = L$. L is estimated from analysis of available historical stream monitoring data, as presented in the RI and summarized in the FS (USEPA 2000a and 2000b).

For any source type j, load L_j is the net loading from all upstream discrete or "site-specific" instances of source type j. This makes L_j a spatial sum, which, for a total of "S" discrete sites of source type j, is formalized as $\Sigma_{s=1,S}L_{js} = L_j$, where L_{js} is the loading from site "s." The volumes

of these sources, V_{js}, sum to V_j; that is: $\mathbf{E}_{s=1,S} V_{js} = V_j$. The ratio L_j/V_j , which equates to RLP_j, is thus given by the following:

$$L_j/V_j = \quad {}_{s=1,S} L_{js} / \quad {}_{s=1,S} V_{js}$$

which equals the harmonic average of the S individual ratios L_{js}/V_{js}. That is, RLP_j is the harmonic average of the S individual ratios L_{js}/V_{js}:

$$\begin{aligned} L_j/V_j &= S / \quad {}_{s=1,S}(L_{js}/V_{js})^{-1} \\ &= \quad {}_{s=1,S} L_{js} / \quad {}_{s=1,S} V_{js} \\ &= RLP_j \end{aligned}$$

The loadings L_{js}, and their sum L_j, are averaged over a water year and thus represent temporal averages. By averaging over a water year, typical seasonal effects and variations are averaged or "integrated-out" of loadings L_{js} and L_j, which are time-dependent over the water year. The time-dependence of loadings are symbolized $L_{js}(\mathbf{J})$ and $L_j(\mathbf{J})$, where $\mathbf{J}$ represents time over the water year ($\mathbf{J}$ should not to be confused with time t, which represents time after remedial action). $L_{js}(\mathbf{J})$ and $L_j(\mathbf{J})$ follow from the corresponding time-dependence of discharge, $Q_{js}(\mathbf{J})$ and $Q_j(\mathbf{J})$, and concentration, $C_{js}(\mathbf{J})$ and $C_j(\mathbf{J})$. These relationships are formalized as follows for a discrete number of individual sources S:

$$\begin{aligned} L_j &= yr^{-1} \quad {}_{\text{water year}} L_j(\mathbf{J})\, d\mathbf{J} &= yr^{-1} \quad {}_{s=1,S} \quad {}_{\text{water year}} L_{js}(\mathbf{J})\, d\mathbf{J} \\ &= yr^{-1} \quad {}_{\text{water year}} Q_j(\mathbf{J}) C_j(\mathbf{J})\, d\mathbf{J} &= yr^{-1} \quad {}_{s=1,S} \quad {}_{\text{water year}} Q_{js}(\mathbf{J}) C_{js}(\mathbf{J})\, d\mathbf{J} \end{aligned}$$

The choice, or definition, of site-specific discrete sources is dependent on geographic scale. That is, each discrete source s is itself composed of multiple "sub-sources" that compose its spatial volume, represented by the local spatial coordinates x, y, and z. A full definition of the load from source s would thus require appropriate volume integrals, represented as follows, where

$L_{js}(\mathbf{J},x,y,z)$ is the loading from point x,y,z in source s at time **J**, reflecting the corresponding discharge and concentrations, $Q_{js}(\mathbf{J},x,y,z)$ and $C_{js}(\mathbf{J},x,y,z)$, respectively; that is:

$$L_{js} = yr^{-1} \quad {}_{\text{Volume of s}} \quad {}_{\text{water year}} L_{js}(\mathbf{J},x,y,z)\, d\mathbf{J}\, dx\, dy\, dz$$

$$= yr^{-1} \quad {}_{\text{Volume of s}} \quad {}_{\text{water year}} Q_{js}(\mathbf{J},x,y,z) C_{js}(\mathbf{J},x,y,z)\, d\mathbf{J}\, dx\, dy\, dz$$

The discharges, concentrations, and loadings from each site-specific source are generally composed of multiple metal transport pathways from the source to surface water. These metal transport pathways are primarily associated with erosion via surface water runoff and leaching via infiltration, percolation, and groundwater flow. Leaching may generally include effects from percolation of infiltrated surface water, groundwater table fluctuations, and saturated groundwater flow. For any particular source s, these effects may be formalized as follows (implicitly considering spatial [x,y,z] effects over the source volume)

$$L_{js}(\mathbf{J}) = L_{js}(\mathbf{J})_E + L_{js}(\mathbf{J})_P + L_{js}(\mathbf{J})_{GWT} + L_{js}(\mathbf{J})_{GW}$$

where $L_{js}(\mathbf{J})_E$ signifies loading primarily associated with erosion; $L_{js}(\mathbf{J})_P$ signifies loading primarily associated with leaching to unsaturated groundwater flow from percolation of infiltrated surface water; $L_{js}(\mathbf{J})_{GWT}$ signifies loading from leaching primarily associated with groundwater table fluctuations; and $L_{js}(\mathbf{J})_{GW}$ signifies loading primarily associated with leaching from saturated groundwater flow below the groundwater table (phreatic surface). Besides spatial effects, these **J**-dependent factors would also generally experience year-to-year temporal variations between water years. This inter-year variability would likely be strongly correlated between the different sources and source types (affecting each in generally similar ways) and average out over the long-term, assuming no long-term systematic temporal trends.

The forgoing discussion briefly suggests the true complexity of loading. In theory, it would be possible to rigorously estimate loadings based on source-by-source characterization and modeling. The modeling could include solutions to advection-diffusion/dispersion equations that included appropriate geochemical interactions, erosion effects, and boundary effects.[4] Such an

[4] Such modeling could require coupled sets of multi-dimensional partial differential equations representing chemical and physical effects in the vadose zones and saturated zones within the metal transport pathways from sources to surface water monitoring locations. Solutions to these equations would require estimating numerous spatially and temporally variable input parameters.

effort would be beyond the state of the art and require site-specific data collection, analysis, and modeling efforts of extraordinary and unprecedented scope, complexity, time, and cost.

In actual practice, such an effort would be not only highly impracticable but also wastefully excessive and unnecessary for the purpose this analysis. In contrast, within the scope and purpose of the analysis, RLPs provide a practical alternative to estimating current, "pre-remediation" loads from the various sources based on the concept of relative loading potentials. RLPs were estimated in practical terms using available data as follows.

Without loss of generality, RLPs were operationally defined using the most severely affected floodplain sediments (FP) as a reference source type, such that RLP_{FP} would be unity (1.0); that is:

[4] $RLP_j = L_j/V_j \,/\, L_{FP}/V_{FP}$ = 1.0 for j=FP, the reference source type
$= L_j V_{FP} \,/\, L_{FP} V_j$
$= L_j \,/\, L_{FP}$ per unit volume of source type j and FP
$= V_{FP} \,/\, V_j$ per unit load of source type j and FP

Although any source type could be used as the reference source type, the most severely affected floodplain sediments was selected because it is expected to be the source type contributing the maximum load per unit volume, as will be discussed. In terms of impacted floodplain sediment having an RLP of 1.0, the pre-remediation equivalent total source volume of all K source types, or simply the "equivalent total source volume," symbolized RLV', is defined as RLV'= $\mathbf{E}_{j=1,K} RLP_j V_j$.

Because RLV' represents the net source of loading measured by L (at a given monitoring location, x), it follows that the pre-remediation load from source type, L_j, is given by:

[5] $$L_j = \frac{RLP_j V_j}{{}_{j=1,K} RLP_j V_j} L$$

By its existence, Eq 5 shows that pre-remediation loadings from any source type can be estimated from available data. Also, through algebraic substitution of Eq 5 into Eq 3, the proportionality factor from Eq 3, $"_j$, is $\mathbf{E}_{j=1,K} RLP_j V_j L^{-1}$; thus making the provisional and operational definitions of RLP equivalent for any given source type.

Operational estimates of RLPs for the source types are given in Appendix C based on engineering interpretation and synthesis of available information. The RLPs represent estimates of true RLPs, which are uncertain because of natural variability and imperfect information. The RLP estimates were treated probabilistically to formally and rigorously account for this uncertainty, which was characterized using lognormal distributions. Because the estimates represent spatial and temporal averages, temporal and source-by-source variations are implicit in the estimates. One source type requires special consideration, as discussed next.

Deeper Impacted Floodplain Sediments. As discussed in Appendix C, this source type collectively represents potential zinc sources not subject to reduction by any potential remedial action documented in the draft FS. These deeper impacted sediments principally include aquifer materials below the depth of floodplain sediments identified in Table 1-1 (main text of the technical memorandum) that are subject to potential remedial action as well as impacted material within floodplains not otherwise included in the draft FS volume estimates (e.g., road and former railroad embankments). For purpose of the analysis, these deeper sediments also represent potential zinc sources associated with possible groundwater loading to streams through deep bedrock or from deeper parts of the alluvial system that would not be reduced by remedial action.

Historical mining-related zinc that has sorbed, precipitated, or otherwise become contained within these deeper aquifer materials is expected to be a significant zinc source to surface water. This expectation is based on analysis and interpretation of information in the RI and the technical memorandum on background concentrations (USEPA 2001), reports by others (Paulson, Balderrama, and Zahl 1996), and communications with USGS (S. Box, USGS Spokane, Washington).

For the purpose of the analysis, a RLP for these deeper impacted sediments, RLP_{DS}, was estimated as proportional to the volumetric average metal (zinc) concentrations in the deeper impacted sediments (C_{DS}) relative to the volumetric average concentration in the impacted sediments having an RLP=1 (C_S); that is:

[6] $RLP_{DS} = >C_{DS}/C_S$

Factor > depends on details of the C_{DS}/C_S estimate and also generally allows RLP_{DS} to be other than directly proportional to C_{DS}/C_S. Further practical details of the estimate are documented in Appendix C.

An RLP_{DS} estimate that is proportional to C_{DS}/C_S is considered appropriate for impacted floodplain sediments (or the like) because of the general similarity in the source type and exposure conditions to metal transport by leaching and erosion. Although impacted material within floodplains not otherwise included in the FS volume estimates (e.g., road and former railroad embankments) may have higher concentrations on average than the deeper floodplain sediments, they are not as exposed to water transport effects. In the absence of any known contrary information, they were therefore assumed to have, on average, RLPs equal to the deeper sediments. Uncertainty in the estimates is considered in Section B.3.

Adit Loading

For adit loading, the equivalent volume of impacted floodplain volume having RLP=1, V_{adits}, is given as follows, where A_L is the proportion of total load (L) contributed by adits; that is:

[7] $V_{adits} = A_L/(1-A_L)*\Sigma(RLP*V)_{\text{solid source types}}$

A_L = Adit total load / Stream total loading, L

The volume given by Eq 7 results in adits having a load contribution exactly equal to their measured loading.

Total Effective Metal (Zinc) Mass, TEM

The total effective metal (zinc) mass from all source types that causes loading *before* any remediation (t<0), symbolized TEM', is estimated from the equivalent total source volume of impacted floodplain sediments having an RLP of 1.0. TEM' is estimated as follows, where the summation is over all source types:

[8] $TEM' = \Sigma(RLP*V)\,\rho\,C_S$
$= RLV'\,\rho\,C_S$

where:

ρ = Volumetric average unit weight of floodplain sediments
C_S = Volumetric average zinc concentration of floodplain sediments having RLP=1

RLP and V, the relative loading potential and volume, of each source type are the same as used in Eq 1 for estimating R_0. Variable TEM' is thus the product of the pre-remediation relative loading volume, RLV', and the average unit weight and concentration of the floodplain sediments having RLP=1. TEM' is independent of any specific remedial action.

The total effective metal (zinc) mass from all source types that causes loading *after* remediation at time t=0 is also estimated in terms of the equivalent total source volume of impacted floodplain sediments having an RLP of 1.0. Symbolized TEM", this quantity is estimated as follows, where the summation is over all source types and remedial actions and **(** and C_S are as defined for TEM':

$$[9] \quad TEM'' = \mathbf{E}(RLP*\mathbf{E}VR)\,\mathbf{(}C_S = RLV''\,\mathbf{(}C_S$$

In contrast to TEM', TEM" depends on the remedial action. It is also obvious that R_0 is the ratio of the total effective metal mass immediately after (t=0) and before remediation (t<0); that is (noting that **(**C_S cancel in the ratio of TEM"/TEM') :

$$[10] \quad R_0 = \frac{TEM''}{TEM'} = \frac{\mathbf{E}(RLP*\mathbf{E}VR)\,\mathbf{(}C_S}{\mathbf{E}(RLP*V)\,\mathbf{(}C_S} = \frac{RLV''}{RLV'}$$

Note that for no action TEM' and TEM" are equal, making R_0=1. With or without remediation, source depletion will cause the total effective metal mass to decrease over time after remediation (t>0), making TEM" a function of time, TEM"(t). This effect is simulated as documented in the following section.

B.2.2.2 Source Depletion (Decay) Factor and Source Depletion (Decay) Rate

As developed in this section, the source depletion factor, $e^{-\$t}$, follows from the fundamental working hypothesis that loading is proportional to total metal mass.[5] Specifically, post-remediation loading, F(t), is formally related to TEM" as follows, where **$**, the source decay rate, is the proportionality factor; that is:

$$[1] \quad F(t) = \mathbf{\$}\,TEM''(t)$$

[5] A similar relationship would hold for concentrations.

Factor **$** is a spatial and temporal average and is both chemical-specific and location-specific. It is assumed that **$** does not change over time due to changes in geobiochemistry or hydrology. Furthermore, since F(t) is the time rate of change (decrease) of TCM"(t), it follows that the time rate of change of TEM"(t), after remediation ($t \geq 0$), is also proportional to TEM"(t); that is:

[2] $\quad dTEM''(t)/dt = -\mathbf{\$}\, TEM''(t)$

The numerical equality implied by Eqs 1 and 2 neglects the metal mass in groundwater flow not measured at stream location x. However, since the primary concern is with metal loading in the stream, Eqs 1 and 2 are interpreted to represent that portion of TEM"(t) that loads the stream at or along location x (this interpretation can be refined to explicitly consider the groundwater metal mass at x). The equality of loading F(t) and the time rate of decrease of TEM"(t) is formalized as follows:

[3] $\quad F(t) = -dTEM''(t)/dt$

The proportionality factor, **$** (which is also the source decay rate), is averaged over a water year to account for daily and seasonal variations in the proportionality factor. One year is therefore the minimum time increment (dt) for all time-dependent equations. By averaging over a water year, typical seasonal flow effects on concentrations and loadings are averaged or "integrated-out" of **$**, as well as R(t). This integration is formally represented as follows where, as before, **J** represents time over the water year (and should not to be confused with time t, which represents time after remedial action); that is:

$$\mathbf{\$} = yr^{-1} \int_{\text{water year}} \mathbf{\$}(\mathbf{J})\, d\mathbf{J}$$

Without this water-year averaging **$** would be a function of discharge, Q(**J**), and potentially other **J**-dependent factors, which would make the analysis highly but needlessly complex for practical purposes. Factor **$** is derived by first solving Eq 2 for TEM"(t) as follows:

[4] $\quad dTEM''(t)/dt = -\mathbf{\$}\, TEM''(t)$

$$dTEM''(t)/TEM''(t) = -\mathbf{\$}\, dt$$
$$\ln\{TEM''(t)\} = -\mathbf{\$}\, t + k$$

$$TEM''(t) = e^{-\$t + k}$$
$$TEM''(t) = TEM''(t=0)\, e^{-\$t}$$
$$= TEM''\, e^{-\$t} \quad \text{(see footnote}^{6}\text{)}$$

[4a] $$TEM''(t) = TEM''\, e^{-\$t}$$

Substituting TEM"(t) from Eq 4a into Eq 3 and carrying out the differentiation yields the following for F(t):

[5] $$F(t) = -\, dTEM''e^{-\$t}/dt$$
$$= \$\, TEM''e^{-\$t}$$

Solving Eq 5 for F(t) at time t=0 yields F_0 :

[6] $$F_0 = \$\, TEM_0''$$

Substituting Eq 6 into Eq 5 yields F(t), which is identical to B.2.2 Eq 2 (where $F_0=R_0L$); that is:

[7] $$F(t) = F_0\, e^{-\$t}$$

Eqs 5 and 7 are mathematically equivalent and describe the time-evolution of loading, F(t). Eq 5 describes F(t) in terms of the initial (t=0) total effective metal mass, TEM"; Eq 7 describes F(t) in terms of the initial (t=0) loading, F_0.

To be physically valid there must be conservation of total metal mass such that TEM"(t) + ∫F(t)dt is constant for all t. In particular, mass balance requires that the integral of F(t) over all time t must equal TEM"; that is:

[8] $$TEM'' = {}_{0\text{ to }4}\, F(t)\, dt = {}_{0\text{ to }4}\, F_0\, e^{-\$\, t}\, dt = F_0 / \$$$

[6] The mathematically equivalent notation TEM", TEM"(t=0), and TEM_0" are identities that each represent the same quantity: the total effective metal mass at time t=0; i.e., TEM"=TEM"(t=0)=TEM_0". TEM", the simplest notation, is typically used in the analysis.

Solving for **$** yields:

[9] **$** = F_0 / TEM"

Consistent with **$** being averaged over a water year, F_0 in Eq 9 would be the total load for the first water year following remediation. Substituting $F_0=R_0L=$(TEM"/TEM')L, where $R_0=$TEM"/TEM' from B.2.2.1 Eq 10, into Eq 9 shows that **$** is the same for all remedial actions, including no action, and equal to the following; that is:

[10] **$** = F_0 / TEM"
= R_0L / TEM"
= (TEM"/TEM')L / TEM"
= L / TEM'

Noting that the correct load for **$** is the *yearly average* load and that TEM'=RLV'**(**C_S, Eq 10 is recast as follows:

[10a] **$** = *L* / TEM'
= *L* / RLV'**(**C_S

where:

L = Yearly average metal (zinc) loading prior to remediation (t<0)
TEM' = Total effective metal (zinc) mass prior to remediation (t<0)

L is the pre-remediation load summed over a water year and estimated from historical stream monitoring data. The symbol *L* (italic L) is used to distinguish that the correct load for **$** is the *yearly average* load.

From Eqs 4a and 7 it follows that the "half-life," $t_{½}$, of both the total effective metal mass, TEM(t)", and the stream loading, F(t), are related to **$** as follows:

[11] $t_{½}$ = ln2 / **$**

The half-life $t_{½}$ is the time it would take for TEM(t)" and F(t) to decrease by a factor of two from their initial values at time t=0. As will be discussed in the following two sections, the same equations that are used to predict metal loading can be used to predict metal concentrations (assuming constant discharge, Q). Therefore, $t_{½}$ is also the half-life for concentrations.

However, before proceeding to subsequent sections, it may be helpful to provide further detail and justification for the working hypothesis that loading is proportional to total metal mass, as represented by F(t) = **$**TEM"(t), and used in the probabilistic analysis of post-remediation metal loadings. The following exercise is intended to demonstrate that F(t) = **$**TEM"(t) is the most practical hypothesis, if not the most credible, and fully supports the purpose of the probabilistic analysis.

Further Evaluation of F(t) = $TEM"(t)

The relationship F(t) = **$**TEM"(t), as generally used in the analysis, is a particular case of the following more general relationship, where $\$_n = F_0/TEM"^n$ (see footnote[7]); that is:

[12] $F(t) = \$_n\ TEM"(t)^n$

By varying exponent n, the relationship $F(t) = \$_n TEM"(t)^n$ could allow loading to be any hypothetical yet plausible continuous function of total effective metal mass. The relationship's generality implicitly includes piece-wise time-dependence in parameter $\$_n$, such that $\$_n=\$_n(\mathbf{I})$, where **I** represents successive time periods over which $\$_n$ is reasonably constant (these time periods could be divided as finely as desired—e.g., into one-year increments—to help assure that $\$_n$ is constant over the time period). In particular, it would tentatively appear that F(t) = $\$_n TEM"(t)^n$ could approximate the *net* effect of any plausible theory of geochemical dependence between metal mass and loading. However, under no physically meaningful conditions can exponent n be negative.

The evaluation of the implications of $F(t)=\$_n TEM"(t)^n$ begins by recognizing that $\$_n=\$/TEM"^{n-1}$, where **$**, as generally used in the analysis, is the same as $\$_n$ for n=1. Next, since F(t) is the time rate of decrease of TCM"(t), it follows that:

[13] $$dTEM"(t)/dt = -F(t) = -\$_n\ TEM"(t)^n$$

[7] See previous footnote: TEM"=TEM"(t=0)=TEM_0" and are identical in meaning.

TEM"(t) is evaluated from Eq 13 for all non-negative values of n except n=1 (which was evaluated in the preceding development of F(t) = **$**TEM"(t)—where implicitly n=1); that is:

[14] $\text{TEM"}(t)^{-n}\, d\text{TEM"}(t) = -\$_n dt$

$\text{TEM"}(t)^{(1-n)}\,(1-n)^{-1} = -\$_n t + k$

[14a] $\text{TEM"}(t) = \{(-\$_n t + k)(1-n)\}^{1/(1-n)}$

Solving Eq 14a at time t=0 for integration constant k yields $k=\text{TEM"}^{1-n}(1-n)^{-1}$, where TEM"(t=0) is TEM"; substituting back into Eq 14a yields:

[15] $\text{TEM"}(t) = \{\text{TEM"}^{1-n} + \$_n\, t\,(n-1)\}^{1/(1-n)}$

Substituting $\$_n = \$/\text{TEM"}^{n-1}$ (where $\$ = F_0/\text{TEM"}$) into Eq 15 yields, after algebraic simplification:

[16] $\text{TEM"}(t) = \text{TEM"}\,\{1 + \$t\,(n-1)\}^{-[1/(n-1)]}$

Eq 16 describes how the total effective metal mass, TEM"(t), would diminish over time due to source depletion for any non-negative value of n not equal to one. Eq 16 indicates that for zero or positive n<1, TEM"(t) would continuously diminish over time and eventually reach zero, TEM"(t)=0, at time $t = t_{TEM=0}$; that is:

[17] $t_{TEM=0} = 1/\$(1-n)$ for $0 \le n < 1$
$= \text{TEM"}/F_0(1-n)$

For n>1, Eq 16 says that the total metal mass would also continuously diminish with time, but would never reach the absolute (asymptotic) limit of zero. Note that the same is true for n=1, where the metal mass diminishes exponentially over time towards the asymptotic limit of zero.

Substituting TEM"(t) from Eq 16 into Eq 12, followed by substitution of $\$_n = F_0/\text{TEM"}^n$, yields the desired relationship defining F(t) for all non-negative n except n=1; that is:

[18] $$F(t) = \$_n \, TEM''(t)^n = \$_n \, TEM''^n \{1 + \$t\,(n-1)\}^{-[n/(n-1)]} = F_0 \{1 + \$t\,(n-1)\}^{-[n/(n-1)]}$$

Eq 18 describe how F(t) would decrease with time for the general relationship $F(t) = \$_n TEM''(t)^n$ for any non-negative value except n=1 (in which case $F(t) = \$TEM''(t) = F_0 e^{-\$t}$). The following equation verifies that the integral of F(t) by Eq 18 over all time equals TEM", the total effective metal mass at time t=0, as it should. The time integral of F(t) is from t=0 to t=t*, where t* = $t_{TEM=0} = 1/\$(1-n)$ for $0 \le n < 1$ and t* = 4 (infinity) for n>1; that is:

[19] $$\int_{0 \text{ to } t^*} F(t)\,dt = \int_{0 \text{ to } t^*} F_0 \{1 + \$t\,(n-1)\}^{-[n/(n-1)]}\,dt = F_0 \{1 + \$t\,(n-1)\}^{1/(1-n)} (1-n) / \$t\,(n-1) \Big|_{t=0}^{t=t^*} = F_0 / \$ = TEM''$$

Eq 19 indeed verifies that the integral of F(t) by Eq 18 over all time equals the total effective metal mass at time t=0, TEM". It is again noted that to be physically valid there must be conservation of total metal mass such that TEM"(t) + ∫F(t)dt is constant for all t.

It remains to compare F(t) predicted for n=1 (the exponential decay of loading over time, as generally used in the probabilistic analysis) with F(t) predicted for other $n \ge 0$. The relevant two equations, as previously derived, are as follows:

$$F(t) = F_0 e^{-\$t} \qquad n=1$$
$$F(t) = F_0 \{1 + \$t\,(n-1)\}^{-[n/(n-1)]} \qquad \text{all other } n \ge 0$$

For comparison, both equations can be expressed in terms of the loading ratio $F(t)/F_0$ and graphed against $\$t$ and the related time ratio $t/t_{½}$, where $t_{½}$ is the loading half-life for n=1, the time required to reduce loading by a factor of 2 ($t_{½}=\ln 2/\$$). The dimensionless load ratio $F(t)/F_0$ is the reduction in loading over time relative to the loading at time t=0. The dimensionless factors $\$t$ and time ratio $t/t_{½}$ represent relative time. The time ratio is in multiples of $t_{½}$; that is, $t/t_{½}=1$ represents one half-life (load reduction by a factor of 2 for n=1), $t/t_{½}=2$ represents two half-lives (load reduction of 2x2=4 for n=1), and so on. As a point of reference, the half-life for loading in the upper basin is estimated to be on the order of three centuries.

Figure B-1 graphs $F(t)/F_0$ against **$**t for the following cases:

- n=1, which simulates net or total loading that is directly proportional to the effective total metal mass. This case is the working hypothesis used in the probabilistic analysis because it is both most practical and expected to reasonably approximate true conditions.

- n=0, which simulates a hypothetical situation where there is no (zero) reduction in loading until the effective metal mass has been completely depleted, after which time loading abruptly stops (i.e., zero loading for all subsequent time). As a net effect, this case is not considered plausible (except as a potential transition state to n>0 during the lag time very soon after remedial action and immediately *preceding* time t=0). Note that this implausibility should not be confused with a situation where *all* source material is completely remediated (i.e., TEM"=0) so that post-remediation loading is effectively zero at time t=0.

- n=1/2, which simulates a hypothetical situation where loading is proportional to the square root of the effective metal mass, until the effective metal mass has been completely depleted, after which time loading stops (i.e., zero loading for all subsequent time). This case could represent a tentatively plausible maximum limit on the net rate of load reduction over times t>0, at least within the first one or two half-lives.

- n=2, which simulates a hypothetical situation where loading is proportional to the square of the effective metal mass. This case could represent a tentatively plausible minimum limit on the net rate of load reduction over times t>0.

A reasonable interpretation of the results presented in Figure B-1 is that there is no practical difference between loading predictions based on n=1, n=1/2, or n=2 (including superpositions or combinations of these or similar cases).[8] Several practical considerations would appear to support such a conclusion. To begin with, existing empirical data or theoretical models to reliably distinguish a "true" value of n for post-remediation conditions are obviously lacking. It is also noted that differences in results between n=1 and other values of n continuously decrease as n approaches one. Further, it would appear questionable that differences of the magnitudes predicted could be detected within several human generations (each tick mark on the time scale of Figure B-1 is equivalent to about 20 to 30 years). Moreover, practical risk-management conclusions regarding potential remedial performance would remain unchanged within tentatively plausible variations of n. Very simply: the exact value of n is not critical to the purpose of the analysis and would not change its principal results.

[8] For example, arguing that the effective value of n could be time dependent.

These important considerations, by themselves, provide adequate justification for adopting n=1, as generally used in the analysis. However, there are additional technical reasons for using n=1. Besides being the simplest *a priori* credible form of the relationship between mass loading and metal source mass,[9] n=1 is the only non-zero value of n that is regenerative and invariant under arbitrary variations in loading history.[10] That is, all non-zero constant values of n except n=1 predict loadings F(t) and effective metal masses TEM"(t) that vary with *arbitrary* changes in the loading history, as explained next.

That all constant non-zero values of n except n=1 predict loadings F(t) and effective metal masses TEM"(t) that vary with arbitrarily changes in the loading history can be deduced as follows from F(t) calculated by Eq 18 (valid for all positive n except n=1) for any continuous time period starting at t=0 and ending at time $t = t_{PX}$ (to be defined). Start by calculating F(t) at any arbitrary time t_{P1}; this yields loading $F(t_{P1})$ at time t_{P1}. Next, use $F(t_{P1})$ as F_0 in calculating subsequent loadings at times $t>t_{P1}$. The subsequent loadings are F(t) evaluated at time $t = t–t_{P1}$; that is, $F(t=t–t_{P1})$ by Eq 18. Repeat this process for any arbitrary number (say X) of arbitrary time periods such that $t_{P1} < t_{P2} < t_{P3} < \ldots < t_{PX}$[11] (and $t_{PX} < t_{TEM=0}$ for n<1). For any given final time $t = t_{PX}$, the results demonstrate that load $F(t_{PX})$ at time t_{PX} depends on the arbitrary time periods t_{P1} through t_{PX}. This same result would follow for TEM"(t) using Eq 16 (valid for all positive n except n=1). These results would appear inconsistent with physical reality and suggest that values of n other than one are not credible.

In contrast, conducting the same exercise for arbitrary time periods t_{P1} through t_{PX} using the corresponding equations for n=1—i.e., Eq 7 for F(t) and Eq 4a for TEM(t)"—would show that $F(t_{PX})$ and $TEM"(t_{PX})$ using n=1 are *not* affected by arbitrary changes in loading history. It is noted that $F(t_{PX})$ and $TEM"(t_{PX})$ for n=0 are also not affected by arbitrary changes in loading history (for $t_{PX} < t_{TEM=0}$); however, n=0 is not considered credible for long-term net loading.

Since n=1 is the only non-zero value of n that yields physically reasonable results that are independent of arbitrary changes in loading history, it would appear that n=1 is the most credible value of n for $F(t) = \$_n TEM"(t)^n$. This is considered a tentative conclusion (and thus a working hypothesis) because it is uncertain *exactly* how loadings from real sources interact and aggregate into net loadings over time. This means that there is "model uncertainty" in F(t). For example, idealized models of chemical desorption including diffusion effects result in theoretical time-dependent behavior that is generally more complex than exponential decay (e.g., Crank 1989,

[9] Consistent with the scientific principle (called Occam's razor) to favor the simplest explanations consistent with available information (i.e., provided it is not "simpler than possible," to paraphrase Einstein).
[10] It should be understood that conservation of total metal mass is implicit; i.e., TEM"(t)+∫F(t)dt is constant for all t. The "variations in loading history" do not represent real changes in the system from external or internal physical causes. The loading history is a purely mathematical effect.
[11] X can be any integer greater than 1.

Carslaw and Jaeger 1989).[12] However, it is hypothesized that, more consistent with realistic basin conditions, appropriately randomized aggregations of such "idealized sources" undergoing subsequent successive random dilutions would effectively predict net loads of the form $F(t) = \$_n TEM''(t)^n$ that average out to approximately $F(t) = \$TEM''(t) = F_0 e^{-\$t}$. In this aggregated form, **$** would represent the net cumulative effect (averaged over a given water year) of chemical reactions and interactions including diffusion and advective effects. Be that as it may, in practical terms, model uncertainty in F(t) is dealt with probabilistically, using parameter uncertainty in **$**, as introduced in the following and developed mathematically in Section B.3.

From a practical standpoint, the most appropriate approach, as adopted in the analysis, was to accept $F(t) = \$TEM''(t) = F_0 e^{-\$t}$ as a working hypothesis and treat the uncertainty in F(t) probabilistically. Figure B-2 illustrates the effect of including parameter uncertainty in **$**. Figure B-2 repeats the results of Figure B-1 (absent n=0) and adds the 95 percent upper and lower bounds, UB and LB, respectively, for $F(t)/F_0 = e^{-\$t}$ (n=1) using a representative range of coefficients of variation for **$**, CV[**$**], from 0.50 to1.0. The upper bounds are above the deterministic case for n=1; the lower bounds are below. The widest bounds are for CV[**$**]=1.0. The results in Figure B-2 illustrate how uncertainty in $F(t) = F_0 e^{-\$t}$ manifests through parameter uncertainty in **$** via CV[**$**] and suggests how that uncertainty can effectively incorporate the model uncertainty inherent in the true value of F(t). Simply put, this practical approach was considered adequate for the purpose of the analysis.

B.2.3 LOAD RATIO, Lr

The loading capacity, symbolized C_L, is used as a metric for estimating compliance with AWQC. Specifically, AWQC are estimated using the load ratio, Lr, defined as follows, where TMDL loading capacity C_L is the product of AWQC and stream discharge Q; that is:

[1] $Lr(t) = F(t) / C_L$

where:

$C_L = AWQC \cdot Q$

[12] These models, although mathematically exact, generally assume idealized single homogenous "sources" with boundary conditions and interactions that are vastly simpler and more uniform than conditions in the basin. The solutions have the general form $F(t) = F_0 \mathbf{E}(k'+D')^{-1}(k'+D'e^{-\$t})$, where k', D', and **$** depend on desorption rate k and diffusion rate D (e.g., Crank 1989 Section 14).

Lr(t) is the number of times that the metal loading, F(t), exceeds the loading capacity, C_L. When Lr(t)=1, F(t)=C_L and the loading equals the loading capacity. For Lr(t)$\leq$1, F(t)$\leq C_L$ and the loading is at or below the loading capacity. Lr=1 is considered the *target load ratio* for complying with AWQC.

AWQC are continuous functions of water hardness, H (USEPA 1999 and 2000c); the functional relationship between AWQC and hardness is symbolized as AWQC(H). In the basin, hardness H, and therefore AWQC(H), are strongly correlated with stream location and discharge Q (USEPA 2000c). For example, the correlation coefficient between the natural logs of Q and H is 0.96 for the SFCDR at SF271, based on the data presented in USEPA 2000c (Appendix I), as used for the TMDL loading capacity. Because AWQC is a direct function of hardness, the correlation coefficient between the natural logs of AWQC and Q at SF271 is also 0.96 for the TMDL data set.

This correlation means that hardness and AWQC are *probabilistic* (or "stochastic") functions of Q. At SF271 for example, the simplest relationships with the highest correlation coefficients (0.96) were as follows, where m and b are coefficients estimated in the usual way by linear regression; that is:

$$\ln H = m_H \ln Q + b_H \text{ or } \qquad H = e^{bH} Q^{mH}$$

$$\ln AWQC = m_{AWQC} \ln Q + b_{AWQC} \quad \text{or} \quad AWQC = e^{bAWQC} Q^{mAWQC}$$

These probabilistic relationships between hardness and AWQC and Q are symbolized, respectively, as H(Q) and AWQC(Q).

The analysis used the same discharge and H(Q) and AWQC(Q) relationships used in USEPA 2000c for the TMDL. This means that the loading capacities used in the analysis are equivalent to the loading capacities used in the TMDL. It should be clear, however, that these loading capacities include no load allocations or explicit safety factors as otherwise used in the TMDLs. The loading capacities are the product of discharge Q and AWQC(Q).

Both H(Q) and AWQC(Q) appeared to be continuous functions of discharge Q, which varies with time over the water year. These relationships were assumed in the analysis to hold in the future and to remain unchanged in response to remedial actions (time t>0). The analysis of alternatives therefore postulated the following *continuous* relationship for $C_L(t)$ based on AWQC(Q) and discharge Q(t):

[2] $C_L(t) = AWQC(Q)\mathcal{C}Q(t)$

Eq 2 is a refinement of the TMDL loading capacity, the product of AWQC and discharge Q at various discrete average daily "flow tiers," as developed in the TMDL documents (USEPA 2000c). Defining $C_L(t)$ by Eq 2 means that loading estimates relate directly to AWQC estimates at any given time.

More generally, at any time and location, metal mass loading is the product of metal concentration and discharge. Post-remediation loading, F(t), is the product the post-remediation metal concentration, C"(t), and discharge Q(t); that is:

[3] $F(t) = C''(t) \cdot Q(t)$

Substituting Eqs 2 and 3 into Eq 1 yields the following important result:

[4] $Lr(t) = C''(t) \cdot Q(t) / AWQC(Q) \cdot Q(t)$
$= C''(t) / AWQC(Q)$

Eq 4 demonstrates that load ratio Lr(t) is numerically equivalent to an "AWQC ratio," defined as the ratio of concentration, C"(t), to AWQC. Lr(t) is thus *both* the number of times the loading exceeds the loading capacity, C_L, *and* the number of times the concentration exceeds the AWQC. In particular, for $Lr(t) \leq 1$, $C''(t) \leq AWQC$ and metal concentrations are expected to meet AWQC. Lr=1 is thus the *target load ratio* for compliance with AWQC.

It is reiterated that the relationships between loading capacity and AWQC, as well as load ratios, are true only in a probabilistic sense. This limitation provides part of the motivation for using a probabilistic analysis, as documented in Section B.3.

B.2.3.1 Approximate Load Ratio Scaling Factors

Dissolved chemicals that behave similarly to zinc and share source materials, can be assumed to have remediation factors approximately equal to those for zinc. Cadmium (Cd) is an example. R(t) for cadmium is approximately the same as R(t) for zinc (Zn). For R(t) the same for Cd and Zn, the load ratios (for the same location) would scale—that is:

[5] $Lr = F(t)/TMDL$
$= C''(t)/AWQC$

$$\text{Y} \quad Lr(t)_{Cd} = Lr(t)_{Zn}\{F(t)_{Cd}/F(t)_{Zn}\}\{TMDL_{Zn}/TMDL_{Cd}\}$$
$$= Lr(t)_{Zn}\{C''(t)_{Cd}/C''(t)_{Zn}\}\{AWQC_{Zn}/AWQC_{Cd}\}$$
$$= Lr(t)_{Zn}\{L_{Cd}/L_{Zn}\}\{TMDL_{Zn}/TMDL_{Cd}\}$$
$$= Lr(t)_{Zn}\{C_{Cd}/C_{Zn}\}\{AWQC_{Zn}/AWQC_{Cd}\}$$

In certain cases, the load ratios at locations that do not have independently estimated load capacities could be roughly approximated from load ratios extrapolated and interpolated from locations that do. In particular, take the hypothetical case of two nearby locations on the same stream (such that AWQC are the same at both stations)[13] and between which there is negligible remedial action (such that the R(t) are the same at both locations). In this case, the load ratio at location "2" could be approximated from the load ratio at location "1"; that is (where the equalities are approximate):

$$[6] \quad Lr_2/Lr_1 = (C_2/AWQC_2) / (C_1/AWQC_1)$$
$$= (C_2/C_1)(AWQC_1/AWQC_2)$$
$$= \{C_2/C_1\}, \text{ i.e., provided } AWQC_1 = AWQC_2$$
$$\text{Y} \quad Lr_2 = Lr_1\{C_2/C_1\}$$

A more practical situation is represented by locations 1 and 2 where the interest is in approximating the remediation factors required to achieve equal load ratios, Lr (e.g., Lr=1, 3, 7, 10 or 20), at both locations. In this case, $Lr_1 = Lr_2$, and the interest is in the ratio R_2/R_1 required to make $Lr_2/Lr_1 = 1$; that is:

$$[7] \quad Lr_2/Lr_1 = R_2C_2/AWQC_2) / (R_1C_1/AWQC_1)$$
$$= (R_2/R_1)(C_2/C_1)(AWQC_1/AWQC_2)$$
$$= 1$$
$$\text{Y} \quad R_2/R_1 = (C_1/C_2)(AWQC_2/AWQC_1)$$
$$\text{Y} \quad R_2 = R_1(C_1/C_2)(AWQC_2/AWQC_1), \text{ for } Lr_1 = Lr_2$$

[13] AWQC for a given metal is a function of hardness, which generally varies with location and discharge.

The relationship for equal load ratios stated in Eq 7 was exploited to make probabilistic estimates of remediation factors required to achieve target load ratios (expressed as equivalent C"/AWQC ratios) at stream locations upstream of TMDL locations. The probabilistic relationship is developed in Section B.3.4.4.

B.2.3.2 Gross Allocation Loads

Loading allocations as used in the TMDL were not included in the analysis. If they were, Eq 1, and associated equations, would be based on a suitable "gross allocation" (i.e., loading capacity minus background and allocations). If necessary, gross allocation could be dealt with mathematically by defining a gross allocation factor as the ratio of gross allocation to loading capacity. The gross allocation ratio would be equivalent to the gross allocation (loading) as a decimal fraction or percentage of the loading capacity. For example, the TMDL documents (USEPA 2000c) indicate gross allocations of about 30 percent to 60 percent (depending on location and flow tier) of the loading capacity. These percentages would suggest an average gross allocation factor of about 50 percent.

In all cases, the time to reach post-remediation loadings equal to gross allocations would exceed the time to meet loading capacities. The time *increment* to reduce loading from the loading capacity to the gross allocation could be estimated from $F(t)=F_0e^{-\$t}$ by setting the load capacity equal to F_0 and the gross allocation to F(t)=F and solving for t; this results in $t = \ln\{F_0/F\}/\$$, where F_0/F is the reciprocal of the gross allocation ratio (note that for a gross allocation factor of 50 percent , such that $F_0/F=2$, the time increment is the half life, i.e., $t = t_{½} = \ln 2/\$$).

Estimates of the time to reach a gross allocation loading could use estimates of the loading ratio Lr(t) and setting the gross allocation factor as the target loading ratio (which is presently 1.0, as used in the analysis for the loading capacity). The time at which Lr(t) reaches the gross allocation factor is the estimated time at which the loading would reach the gross allocation loading.

For example, using the graphical results of the loading ratio Lr(t) presented in Section B.2, the estimated time to reach a gross allocation loading of 0.5 times the loading capacity would be the time at which Lr(t) equaled 0.5. More refined estimates could be made by formally including the gross allocation factor as another uncertain variable, with its effects appropriately propagated mathematically through the analysis.

B.2.3.3 Background Loads and Concentrations

Analysis results for load ratios can be reasonably assumed, for practical purposes, to implicitly include background effects—both for loading and AWQC (recall that load ratios are numerically equivalent to concentration ratios in terms of AWQC). This assumption is justified as follows.

First, although variable, background loadings are approximately 10 percent of loading capacities based on (USEPA 2000c). Second, EPA's TMDL loading capacities (USEPA 2000c), and thus the loading capacities used in the analysis, use a conservative interpretation of the H(Q) relationships (i.e., the lower bound of the 90th percentile confidence interval for hardness estimates). This conservative interpretation of H(Q) results in underestimates of AWQC(H). The degree of underestimation, although variable, appears to be at least equal to background (zinc) concentrations. The loading capacities and equivalent AWQC, and thus load ratios, can therefore be reasonably assumed for the analysis to implicitly include background effects.

B.2.4 POST-REMEDIATION CONCENTRATIONS, C"(t)

As demonstrated in Section B.2.3, load ratios, Lr(t), represent both the number of times the loading, F(t), exceeds the loading capacity, C_L, and the number of times the post-remediation concentration, C"(t), exceeds AWQC. Because post-remediation concentrations are implicit in the loading ratio, Lr(t), they are not explicitly presented in the technical memorandum.

Post-remediation concentrations can, however, be approximated with the same equations used to estimate post-remediation loading. That is, C"(t) can be estimated from pre-remediation concentrations C using the relationships already defined; that is:

[1] $$\begin{aligned} C''(t) &= F(t) / Q(t) \\ &= R(t)\, L / Q(t) \\ &= R(t)\, C Q / Q(t) \\ &= R(t)\, C\, r_Q \end{aligned}$$

where:

r_Q = Q/Q(t) is the pre-to-post-remediation discharge ratio
= 1 to a first approximation

Eq 1 says that remediation factors R(t) also apply to post-remediation *concentrations* when adjusted, if and as appropriate, by the pre-to-post-remediation discharge ratio, r_Q. For r_Q=1 remediation factors for loadings and concentrations are equal,[14] and yield equal pre-to-post-remediation reductions in loadings and concentrations. Assuming r_Q=1 (i.e., equating pre-remediation and post-remediation discharges) is a reasonable first approximation that results in the following practical simplification of Eq 1:

[14] More precisely, both remediation factors would be lognormally distributed and have the same expected value and coefficient of variation.

[2] $C''(t) = R(t)\,C$

The form of Eq 2 and B.2.1 Eq 1, F(t)=R(t)L, are identical.[15] This important result means that post-remediation concentrations, C"(t), can be approximated with the same equations used to estimate post-remediation loading, F(t), by substituting C for L and C" for F. It may be noted that the resulting equations are consistent with first-order chemical kinetics. However, first-order kinetics, although plausible, is not assumed in the analysis.

[15] This identity would hold for any r_Q by simply defining a remediation factor for concentrations as $R_C(t)=R(t)r_Q$.

B.3.0 PROBABILISTIC FORMULATION

The forgoing development in Section B.2 comprises the deterministic portion of the analysis. The deterministic equations were formulated probabilistically, as documented in this section. The probabilistic formulation quantifies the uncertainty inherent in the deterministic analysis and resulting remedy performance measures so as to help support risk management decisions.

The probabilistic formulation extends the deterministic formulation using the mathematics of probability theory and error propagation. The probabilistic formulation is based on the ideas and methods documented in Appendix A. It characterizes the uncertainty, including natural variability, in pre-remediation and post-remediation loadings and concentrations as lognormal distributions. The implicit location dependence of all parameters remains tacit.

B.3.1 PRE-REMEDIATION LOADING, L

This section focuses the background information presented in Appendix A to document the equations used in the statistical analysis of historical stream monitoring data to estimate pre-remediation loading L. The loadings are characterized as lognormal distributions defined by their probabilistic parameters: expected value, E[L], and coefficient of variation, CV[L]. The historical data used to make the estimates, as well as the principal results of the statistical analyses, are discussed and documented in the RI (USEPA 2000a) and summarized in the FS (USEPA 2000b). The results shown in the RI, based on the following lognormal statistical analyses, demonstrate that metal loading L and metal concentrations C, are lognormally distributed,[16] in agreement with relevant theory (Ott 1995).

Estimates of E[L] and CV[L] were based on lognormal linear least-squares regression analyses of load data L_i calculated from a total of n paired measurement sets (index i=1 to n) of stream discharge Q_i and concentration C_i using the basic relationship $L_i = Q_iC_i$. To the extent there was correlation between Q_i and C_i, it was thus implicit in the calculation of L_i. The lognormal statistical data analyses reported in the RI demonstrate that metal loading is lognormally distributed.

The normal standard variate u was regressed against the natural log of the L_i data, $\ln L_i$, using the linear model $u=b+m*\ln L_i$, where m and b are the slope and intercept, respectively, of the regression line. The L_i data were ranked from lowest to highest (i=1 to n) and assigned a "plotting point" as an estimate of the cumulative frequency or probability of each ranked L_i (or

[16] Generally, the closest or "best" fits occur for dissolved metal concentrations and loadings.

$\ln L_i$). The plotting point, PP, (for a normal distribution) for each L_i was as follows (again, i is the rank of L_i):

[1] $PP_i = (i - 3/8) / (n + 1/4)$

The value of u for each L_i was calculated as the inverse standard normal cumulative distribution, CDF^{-1}, evaluated at the plotting point; that is:

[2] $u_i = CDF^{-1}\{(i - 3/8) / (n + 1/4)\}$
$= CDF^{-1}\{PP_i\}$

The u_i were regressed against $\ln L_i$ to calculate the least squares ("best fit") regression line, $u=b+m*\ln L_i$. E[L] and CV[L] were estimated from the slope and intercept of the regression line using the following relationships for a lognormal variate:

[3] $E[L] = \exp[-b/m + ½m^{-2}]$

[4] $CV[L] = \{\exp[m^{-2}] - 1\}^{1/2}$

The same approach was used in the RI to estimate the expected value and coefficient of variation of discharge, E[Q] and CV[Q], and concentration, E[C], and CV[C], from the available historical data. In addition, the sample correlation coefficient, r, was used an estimate of p, the correlation between Q and C. In some cases, these results were used in the probabilistic analysis to estimate loadings at monitoring locations using the following relationships for lognormal distributions:

[5] $E[L] = E[Q]\,E[C]\,\mathbf{S}$

[6] $CV[L] = \{(CV[C]^2+1)(CV[Q]^2+1)\mathbf{S}^2 - 1\}^{1/2}$

$\mathbf{S} = \exp\{p\{\ln(CV[Q]^2+1)\ln(CV[C]^2+1)\}^{1/2}\}$ see footnote[17]

[17] In general, expressions of the form $\exp\{A\}^{1/2}$ should be read as $(\exp\{A\})^{1/2}$; this convention is similar to what is used in MS Excel and has been adopted in the text to help reduce algebraic complexity.

p	=	Correlation coefficient between the natural log of Q and the natural log of C
	=	$p_{lnQ,lnC}$

Potentially, Eqs 5 and 6 could be used to make refined load estimates where the historical Q data and C data are more extensive and representative than that included in the available *paired* values of Q_i and C_i. In this case, the larger data sets could be used to estimate the probabilistic parameter of Q and C, although it would still be necessary to estimate $p_{lnQ,lnC}$ from the paired data. For the special case that $p_{lnQ,lnC}$ is zero, but only in this special case, Eqs 5 and 6 would reduce to the following:

[7] $E[L] = E[Q]\,E[C]$

[8] $CV[L] = \{(CV[C]^2+1)(CV[Q]^2+1) - 1\}^{1/2}$

Although not used in the analysis, possible modifications to the pre-remediation estimates E[L] and CV[L], which are based on analysis of the historical measurement data as discussed to this point, could be accomplished using factor **M** discussed as follows.

Modifications to E[L] and CV[L]

In the hypothetical situation that it was considered appropriate, modifications or refinements to the pre-remediation estimates E[L] and CV[L] could be accomplished using a modifying factor **M** with E[L] and CV[L] implemented as follows, where E[L"] and CV[L"] represent the modified estimates; that is:

[9] $E[L''] = E[L]\,E[\mathbf{M}]\,\mathbf{S}$

[10] $CV[L''] = \{(CV[L]^2+1)(CV[\mathbf{M}]^2+1)\mathbf{S}^2 - 1\}^{1/2}$

S	=	$\exp\{p\{\ln(CV[L]^2+1)\ln(CV[\mathbf{M}]^2+1)\}^{1/2}\}$
p	=	Correlation coefficient between the natural log of L and the natural log of **M**
	=	$p_{lnL,ln\mathbf{M}}$

The probabilistic parameter E[**M**] would represent a systematic change from the loading reflected in the measured historical data, as reported in the RI. The probabilistic parameter CV[**M**] would represent additional uncertainty or variability not reflected in that historical data. If Eqs 9 and 10 were evoked, E[L"] and CV[L"] would be used in place of E[L] and CV[L] in all affected equations as documented in the appendix. The results in the analysis tacitly used E[**M**]=1 and CV[**M**]=0, consistent with the interpretation of available information and the purpose of the analysis.

It is important to recognize that temporal loading variations, reflecting the natural variability (including daily, seasonal, and year-to-year) inherent in the historical data, are implicit in CV[L]. This variability is assumed constant over time after remediation (except for correlation effects with remedial action). Thus, although the potential variability within a given water year is included in L, L is not otherwise a function of time after remediation. The pre-remediation load estimates L are used with estimated remediation factors, R(t), to estimate post-remediation loading, F(t), as presented in the following section.

B.3.2 POST-REMEDIATION LOADING, F(t)

Based on the relationship for post-remediation loading at any given time t expressed in the following Eq 1, the expected value and coefficient of variation of post-remediation loading, E[F(t)] and CV[F(t)], were estimated as follows:

[1] $F(t) = R(t) \cdot L$

[2] $E[F(t)] = E[R(t)]\, E[L]\, \mathbf{S}$

[3] $CV[F(t)] = \{(CV[R(t)]^2+1)(CV[L]^2+1)\mathbf{S}^2 - 1\}^{1/2}$

$\mathbf{S} = \exp\{p\{\ln(CV[R(t)]^2+1)\ln(CV[L]^2+1)\}^{1/2}\}$ see last footnote

p = Correlation coefficient between the natural log of R(t) and the natural log of L

$= \rho_{\ln L, \ln R}$

The lognormal statistical data analyses reported in the RI empirically demonstrate that L is lognormally distributed, which is in agreement with theory (Ott 1995). Because the underlying physics leading to lognormality will not be changed by remedial action, it is expected that post-remediation loading F(t) will also be lognormally distributed. An important implication of both

L and F(t) being lognormal is that R(t) should also be lognormal, since products and quotients of lognormal distributions are also lognormal.[18]

As previously noted, temporal loading variations, reflecting the natural variability (including daily, seasonal, and year-to-year) inherent in the historical data, are implicit in CV[L]. This natural variability is assumed to continue into the future and affect post-remediation loading. The affect is assumed constant over time after remediation (t>0) except for correlation effects with R(t), as included in Eq 3 and discussed in the following. Although the potential variability within a given water year is included in L, L is not otherwise a function of time t.

R(t) is averaged over one-water-year increments and can be considered to measures the expected yearly average reduction over time from historical loading due to remediation. E[R(t)] and CV[R(t)] measure, respectively, the expected value and the uncertainty in that reduction.

The expected value of F(t) is measured by E[F(t)] with the variability measured as CV[F(t)], which includes the natural variability estimated by CV[L] and the uncertainty in the *prediction* of R(t), estimated by CV[R(t)]. For R(t) and L lognormally distributed, F(t) is also lognormally distributed and Eqs 2 and 3 are mathematically exact. Note that Eqs 1, 2, and 3 also apply where time is implicit—in which case F(t)=F and R(t)=R.

Estimates of the correlation between L and R(t) (as measured by $p_{lnL,lnR}$) were based on professional judgment and interpretation of potential remedial action behavior. Although there is no practical way to quantitatively predict the correlation, it is expected that remedial action will generally be relatively more effective at reducing high loadings (which correlate with high flow conditions) than reducing low loadings (which correlate with low flow conditions) such that L and R(t) will be negatively correlated. The mid-range negative value of $p_{lnL,lnR}$=–0.5 was considered reasonable[19] and used in the analysis for all action alternatives. This –0.5 value is consistent with the maximum entropy estimate for an uncertain variable uniformly distributed between –1 and 0 or any range having –0.5 as a midpoint. Sensitivity analysis could be used to estimate the effect of credible variations in this parameter.

Non-exceedance Estimates of Post-remediation Loading

Non-exceedance estimates of the true value of F(t) are denoted $F(t)_{Pn}$. The subscript "Pn" is the non-exceedance probability, the probability that the *true* value of F(t) *does not exceed* the

[18] R(t) could deviate somewhat from lognormality, particularly at values close to one, and still be compatible with L and F(t) being lognormal. Specifically, for R(t) close to one, it could be argued that R(t) might be more "accurately" represented by a beta distribution over $0 \leq R(t) \leq 1.0$, or even an inverse lognormal distribution constrained at R(t)=1.0. These are considered mathematical fine points that would complicate the analysis without practical significance in the results.

[19] Very extreme events could increase $p_{lnL,lnR}$.

estimate $F(t)_{Pn}$. Equivalently, Pn is the probability that $F(t)_{Pn}$ *overestimates* the true value of F(t), and is thus not an underestimate of the true value of F(t). The complement of the non-exceedance probability, 1-Pn, is the probability that the true value of F(t) does exceed the estimate $F(t)_{Pn}$ or, equivalently, that $F(t)_{Pn}$ underestimates the true value of F(t). The non-exceedance estimate is given by the following equation, where u_{Pn} is the standardized normal variate corresponding to Pn; that is:

[4] $$F(t)_{Pn} = E[F(t)] \exp\{u_{Pn} \ln(CV[F(t)]^2+1)^{1/2}\} / (CV[F(t)]^2+1)^{1/2}$$

A relatively high value of Pn (e.g., 90 percent) is used to make "reliable" estimates of F(t). For the analysis, the reliable estimates used a 90 percent probability of non-exceedance, Pn=90 percent , which makes u_{Pn}=1.28. Substituting Pn=90 percent and u_{Pn}=1.28 into Eq 4 yields:

[5] $$F(t)_{90\%} = E[F(t)] \exp\{1.28 \ln(CV[F(t)]^2+1)^{1/2}\} / (CV[F(t)]^2+1)^{1/2}$$

$F(t)_{90\%}$ has an estimated 90 percent reliability (probability) of overestimating the true value of F(t). Equivalently, there is an estimated 90 percent reliability that the true value of F(t) does not exceed the estimate $F(t)_{90\%}$. $F(t)_{90\%}$ can be thought of as a reasonable *practical upper-bound* estimate of the true value of F(t).

Similarly, $F(t)_{10\%}$ represents a *practical lower-bound* estimate of the true value of F(t). $F(t)_{10\%}$ has an estimated 90 percent probability of *underestimating* the true value of F(t), or equivalently, a 90 percent probability of being *exceeded* by the true value of F(t). $F(t)_{10\%}$ reflects a 10 percent probability of non-exceedance, Pn=10%, making u_{Pn}=–1.28. Substituting Pn=10% and u_{Pn}=–1.28 into Eq 4 yields:

[6] $$F(t)_{10\%} = E[F(t)] \exp\{-1.28 \ln(CV[F(t)]^2+1)^{1/2}\} / (CV[F(t)]^2+1)^{1/2}$$

A probable range of potential future loading was defined by the limits $F(t)_{10\%}$ ("lower-bound") and $F(t)_{90\%}$ ("upper-bound"). Theoretically, these approximate limits would have an 80 percent probability of containing the true value of F(t).

Non-exceedance estimates for other variables follow the same approach, as presented here for loading, and are similarly interpreted.

B.3.3 REMEDIATION FACTOR, R(t)

Based on the relationship for the remediation factor expressed in the following Eq 1, the expected value and coefficient of variation of the remediation factor, E[R(t)] and CV[R(t)], were estimated as follows:

[1] $R(t) = R_0 e^{-\$t}$

[2] $E[R(t)] = E[R_0]E[e^{-\$t}]\ \mathbf{S}$

[3] $CV[R(t)] = \{(CV[R_0]^2+1)(CV[e^{-\$t}]^2+1)\mathbf{S}^2 - 1\}^{1/2}$

$\mathbf{S} = \exp\{-p\{\ln(CV[R_0]^2+1)\ln(CV[e^{-\$t}]^2+1)\}^{1/2}\}$

p = Correlation coefficient between the natural log of R_0 and $\$t$

= $p_{lnR0,\$t}$ (later set to 0, so $\mathbf{S}$=1, implying Eqs 2 and 3 are mathematically exact even if R_0 and $\$t$ are not lognormal)

Eqs 2 and 3 are mathematically exact under two conditions. The first is if R_0 is lognormally distributed and the decay rate $\$$ is normally distributed (making $e^{-\$t}$ lognormally distributed). The second is if there is no correlation between R_0 and $\$$ (i.e., $p_{lnR0,\$t}$=0), in which case Eqs 2 and 3 are mathematically exact regardless of the distributions of R_0 and $\$$. Note that $p_{lnR0,\$t}$=0 for no action, R_0=1.

As previously discussed, R(t) should be lognormal; R_0 should thus also be lognormal, However, based on physical arguments, uncertainty and variability in the decay rate is expected to be more appropriately represented by a lognormal distribution (truncated at the theoretical extreme values), rather than a normal distribution (although truncation could make $\$$ approximately normal). Lognormal $\$$ would mean that Eqs 2 and 3 are mathematical approximations if there is correlation between R_0 and $\$$. Although the mathematical error resulting from such an approximation was considered insignificant for the purpose of the analysis, a sensitivity analysis of Eqs 2 and 3 to credible variations in $p_{lnRo,Bt}$ was conducted. The analysis began by estimating $p_{lnRo,Bt}$, starting with the following definition:

[4] $p_{lnR0,\$t} = Cov[\ln R_0,\$t] / \{V[\ln R_0]V[\$t]\}^{1/2}$

The variance terms in Eq 4 were calculated exactly and the covariance term approximated using a Taylor series expansion method, as documented in Appendix A. The approximation was developed as follows. Dummy variables X_i and X_j were used to represent the uncertain variables

making up R_0 (represented as RLV"/RLV') and **$**t (represented as *L*t/RLV'(Cs) with the partial derivatives evaluated at the mean (m) or expected value of the variables; that is:

[5] $$\mathrm{Cov}[[\ln R_0,\$t] = \mathbf{E}_{i=1,n}\mathbf{E}_{j=1,n}\{\partial \ln R_0/\partial X_i\}_m\{\partial \$t/\partial X_j\}_m \mathrm{Cov}[X_i,X_j]$$
$$= \mathbf{E}_{i=1,n}\mathbf{E}_{j=1,n}\{\partial \ln(RLV''/RLV')/\partial X_i\}_m\{\partial(Lt/RLV'(Cs)/\partial X_j\}_m \mathrm{Cov}[X_i,X_j]$$

$$= Lt\ V[RLV'] / (RLV'^3(Cs) - Lt\ \mathrm{Cov}[RLV',RLV''] / (RLV'^2RLV''(Cs)$$

$$= \$t\{CV[RLV']^2 - p_{RLV',RLV''}\ CV[RLV']CV[RLV'']\}$$

The variances of $\ln R_0$ and **$**t were estimated as follows, assuming R_0 to be lognormally distributed; that is:

[6a] $$V[\ln R_0] = \ln(CV[R_0]^2+1)$$

[6b] $$V[\$t] = E[\$t]^2CV[\$t]^2$$
$$= E[\$t]^2CV[\$]^2$$

Substituting Eqs 6a,b and Eq 5 into Eq 4 yielded, after algebraic simplification, the estimate for $p_{lnRo,Bt}$; that is:

[7] $$p_{lnRo,Bt} = \{CV[RLV']^2 - p_{RLV',RLV''}\ CV[RLV']CV[RLV'']\} / CV[\$]\ln(CV[R_0]^2+1)^{1/2}$$

Eq 7 was used with the results from Section B.3.3.1 for CV[RLV'], CV[RLV"], CV[R_0] and $p_{RLV',RLV''}$ (equated with $p_{lnRLV',lnRLV''}$[20]) and the results from Section B.3.3.2 for CV[**$**]. Estimates of $p_{lnRo,Bt}$ for the alternatives were made using Eq 7 and evaluated. The results indicated that $p_{lnRo,Bt}$ lies within a probable range of approximately –0.1 to 0.1 for the alternatives with negligible net effect on predicted post-remediation results compared to $p_{lnR0,\$t}$=0. Given these results, $p_{lnR0,\$t}$ was set to zero in subsequent analyses. Besides helping to simplify the analysis, setting $p_{lnR0,\$t}$ to zero resulted in Eqs 2 and 3 being mathematically exact.

[20] For expediency and with minor error.

Non-exceedance Estimates of Remediation Factor

Similar to non-exceedance estimates for F(t), non-exceedance estimates of the true value of R(t) are denoted $R(t)_{Pn}$. Non-exceedance probability Pn is both the probability that the true value of R(t) does not exceed the estimate $R(t)_{Pn}$ and the probability that $R(t)_{Pn}$ overestimates the true value of R(t). $R(t)_{Pn}$ is given by the following equation, where u_{Pn} is the standardized normal variate corresponding to Pn; that is:

[8] $$R(t)_{Pn} = E[R(t)] \exp\{u_{Pn} \ln(CV[R(t)]^2+1)^{1/2}\} / (CV[R(t)]^2+1)^{1/2}$$

For 90 percent and 10 percent probabilities of non-exceedance Eq 8 yields:

[9] $$R(t)_{90\%} = E[R(t)] \exp\{1.28 \ln(CV[R(t)]^2+1)^{1/2}\} / (CV[R(t)]^2+1)^{1/2}$$

[10] $$R(t)_{10\%} = E[R(t)] \exp\{-1.28 \ln(CV[R(t)]^2+1)^{1/2}\} / (CV[R(t)]^2+1)^{1/2}$$

$R(t)_{90\%}$ has an estimated 90 percent reliability (probability) of *overestimating* the true value of R(t), and $R(t)_{10\%}$ has an estimated 90 percent probability of *underestimating* the true value of R(t). $R(t)_{90\%}$ and $R(t)_{10\%}$ can be thought of as practical upper and lower bounds on the true value of R(t). The range of potential remediation factors between the limits $R(t)_{10\%}$ (lower-bound) and $R(t)_{90\%}$ (upper-bound) would, theoretically, have an 80 percent probability of containing the true value of R(t).

B.3.3.1 Remediation Factor at Time Zero (t=0), R_0

The remediation factor at time t=0, R_0, is estimated as the ratio of the post-remediation relative loading volume, RLV"=**E**(RLP*EVR), to the pre-remediation relative loading volume, RLV'=**E**(RLP*V). Both of these volumes are the aggregate sums of the various RLP, V, and R associated with a given remedial alternative. The uncertainty in the RLP, V, and R estimates are represented by lognormal distributions. The uncertainty in the aggregate sums RLV" and RLV' are also approximated as lognormal distributions because their ratio, R_0, is lognormal (as previously discussed), which means the sums should be lognormal, or approximately lognormal. The approximate lognormality of sums of lognormal distributions is discussed in Appendix A.

Based on the relationship for the remediation factor expressed in the following Eq 1, the expected value and coefficient of variation of the remediation factor, $E[R_0]$ and $CV[R_0]$, were estimated as follows:

[1] $$R_0 = \frac{\mathbf{E}_{j=1,K}\, RLP_j\{\mathbf{E}_{i=1,Nj}\, V_{i,j}R_{i,j}\}}{\mathbf{E}_{j=1,K}\, RLP_j\{\mathbf{E}_{i=1,Nj}\, V_{i,j}\}} = \frac{\mathbf{E}(RLP*\mathbf{E}VR)}{\mathbf{E}(RLP*V)} = \frac{RLV''}{RLV'}$$

[2] $$E[R_0] = (E[\mathbf{E}(RLP*\mathbf{E}VR)] / E[\mathbf{E}(RLP*V)])(1+CV[\mathbf{E}(RLP*V)]^2)\mathbf{S}$$
$$= (E[RLV''] / E[RLV'])(1+CV[RLV']^2)\mathbf{S}$$

[3] $$CV[R_0] = \{(CV[\mathbf{E}(RLP*\mathbf{E}VR)]^2+1)(CV[\mathbf{E}(RLP*V)]^2+1)\mathbf{S}^2 - 1\}^{1/2}$$
$$= \{(CV[RLV'']^2+1)(CV[RLV']^2+1)\mathbf{S}^2 - 1\}^{1/2}$$

$$\mathbf{S} = \exp\mathbf{\{}-2p\{\ln(CV[\mathbf{E}(RLP*\mathbf{E}VR)]^2+1)\ln(CV[\mathbf{E}(RLP*V)]^2+1)\}^{1/2}\mathbf{\}}^{1/2}$$
$$= \exp\mathbf{\{}-p\{\ln(CV[RLV'']^2+1)\ln(CV[RLV']^2+1)\}^{1/2}\mathbf{\}}$$

p = Correlation coefficient between the natural log of RLV''=$\mathbf{E}$(RLP*$\mathbf{E}$VR) and the natural log of RLV'=$\mathbf{E}$(RLP*V)

$$= \rho_{lnRLV',lnRLV''}$$

The second equality in the following relations are mathematically exact. Subsequent equalities are mathematically exact provided that all RLP_j, $V_{i,j}$, and $R_{i,j}$ are uncorrelated for a given remedial alternative.

[4] $$E[RLV'] = E[\mathbf{E}(RLP*V)]$$
$$= \mathbf{E}_{j=1,K}\, \mathbf{E}_{i=1,Nj}\, E[RLP_j*V_{i,j}]$$
$$= \mathbf{E}_{j=1,K}\, \mathbf{E}_{i=1,Nj}\, E[RLP_j]E[V_{i,j}]$$

[5] $$E[RLV''] = E[\mathbf{E}(RLP*\mathbf{E}VR)]$$
$$= \mathbf{E}_{j=1,K}\, \mathbf{E}_{i=1,Nj}\, E[RLP_j*V_{i,j}*R_{i,j}]$$
$$= \mathbf{E}_{j=1,K}\, \mathbf{E}_{i=1,Nj}\, E[RLP_j]E[V_{i,j}]E[R_{i,j}]$$

[6] $$CV[RLV'] = CV[\mathbf{E}(RLP*V)]$$

$$= \{V[\mathbf{E}_{j=1,K}\,\mathbf{E}_{i=1,Nj}\,RLP_j*V_{i,j}] \,/\, \mathbf{E}_{j=1,K}\,\mathbf{E}_{i=1,Nj}\,E[RLP_j*V_{i,j}]^2\}^{1/2}$$

$$= \{\mathbf{E}_{j=1,K}\,\mathbf{E}_{i=1,Nj}\,V[RLP_j*V_{i,j}] \,/\, \mathbf{E}_{j=1,K}\,\mathbf{E}_{i=1,Nj}\,E[RLP_j*V_{i,j}]^2\}^{1/2}$$

$$= \{\mathbf{E}_{j=1,K}\,\mathbf{E}_{i=1,Nj}\,(CV[RLP_j*V_{i,j}]E[RLP_j*V_{i,j}])^2 \,/\, \mathbf{E}_{j=1,K}\,\mathbf{E}_{i=1,Nj}\,E[RLP_j*V_{i,j}]^2\}^{1/2}$$

$$= \mathbf{\{}\mathbf{E}_{j=1,K}\,\mathbf{E}_{i=1,Nj}\,\{(CV[RLP_j]^2+1)(CV[V_{i,j}]^2+1)-1\}(E[RLP_j]E[V_{i,j}])^2 \,/$$

$$/\, \mathbf{E}_{j=1,K}\,\mathbf{E}_{i=1,Nj}\,(E[RLP_j]E[V_{i,j}])^2\mathbf{\}}^{1/2}$$

[7] $$CV[RLV''] = CV[[\mathbf{E}(RLP*\mathbf{E}VR)]$$

$$= \{V[\mathbf{E}_{j=1,K}\,\mathbf{E}_{i=1,Nj}\,RLP_j*V_{i,j}*R_{i,j}] \,/\, \mathbf{E}_{j=1,K}\,\mathbf{E}_{i=1,Nj}\,E[RLP_j*V_{i,j}*R_{i,j}]^2\}^{1/2}$$

$$= \{\mathbf{E}_{j=1,K}\,\mathbf{E}_{i=1,Nj}\,V[RLP_j*V_{i,j}*R_{i,j}] \,/\, \mathbf{E}_{j=1,K}\,\mathbf{E}_{i=1,Nj}\,E[RLP_j*V_{i,j}*R_{i,j}]^2\}^{1/2}$$

$$= \{\mathbf{E}_{j=1,K}\,\mathbf{E}_{i=1,Nj}\,(CV[RLP_j*V_{i,j}*R_{i,j}]E[RLP_j*V_{i,j}*R_{i,j}])^2 \,/\, \mathbf{E}_{j=1,K}\,\mathbf{E}_{i=1,Nj}\,E[RLP_j*V_{i,j}*R_{i,j}]^2\}^{1/2}$$

$$= \mathbf{\{}\mathbf{E}_{j=1,K}\,\mathbf{E}_{i=1,Nj}\,\{(CV[RLP_j]^2+1)(CV[V_{i,j}]^2+1)\,)(CV[R_{i,j}]^2+1)-1\}(E[RLP_j]E[V_{i,j}]\,E[R_{i,j}])^2 \,/$$
$$/\, \mathbf{E}_{j=1,K}\,\mathbf{E}_{i=1,Nj}\,(E[RLP_j]E[V_{i,j}]E[R_{i,j}])^2\mathbf{\}}^{1/2}$$

There is a practically significant correlation between the numerator of Eq 1, RLV"= $\mathbf{E}$(RLP*$\mathbf{E}$VR), and the denominator, RLV'=$\mathbf{E}$(RLP*V), because of the common RLP and V terms. The correlation is measured by $\rho_{lnRLV',lnRLV''}$, and defined as follows:

[8] $$\rho_{lnRLV',lnRLV''} = \frac{Cov[\ln\mathbf{E}(RLP*\mathbf{E}VR),\ \ln\mathbf{E}(RLP*V)]}{\{V[\ln\mathbf{E}(RLP*\mathbf{E}VR)]V[\ln\mathbf{E}(RLP*V)]\}^{1/2}}$$

$$= Cov[lnRLV'', lnRLV'] \,/\, \{V[lnRLV'']V[lnRLV']\}^{1/2}$$

The variance terms in the denominator of Eq 8 were calculated exactly from the basic relationships defining lognormal distributions. The covariance term in the numerator was mathematically approximated using a Taylor series expansion method, as documented in Appendix A. The approximation is as follows where X_i and X_j represent RLP, V, and R for the various source types, remedial actions, and remediated volumes for a given remedial alternative and the partial derivatives are evaluated at the mean (m) or expected value of the variables; that is:

[9] $\mathrm{Cov}[\ln RLV'', \ln RLV'] = \mathrm{Cov}[\ln \mathbf{E}(RLP^*\mathbf{E}VR), \ln \mathbf{E}(RLP^*V)]$

$$= \mathbf{E}_{i=1,n}\mathbf{E}_{j=1,n}\ \{\mathbb{M}\ln \mathbf{E}(RLP^*\mathbf{E}VR)/\mathbb{M}X_i\}_m\{\mathbb{M}\ln \mathbf{E}(RLP^*V)/\mathbb{M}X_j\}_m\ \mathrm{Cov}[X_i,X_j]$$

Assuming no correlation between the various RLP_j, $V_{i,j}$, and $R_{i,j}$, Eq 9 simplifies to the following:

[10] $\mathrm{Cov}[\ln RLV'', \ln RLV'] = \mathrm{Cov}[\ln \mathbf{E}(RLP^*\mathbf{E}VR), \ln \mathbf{E}(RLP^*V)]$

$$= \mathbf{E}_{j=1,K}\,\mathbf{E}_{i=1,Nj}\{E[V_{i,j}]^2E[R_{i,j}]V[RLP_j] + E[RLP_j]^2E[R_{i,j}]V[V_{i,j}]\} / E[\mathbf{E}(RLP^*\mathbf{E}VR)]E[\mathbf{E}(RLP^*V)]$$

$$= \mathbf{E}_{j=1,K}\,\mathbf{E}_{i=1,Nj}\{E[V_{i,j}]^2E[RLP_j]^2E[R_{i,j}]\ (CV[RLP_j]^2 + CV[V_{i,j}]^2)\} / E[\mathbf{E}(RLP^*\mathbf{E}VR)]E[\mathbf{E}(RLP^*V)]$$

$$= \mathbf{E}_{j=1,K}\,\mathbf{E}_{i=1,Nj}\{E[V_{i,j}]^2E[RLP_j]^2E[R_{i,j}]\ (CV[RLP_j]^2 + CV[V_{i,j}]^2)\} / E[RLV'']E[RLV']$$

The variances of $\ln \mathbf{E}(RLP^*\mathbf{E}VR)=\ln RLV''$ and $\ln \mathbf{E}(RLP^*V)=\ln RLV'$ in the denominator of Eq 8 are estimated assuming that they are lognormally distributed, such that:

[11a] $V[\ln \mathbf{E}(RLP^*V)] = V[\ln RLV'] = \ln(CV[\mathbf{E}(RLP^*V)]^2+1)$

[11b] $V[\ln \mathbf{E}(RLP^*\mathbf{E}VR)] = V[\ln RLV''] = \ln(CV[\mathbf{E}(RLP^*\mathbf{E}VR)]^2+1)$

where $CV[\mathbf{E}(RLP^*V)]$ and $CV[\mathbf{E}(RLP^*\mathbf{E}VR)]$ are given by Eqs 6 and 7. Substituting Eqs 10 and 11a,b into Eq 8 yields the estimate for the needed correlation coefficient; that is:

[12]
$$\rho_{\ln RLV',\ln RLV''} = \frac{\mathbf{E}_{j=1,K}\,\mathbf{E}_{i=1,Nj}\ \{E[V_{i,j}]^2E[RLP_j]^2E[R_{i,j}](CV[RLP_j]^2 + CV[V_{i,j}]^2)\} / E[\mathbf{E}(RLP^*\mathbf{E}VR)]E[\mathbf{E}(RLP^*V)]}{\{\ln(CV[\mathbf{E}(RLP^*\mathbf{E}VR)]^2+1)*\ln(CV[\mathbf{E}(RLP^*V)]^2+1)\}^{1/2}}$$

$$= \frac{\mathbf{E}_{j=1,K}\,\mathbf{E}_{i=1,Nj}\ \{E[V_{i,j}]^2E[RLP_j]^2E[R_{i,j}](CV[RLP_j]^2 + CV[V_{i,j}]^2)\}}{\{\ln(CV[RLV'']^2+1)*\ln(CV[RLV']^2+1)\}^{1/2}\ E[RLV'']E[RLV']}$$

In cases where E[RLV'] and E[RLV"] are approximately equal, such that $E[R_0]$ is close to 1.0, it appeared that Eq 10 somewhat underestimated the exact value of the covariance, and therefore the correlation coefficient $p_{lnRLV',lnRLV"}$ by Eq 12. The underestimate likely occurs because of the neglected covariance increment from correlations between source type RLPs, which is not accounted for in Eq 10.

The net effect of the underestimation of $p_{lnRLV',lnRLV"}$ was that for $E[R_0]$ close to 1.0 the resultant estimated probability that the true value of R_0 exceeded 1.0, $P[R_0>1.0]$, was considered excessive and therefore problematic since, ideally, the probability should be zero. The soundest mathematical solution appeared to be incorporation of the neglected correlations between the RLPs. This approach was not embraced, however, because of the complexity involved in attempting to estimate the uncertain correlations. Although, more simply, theoretical values of R_0 greater than 1.0 could be truncated at $R_0=1.0$ and $PDF[R_0]$ re-normalized and used to recalculate $E[R_0]$ and $CV[R_0]$, this approach was rejected because the re-normalized $PDF[R_0]$ would not be reasonably lognormal.

To help check that the underestimate was not resulting from the linear approximation inherent in the Taylor Series expression, the coefficient of variation terms in Eq 10 were supplemented by their cross product, $CV[RLP_j]^2CV[V_{i,j}]^2$, which is analogous to the difference between a Taylor Series approximation of a product and a mathematically exact value. This approach, after substitution, would yield:

$$Cov[lnRLV',lnRLV"]=\mathbf{E}_{j=1,K}\,\mathbf{E}_{i=1,Nj}\{E[V_{i,j}]^2E[RLP_j]^2E[R_{i,j}]\{(CV[RLP_j]^2+1)(CV[V_{i,j}]^2+1)-1\}\} \;/\; E[RLV']E[RLV"]$$

This alternative was rejected because it appeared to overestimate the covariance and correlation coefficient for cases where $E[R_0]$ was close to 1.0.

The adopted practical alternative was to use a correlation coefficient that would control the estimated probability that the true value of R_0 exceeded 1.0 to a suitably small value of ". The value of the correlation coefficient, symbolized $p_{"}$, that gave $P[R_0>1.0]="=0.01$ was used. This approach maintained the lognormality of $PDF[R_0]$—consistent with both L and F_0 being lognormal or RLV' and RLV" being lognormal. The approach was also consistent with the intuitive and theoretical notion that RLV" very close to RLV' (such that R_0 was close to 1.0) would be so highly correlated that $P[R_0>1.0]$ must be negligible. Finally, it was expected that formally including estimates of the neglected RLP correlations (which are uncertain) would result in refined estimates of $p_{lnRLV',lnRLV"}$ that matched estimates based on the more straightforward estimate $p_{"}$. Based on these considerations, for $P[R_0>1.0]>"=0.01$ by Eq 12, $p_{lnRLV',lnRLV"}$ was estimated as $p_{"}$ using the relationships documented in Appendix A; that is:

[13] $p_{lnRLV',lnRLV''} = p'' = (A - B^2)/2C^{1/2}$ (if $P[R_0>1.0]>''$)

where:

$$A = \ln\{(CV[\mathbf{E}(RLP*\mathbf{E}VR)]^2+1)*(CV[\mathbf{E}(RLP*V)]^2+1)\}$$
$$= \ln\{(CV[RLV'']^2+1)*(CV[RLV']^2+1)\}$$

$$B = \ln\{E[\mathbf{E}(RLP*V)]/E[[\mathbf{E}(RLP*\mathbf{E}VR)]* * \{(CV[\mathbf{E}(RLP*\mathbf{E}VR)]^2+1)/(CV[\mathbf{E}(RLP*V)]^2+1)\}^{1/2}\} / u_{1-''}$$
$$= \ln\{E[RLV'](CV[RLV'']^2+1)^{1/2} / E[RLV''](CV[RLV']^2+1)^{1/2}\} / u_{1-''}$$

$$C = \ln\{CV[\mathbf{E}(RLP*\mathbf{E}VR)]^2+1\}* \ln\{CV[\mathbf{E}(RLP*V)]^2+1\}$$
$$= \ln\{CV[RLV'']^2+1\}* \ln\{CV[RLV']^2+1\}$$

$u_{1-''}$ = the normal standard variate for a probability of 1-''

In summary, the correlation coefficient used to calculate $\mathbf{S}$ in Eqs 2 and 3, $p_{lnRLV',lnRLV''}$, was estimated using Eq 12 provided the resulting $P[R_0>1.0]<''=0.01$. Otherwise, Eq 13 was used. This approach was practical and considered to introduce negligible error in the estimates of R_0.

RLP for Deeper Floodplain Sediments

The expected value and coefficient of variation for the RLP of deeper sediments, RLP_{DS}, was estimated as follows from the Eq 14 (same as Eq 6 in Section B.2.2.1, and assuming no correlation between >and C_{DS} or C_S):

[14] $RLP_{DS} = >C_{DS}/C_S$

[15] $E[RLP_{DS}] = E[>]E[C_{DS}](1+CV[C_S]^2)\mathbf{S} / E[C_S]$

[16] $CV[RLP_{DS}] = \{(CV[>]^2+1)(CV[C_{DS}]^2+1)(CV[C_S]^2+1)\mathbf{S}^2 - 1\}^{1/2}$

where:

$\mathbf{S}$ = $\exp\{-p\{\ln(CV[C_{DS}]^2+1)\ln(CV[C_S]^2+1)\}^{1/2}\}$

p = Correlation coefficient between the natural log of C_{DS} and the natural log of C_S

= $p_{lnCsd,lnCs}$

Variables C_{DS} and C_S are, respectively, the metal (zinc) concentration of the deeper sediments and floodplain sediments having RLP=1. These sediment concentrations will be positively correlated. A correlation coefficient of $p_{lnCsd,lnCs}$=0.5 was used in the analysis; this represented a maximum entropy estimate for $p_{lnCsd,lnCs}$ uniformly distributed between 0.2 and 0.8, the expected limits of the correlation coefficient.

Non-exceedance Estimates of the Remediation Factor, R_0

Non-exceedance estimates of the true value of R_0 are denoted R_{0Pn}. Similar to other non-exceedance estimates, the relevant equations for R_{0Pn} are as follows:

[17] $R_{0Pn} = E[R_0] \exp\{u_{Pn} \ln(CV[R_0]^2+1)^{1/2}\} / (CV[R_0]^2+1)^{1/2}$

[18] $R_{090\%} = E[R_0] \exp\{1.28 \ln(CV[R_0]^2+1)^{1/2}\} / (CV[R_0]^2+1)^{1/2}$

[19] $R_{010\%} = E[R_0] \exp\{-1.28 \ln(CV[R_0]^2+1)^{1/2}\} / (CV[R_0]^2+1)^{1/2}$

A probable range of the potential remediation factor was defined by the limits $R_{010\%}$ (lower bound) and $R_{090\%}$ (upper bound). Theoretically, these approximate limits would have an 80 percent probability of containing the true value of R_0.

B.3.3.2 Source Depletion Natural Recovery Rate and Factor

Based on the relationship for the source decay rate expressed in the following Eq 1, the expected value and coefficient of variation of the decay rate, E[**$**] and CV[**$**], were estimated as follows:

[1] **$** = L / TEM'

[2] E[**$**] = $E[L](1+CV[TEM']^2)$**S** / E[TEM']

[3] CV[**$**] = $\{(CV[L]^2+1)(CV[TEM']^2+1)$**S**$^2(CV[$**9**$]^2+1) - 1\}^{1/2}$

where:

S = $\exp\{-p\{\ln(CV[L]^2+1)\ln(CV[TEM']^2+1)\}^{1/2}\}$
p = Correlation coefficient between the natural log of L and the natural log of TEM'
= $p_{lnL,lnTEM'}$ (later set to 0, implying Eqs 2 and 3 are mathematically exact even if L and TEM' are not lognormal)

For *L* and TEM' lognormally distributed, **$** is also lognormally distributed and Eqs 2 and 3 are mathematically exact. As discussed in Section B.2.2.2, PDF[**$**] represents both parameter uncertainty in **$**, including effects of natural variability, and model uncertainty in the source depletion factor $e^{-\$t}$ as it affects F(t). The term CV[**9**] represents additional model uncertainty as it affects F(t) through $e^{-\$t}$. An estimate of CV[**9**]=0.5 was used in the analysis. This resulted in the net uncertainty being about equally distributed between parameter uncertainty, due to CV[*L*] and CV[TEM'], and model uncertainty, due to CV[**9**].

Estimates for E[*L*] and CV[*L*] and E[TEM'] and CV[TEM'] are given in the following paragraphs. Since the estimates for *L* and TEM' are independent of each other, $p_{\ln L,\ln TEM'}$ was set to zero in the analysis. This lack of correlation in the *estimates* should not be confused with the positive correlation that must exist in the *true* values of *L* and TEM', and is otherwise inherent in the data used to make the estimates. To the extent there was (positive) correlation between the estimates *L* and TEM', it would decrease both E[**$**] and CV[**$**].

Yearly Average Pre-remediation Load, L

L is the total pre-remediation load summed over a water year and estimated from historical stream monitoring data. Assuming statistically unbiased data, the expected value and coefficient of variation of *L*, E[*L*] and CV[*L*], were estimated from the E[L] and CV[L] of the historical data, as follows, where n represents a measure of the effective number of historical measurements used to estimate E[L] and CV[L]; that is:

[4] $E[L] = E[E[L]*365.25]$
$= E[L]*365.25$

[5] $CV[L] = CV[E[L]]$
$= CV[L] / n^{1/2}$

Eqs 4 and 5 assume that E[L] is a statistically unbiased estimate and that the uncertainty in *L* is equal to the statistical uncertainty in E[L]. Note that multiplying L (lb/d) by 365.25 (days/year) converts *L* to pounds of zinc per year (lb/year).

Pre-remediation Total Contaminant (Zinc) Mass, TEM'

The pre-remediation effective total metal (zinc) mass that is available for loading, TEM', is expressed by Eq 6. The expected value and coefficient of variation of TEM', E[TEM'] and CV[TEM'], were estimated in terms of the equivalent total relative source volume, RLV'=**E**(RLP*V), as follows:

[6] TEM' = $\mathbf{E}(RLP*V)\,\mathbf{(}\,C_S$
= $RLV'\,\mathbf{(}\,C_S$

[7] E[TEM'] = $E[\mathbf{E}(RLP*V)]E[\mathbf{(}]E[C_S]\ \mathbf{S}$
= $E[RLV']E[\mathbf{(}]E[C_S]\ \mathbf{S}$

[8] CV[TEM'] = $\{(CV[\mathbf{E}(RLP*V)]^2+1)(CV[\mathbf{(}]^2+1)\,)(CV[C_S]^2+1)\mathbf{S}^2-1\}^{1/2}$
= $\{(CV[RLV']^2+1)(CV[\mathbf{(}]^2+1)\,)(CV[C_S]^2+1)\mathbf{S}^2-1\}^{1/2}$

where:

$\mathbf{S}$ = $\exp\mathbf{\{}2p\{\ln(CV[\mathbf{E}(RLP*V)]^2+1)\ln(CV[C_S]^2+1)\}^{1/2}\mathbf{\}}^{1/2}$
= $\exp\mathbf{\{}p\{\ln(CV[RLV']^2+1)\ln(CV[C_S]^2+1)\}^{1/2}\mathbf{\}}$

p = Correlation coefficient between the natural log of $RLV'=\mathbf{E}(RLP*V)$ and the natural log of C_S
= $p_{lnRLV',lnCs}$

Eqs 7 and 8 assume that there is no correlation between the average unit weight, $\mathbf{(}$, of floodplain sediments and either the equivalent total source volume, $RLV'=\mathbf{E}(RLP*V)$, or the volumetric average concentration, C_S, of the sediments. That is, only RLV' and C_S are assumed to be potentially correlated. The correlation is negative because the volumetric average concentration of metal (zinc) decreases as the total affected volume increases. A correlation coefficient of $p_{lnRLV',lnCs}=-0.5$ was used in the analysis; this represented a maximum entropy estimate for $p_{lnRLVi,lnCs}$ uniformly distributed between –0.2 and –0.8, the expected limits of the correlation coefficient.

Source Depletion Factor

$E[e^{-\$t}]$ and $CV[e^{-\$t}]$, the expected value and coefficient of variation of the source depletion factor $e^{-\$t}$, were estimated at selected times, t, based on the following fundamental relationships, where PDF[$\$$], the probability density function of $\$$, was assumed to be lognormal. Recall that PDF[$\$$] represents both parameter uncertainty in $\$$ and model uncertainty in $e^{-\$t}$ as it affects F(t).

[9] $E[e^{-\$t}]$ = $_{0\text{ to }4}\ e^{-\$t}\ PDF[\$]\ d\$$

[10] $CV[e^{-\$t}]$ = $V[e^{-\$t}]^{1/2} / E[e^{-\$t}]$

= $\{\ _{0\text{ to }4}\ (e^{-\$t})^2\ PDF[\$]\ d\$ - E[e^{-\$t}]^2\}^{1/2} / E[e^{-\$t}]$

For a given time, t, the integrals on the right sides of Eqs 9 and 10 were numerically approximated as follows. The values of **$** were calculated from the normal standard variate (for a lognormal distribution) for the cumulative density function of **$**, CDF[**$**], as determined by the estimates of E[**$**] and CV[**$**].

One hundred and one (101) values of **$** were calculated from CDF[**$**], starting with a cumulative probability of 0.005 (0.5 percent). An initial cumulative probability increment of 0.005 was used for the second value of **$**, yielding a cumulative probability of 0.01 (1 percent). For the next 98 increments of **$** the CDF was successively increased in 0.01 (1 percent) probability increments to a maximum cumulative probability of 0.99 (99 percent). The last (101st) value of **$** was calculated for a cumulative probability of 0.995 (99.5 percent). The estimates were thus limited to the central 0.99 (99 percent) of the range of potential **$** values, eliminating theoretically potential extreme values at the lower tail (below 0.005 or 0.5 percent) and the upper tail (above 0.995 or 99.5 percent). Realistically, these theoretical extremes have no practical potential significance.

Using this approach, the integral of PDF[**$**]d**$** equaled the cumulative probability increments: 0.005 over the first and last increment and 0.01 over each of the other 98 increments (a total of 100 increments between the lowest cumulative probability, 0.005, and the highest cumulative probability, 0.995). The average **$** over each CDF increment was calculated from the **$** for the CDF at the beginning of the increment and the **$** for the CDF at the end of the increment. The value of $e^{-\$t}$ was calculated for each average **$** over the CDF increment.

The integrals in Eqs 9 and 10 were approximated as the sums, respectively, of $e^{-\$t}$PDF[**$**]d**$** and $(e^{-\$t})^2$PDF[**$**]d**$** over the 100 CDF increments normalized by the sum of PDF[**$**]d**$** over those increments. The normalizing sum was equal to 0.99, since the theoretical **$** values having cumulative probabilities below 0.005 and above 0.995 were considered to have no practical significance. Normalizing by the sum of PDF[**$**]d**$** over all potential **$** values was required to assure that the probability of all potential values summed to one.

B.3.3.3 Approximations for E[R(t)] and CV[R(t)]

This section documents two useful approximations to simplify estimates of E[R(t)] and CV[R(t)]. The approximations are based on results for the upper basin using the equations for E[R(t)] and CV[R(t)] documented to this point and input values documented in Appendix C. The approximations were used in the analysis to simplify estimates of E[R(t)] and CV[R(t)] for the zinc loading contribution from the lower basin.

The upper basin estimates were made for the six remedial alternatives documented in the FS, using the input summarized in Appendix C, for post-remediation zinc loadings at the SFCDR at Pinehurst (SF271). The estimates of E[R(t)] and CV[R(t)] for the upper basin indicated the following.

First, there was minimal practical difference between E[R(t)] estimated using $E[R_0]$ and using point estimates of R_0 (the difference ranged from approximately –0.4 percent to +2.3 percent). It was therefore considered generally adequate to estimate E[R(t)] from point estimates of R_0 in the absence of formally estimating $E[R_0]$ (note $E[e^{-\$t}]$ would still be estimated using the method of Section B.3.3.2). This approximation from the upper basin analysis—estimating E[R(t)] from point estimates of R_0—appeared to be generally valid and was therefore used for the lower basin, where it saved considerable effort.[21]

Second, there was strong correlation between E[R(t)] and CV[R(t)] for the upper basin estimates. The correlation was formalized as follows:

$$
[1] \quad CV[R(t)] = \begin{cases} \exp\{0.19 - 3.3E[R(t)]\} & E[R]<0.85 \\ \exp\{10 - 15E[R(t)]\} & 0.85 \leq E[R]<1.0 \text{ see footnote}^{22} \\ 0 & E[R]=1.0 \text{ (no action at time } t=0) \end{cases}
$$

Eq 1 was developed from the theoretical relationship between E[R] and CV[R] that would assure that P[R>1] would not exceed 0.001 for E[R]<0.85 and would not exceed approximately 0.01 for $R \geq 0.85$. For the upper basin alternatives, including no action, CV[R(t)] from Eq 1 was consistent with the estimates of CV[R(t)] made using the more formal, and much more complex, methods of the previous sections.

The differences between CV[R(t)] estimated from Eq 1 and CV[R(t)] estimated by the methods of previous sections were both well within a plausible range of realistic values of CV[R(t)] and had relatively minor effect on the resulting probabilistic estimates. For these reasons and because of its general theoretical and practical basis, Eq 1 was considered a valid and reasonable approximation for estimating CV[R(t)] for the lower basin, with further savings of effort.

These two simplifying approximations from the upper basin analysis—estimating E[R(t)] from point estimates of R_0 and approximating CV[R(t)] using Eq 1—not only saves effort but could also potentially reduce controversy. That is, they provide necessary probabilistic parameters for any given remedial alternative *without* having to make estimates of the individual coefficients of

[21] As noted, $E[e^{-\$t}]$ was estimated using the method of Section B.3.3.2.
[22] If necessary, for E[R(t)] close to one, CV[R(t)] was simply adjusted so that P[R>1]<0.01.

variation for the relative loading potentials, remediated volumes, and remedial action effectiveness factors that are otherwise required to estimate E[R(t)] and CV[R(t)] by the methods of previous sections.

The two approximations thus suggest that, in practical terms, the probabilistic results are not particularly sensitive to credible variations in these coefficients of variation.[23] Importantly, this apparent lack of practical sensitivity could help answer any appropriate question, or potential controversy, over exactly what coefficient of variation estimates to use at this time (see the preceding footnote).

B.3.4 LOAD RATIO, Lr

Based on the relationship for the load ratio, Lr, expressed in the following Eq 1, the expected value and coefficient of variation of the load ratio, E[Lr(t)]and CV[Lr(t)], were estimated as follows:

[1] $Lr(t) = F(t) / C_L$

[2] $E[Lr(t)] = E[F(t)](1+CV[C_L]^2)\mathbf{S} / E[C_L]$

[3] $CV[Lr(t)]] = \{(CV[F(t)]^2+1)(CV[C_L]^2+1)\mathbf{S}^2 - 1\}^{1/2}$

where:

$\mathbf{S} = \exp\{-p\{\ln(CV[F(t)]^2+1)\ln(CV[C_L]^2+1)\}^{1/2}\}$

p = Correlation coefficient between the natural log of F(t) and the natural log of C_L

$= p_{lnF,lnCL}$

Eqs 2 and 3 assume that F(t) and C_L are both lognormally distributed, such that Lr(t) is also lognormally distributed (the quotient of two lognormal distributions is also lognormal). Loadings are generally expected to maintain lognormality after remediation, since the basic physics and mathematics of the random processes leading to the variability in the observed

[23] In this context, "credible" variations would have to be consistent with the relatively high degree of uncertainty that truly prevails in any current estimate of potential future remedial performance. Over the long term, assuming the evolutionary addition of significant new information, the inherent level of uncertainly will decrease, eventually making Eq 1 obsolete and inappropriate.

loadings[24] would not be altered by remedial action. As will be discussed, the loading capacity C_L as expressed in the TMDL document (USEPA 2000c) and as estimated from AWQC and associated discharges, is adequately lognormal for practical purposes. Note that Eqs 1, 2, and 3 also apply where time is implicit—in which case Lr(t)=Lr and F(t)=F.

The future correlation between lnF(t) and $\ln C_L$, measured by $p_{\ln F,\ln CL}$, is expected to be very high. This expectation is based on an almost perfect correlation (p=1.0) between $\ln C_L$ and lnQ and a virtually certain high future correlation between lnF(t) and lnQ, just as there has been historically between discharge and loading (which, being a function of discharge, induces correlation). In addition, as further discussed in Section B.3.4.1, an independent statistical analysis of the zinc concentrations, water hardnesses, and discharge data corresponding to that used in developing the TMDL loading capacities for SF271 (USEPA 2000c) showed a correlation coefficient of 0.95 between the natural logs of the zinc loadings (computed as the product of concentration and discharge) and the equivalent loading capacities (computed as the product of the zinc AWQC(H) and discharge). Consistent with this information, a value of $p_{\ln F,\ln CL}$=0.9 was used in the analysis.

An appropriate estimate for $p_{\ln F,\ln CL}$ should also consider the correlation between loading capacities, and loading in general, and AWQC. Although very high, that correlation is not perfect, since perfect correlation would require hardness H, and thus AWQC, to be perfectly correlated with discharge Q. From a practical standpoint, any actual or potential weakness in the correlation between loading capacity and AWQC, or imperfections in the lognormality of Lr(t), could be reflected in a reduced estimate of $p_{\ln F,\ln CL}$. From that viewpoint, $p_{\ln F,\ln CL}$=0.9 could be considered "optimistically" high.[25] However, any potential optimism was balanced by a conservative interpretation of the loading capacity under low flow conditions, as discussed in Section B.3.4.1, which further justified the potentially high value of $p_{\ln F,\ln CL}$=0.9 used in the analysis.

Recall that the load ratio Lr(t) is also numerically equivalent to a AWQC ratio, defined as the ratio of post-remediation concentration, C"(t), to AWQC. That is:

$$\begin{aligned} Lr(t) &= F(t) / C_L \\ &= C''(t) / AWQC(Q) \end{aligned}$$

[24] That is, successive random dilutions, proportional effects, and the central limit theorem, as noted in Appendix A and detailed in Ott (1995).
[25] A reduced value of $p_{\ln F,\ln CL}$ would increase the estimated times to meet AWQC for all alternatives.

Lr(t) is *both* the number of times the loading exceeds the loading capacity, C_L, *and* the number of times the concentration exceeds the AWQC. Lr=1 is considered the *target load ratio* for compliance with AWQC.

Non-exceedance Estimates of the Load Ratio

Non-exceedance estimates of the true value of Lr(t) are denoted $Lr(t)_{Pn}$. The subscript "Pn" is the non-exceedance probability, the probability that the *true* value of Lr(t) *does not exceed* the estimate $Lr(t)_{Pn}$. Equivalently, Pn is the probability that $Lr(t)_{Pn}$ *overestimates* the true value of Lr(t) and is thus not an underestimate. The complement of the non-exceedance probability, 1-Pn, is the probability that the true value of Lr(t) does exceed the estimate $Lr(t)_{Pn}$ or, equivalently, that $Lr(t)_{Pn}$ underestimates the true value of Lr(t). Non-exceedance estimates $Lr(t)_{Pn}$ are given by the following equations, where all variables and associated concepts have been discussed:

[4] $$Lr(t)_{Pn} = E[Lr(t)] \exp\{u_{Pn} \ln(CV[Lr(t)]^2+1)^{1/2}\} / (CV[Lr(t)]^2+1)^{1/2}$$

[5] $$Lr(t)_{90\%} = E[Lr(t)] \exp\{1.28 \ln(CV[Lr(t)]^2+1)^{1/2}\} / (CV[Lr(t)]^2+1)^{1/2}$$

[6] $$Lr(t)_{10\%} = E[Lr(t)] \exp\{-1.28 \ln(CV[Lr(t)]^2+1)^{1/2}\} / (CV[Lr(t)]^2+1)^{1/2}$$

A probable range of potential ratios was defined by the limits $Lr(t)_{10\%}$ (lower bound) and $Lr(t)_{90\%}$ (upper bound). Theoretically, these approximate limits would have an 80 percent probability of containing the true value of Lr(t).

B.3.4.1 Loading Capacity, C_L

$E[C_L]$ and $CV[C_L]$ were approximated using a linear least-squares regression analysis and interpretation of the loading capacity over time. Using the probabilistic interpretation of the TMDL flow tiers discussed in the following, the normal standard variate corresponding to the percentile average daily flow of each flow tier was regressed against the natural log of the corresponding loading capacity.

The loading capacities correspond to flow tiers representing the 10th, 50th, and 90th percentile average daily flow, as developed in USEPA 2000c. For analysis, because the 7Q10 flow does not have a flow tier percentile in the TMDL document, the 7Q10 load capacity was assigned a 1st percentile average daily flow. This overestimated the true frequency of 7Q10 flows and underestimated the corresponding loading capacity at the 1st percentile average daily flow. Assuming the correspondence between the 7Q10 load capacity and the 1st percentile average daily flow conservatively supports the probabilistic interpretation of the flow tiers, as discussed next.

The flow tier average daily flow percentiles were interpreted as the probabilities that the loading capacity on any given day (C_L) would be at or below the loading capacity corresponding to the given percentile. That is, C_L has a 90 percent probability of being at or below the loading capacity for the 90th percentile flow tier, a 50 percent probability of being at or below the loading capacity for the 50th percentile flow tier, and so on for the 10th and 1st percentile flow tiers. The resulting cumulative probability distribution of C_L, CDF{C_L}, was postulated for the purpose of the analysis to be a continuous lognormal distribution.

Using this probabilistic interpretation of the TMDL flow tiers, the normal standard variate (u) corresponding to the percentile average daily flow of each flow tier was regressed against the natural log of the corresponding loading capacity. The regression used the linear model $u=b+m*\ln C_L$, where m and b are the slope and intercept, respectively, of the regression line.

Specifically for the SFCDR at Pinehurst, SF271, the resulting r^2 for the regression[26] was 1.00 using the 10th, 50th, and 90th percentile flow tiers (n=3) and 0.98 when a 1st percentile was used for the 7Q10 load capacity (n=4). For the CDR at Harrison, LC60, the corresponding r^2's were 0.98 (n=3) and 0.95 (n=4). Values of r^2 for other TMDL sites in the basin were all above 0.96 (n=3), and 0.94 (n=4) when the 7Q10 values were included.

The regression results were interpreted to adequately support the postulate that loading capacity, C_L, can be reasonably approximated as a lognormal distribution for the practical purpose of the probabilistic analysis documented in this technical memorandum. It is reiterated that CDF{C_L} as used in the analysis is a *continuous* distribution between and beyond the four TMDL flow tiers. Besides its analytical practicality, a continuous distribution supports the loading capacities being a metric for compliance with AWQC, and is considered appropriately consistent with the TMDL document (USEPA 2000c). Although alternative (i.e., step-wise discrete) formulations are possible, they would add complexity and seriously detract from the analysis, particularly for AWQC estimates.

Importantly, for all sites, the regression-predicted estimate of C_L corresponding to a 1 percent cumulative probability of occurrence was *less* than the 7Q10 loading capacity. This indicates that the analysis is conservative (underestimates C_L) under extreme low flow conditions, which is considered the most critical condition for AWQC and ecological risk.

[26] The square of the sample correlation coefficient, r^2, measures the strength of the relationship between any variables x and y in the regression y=mx+b; r^2 may be interpreted as the proportion of total variability measured by the regression. The square root of r^2 is also an estimate of the correlation coefficient $p_{x,y}$ as used in the probability equations.

Given the regression results, $E[C_L]$ and $CV[C_L]$ were estimated from the slope and intercept of the regression line using the following relationships for a lognormal variate, as documented in Appendix A:

[1] $E[C_L] = \exp[-b/m + 1/2m^2]$

[2] $CV[C_L] = \{\exp[1/m^2] - 1\}^{1/2}$

The forgoing probabilistic interpretation of the loading capacity is supported by independent statistical analysis of the measured water hardness, H, and discharge, Q, data used in developing the TMDL loading capacities for the SFCDR at Pinehurst (USEPA 2000c, Appendix I, SF271). These data were analyzed together with the measured zinc concentrations, [Zn], corresponding to the H and Q data (for a total of 27 concurrent measurements of H, Q, and [Zn], including two duplicates that used averaged values for an effective n=25). The analysis computed zinc loadings as the product of measured concentration and discharge; equivalent loading capacities were computed as the product of the zinc AWQC(H), calculated from measured hardness, and measured discharge. The analysis showed the following principal results:

- Both the AWQC(H) and the equivalent loading capacities were lognormally distributed with respective r^2's of 0.94 and 0.97.
- The ratios of zinc loadings to equivalent loading capacities were equal to the ratios of zinc concentrations to AWQC(H) (i.e., the load ratio Lr was the same for loading and concentrations, as generally assumed) and were lognormally distributed with an r^2 of 0.95.
- Zinc concentrations, loadings, and hardness were lognormally distributed with respective r^2's of 0.960, 0.962, and 0.93.
- The correlation coefficient between the natural logs of hardness and discharge and between the natural logs of AWQC(H) and discharge were both equal to 0.96.
- The correlation coefficient between the natural logs of the zinc loadings and the equivalent loading capacities was 0.95.

These results support the assumptions that loading capacities, AWQC, and load ratios are lognormally distributed, and that load ratios are the same for loadings and concentrations. The results also support a high value of $p_{lnF,lnCL}$, as discussed in Section B.3.4.

B.3.4.2 Probability True Load Ratio Does Not Exceed a Multiple of the Loading Capacity or AWQC

The probability that the true (but uncertain) load ratio does not exceed a given multiple X of the loading capacity or AWQC (at any time on average over a water year) was estimated from the probabilistic parameters of Lr(t). The estimate was based on the following Eq 1, as documented in Appendix A, Section A.3.1. Eq 1 quantifies the probability that the true post-remediation load F(t) will be less than X times the loading capacity, and, equivalently, the probability that the post-remediation concentrations C"(t) will be less than X times the AWQC.

[1] $P[Lr(t) \leq X] = CDF_{Lr(t)}\{u\}$
$= CDF_U\{u\}$

where:

$CDF_{Lr(t)}\{u\}$ = Standard normal cumulative distribution function of Lr(t) evaluated at u

u = Standard normal variate for a lognormal variable
$= \ln\{X(CV[Lr(t)]^2+1)^{1/2}/E[Lr(t)]\}\ln(CV[Lr(t)]^2+1)^{-1/2}$

X = Target multiple of loading capacity or AWQC (X>0)

Values of $CDF_U\{u\}$ were obtained from the spreadsheet function NORMSDIST(u) in MS Excel 97 (values could also be obtained from statistical tables for standard normal variates, e.g., Natrella 1966).

B.3.4.3 Load Ratio Estimate Required to Achieve a Target True Load Ratio With Specified Probability

For any given time t, the expected value of the true load ratio required to achieve a specific target value X with specified probability, Ps, can be calculated from B.3.4.2 Eq 1 for a given value of CV[Lr]. The standard normal variate in B.3.4.2 Eq 1 is that for probability Ps, u_{Ps}; that is, from B.3.4.2 Eq 1:

[1] $u_{Ps} = \ln\{X(CV[Lr(t)]^2+1)^{1/2}/E[Lr(t)]\}\ln(CV[Lr(t)]^2+1)^{-1/2}$

The required expected value of Lr, symbolized $E[Lr]_{Ps}$, is estimated from Eq 1 by setting Lr=X, $E[Lr]_{Ps}$=E[Lr(t)], and CV[Lr]=CV[Lr(t)] and solving for $E[Lr]_{Ps}$; the result is as follows:

$$[2] \qquad E[Lr]_{Ps} = \frac{Lr\,(CV[Lr]^2+1)^{1/2}}{\exp\{u_{Ps}\,\ln(CV[Lr]^2+1)^{1/2}\}}$$

where:

Lr	=	Specific target value of the true Lr (e.g., Lr=1)
u_{Ps}	=	the normal standard variate for a probability of Ps
$CV[Lr]^2+1$	=	$(CV[L]^2+1)(CV[R]^2+1)(CV[C_L]^2+1)\mathbf{S}_{lnL,lnR}{}^2\mathbf{S}_{lnF,lnCL}{}^2$
$\mathbf{S}_{lnL,lnR}$	=	$\exp\{p\{\ln(CV[L]^2+1)\ln(CV[R]^2+1)\}^{1/2}\}$
p	=	Correlation coefficient between the natural log of L and the natural log of R
	=	$p_{lnL,lnR}$ (set at –0.5, as discussed)
$\mathbf{S}_{lnF,lnCL}$	=	$\exp\{-p\{\ln(CV[F]^2+1)\ln(CV[C_L]^2+1)\}^{1/2}\}$
	=	$\exp\{-p\{\ln\{(CV[L]^2+1)(CV[R]^2+1)\mathbf{S}_{lnL,lnR}{}^2\}\ln(CV[C_L]^2+1)\}^{1/2}\}$
p	=	Correlation coefficient between the natural log of F and the natural log of C_L
	=	$p_{lnF,lnCL}$ (set at 0.9, as discussed)

The quantity $E[Lr]_{Ps}$ together with CV[Lr] would have a probability of Ps of underestimating a true load ratio equal to Lr. Used in a parametric or sensitivity analysis, these estimates are useful for identifying values of the remediation factor required to achieve a given target value of Lr—in particular, Lr=1 (i.e., loading equal to the loading capacity and concentration equal to AWQC) or Lr>1 for interim benchmarks. This is further developed in the next section for the remediation factor.

B.3.4.4 Remediation Factor Required to Achieve a Target True Load Ratio With Specified Probability

The expected value of the remediation factor required to achieve a specific target true value of the load ratio Lr with specified probability Ps is symbolized $E[R]_{Ps}$. The target load ratio can be any value, including Lr=1 for meeting AWQC, or Lr>1 for meeting interim benchmarks (e.g., Lr=3, 7, 10, or 20). The target load ratio can be for any time $t \geq 0$.

$E[R]_{Ps}$, the expected value of the remediation factor required to achieve a specific target load ratio with specified probability Ps, can be calculated as follows. Starting with B.3.4 Eq 2 for E[Lr] and substituting the left-hand side of B.3.2 Eq 2 for E[F] results in the following:

$$[1] \quad E[Lr] = E[F](1+CV[C_L]^2)\mathbf{S}_{lnF,lnCL} / E[C_L]$$
$$= E[L]E[R]\mathbf{S}_{lnL,lnR}(1+CV[C_L]^2)\mathbf{S}_{lnF,lnCL} / E[C_L]$$

Solving for E[R]:

$$[2] \quad E[R] = E[Lr]E[C_L] / E[L](1+CV[C_L]^2)\mathbf{S}_{lnL,lnR}\,\mathbf{S}_{lnF,lnCL}$$

Equating $E[Lr]_{Ps}$ from B.3.4.3 Eq 2 with E[Lr] in Eq 2 results in the desired quantity $E[R]_{Ps}$, the expected value of the remediation factor required to achieve a specific target true value of the load ratio with specified probability; that is:

$$[3] \quad E[R]_{Ps} = \frac{E[Lr]_{Ps}\,E[C_L]}{E[L](CV[C_L]^2+1)\mathbf{S}_{lnL,lnR}\,\mathbf{S}_{lnF,lnCL}}$$

An equivalent version of Eq 3 can be formulated by substituting first the right hand side of B.3.4.3 Eq 2 for $E[Lr]_{Ps}$ followed by $(CV[Lr]^2+1)$ from that same equation[27]; these substitutions yield:

$$[4] \quad E[R]_{Ps} = \frac{Lr\,(CV[Lr]^2+1)^{1/2}\,E[C_L]}{E[L](CV[C_L]^2+1)\exp\{u_{Ps}\ln(CV[Lr]^2+1)^{1/2}\}\mathbf{S}_{lnL,lnR}\mathbf{S}_{lnF,lnCL}}$$

$$= \frac{Lr\,E[C_L]\,\{(CV[L]^2+1)(CV[R]^2+1)\}^{1/2}}{E[L](CV[C_L]^2+1)^{1/2}\exp\{u_{Ps}\ln\{(CV[L]^2+1)(CV[R]^2+1)(CV[C_L]^2+1)\mathbf{S}_{lnL,lnR}{}^2\mathbf{S}_{lnF,lnCL}{}^2\}^{1/2}\}}$$

Eqs 3 and 4 are equivalent. Given estimates of the current loading (i.e., E[L] and CV[L]) and the loading capacity (i.e., $E[C_L]$ and $CV[C_L]$), the expected value of the remediation factor given by Eqs 3 or 4, $E[R]_{Ps}$, is the value required to achieve $E[Lr]_{Ps}$ for a target true load ratio Lr with probability Ps. For the case of Lr=1 (i.e., loading equal to the loading capacity and concentrations equal to AWQC), the remediation factor given by Eqs 3 or 4 is the value required to reduce current loading to the loading capacity and achieve AWQC. Simultaneously solving

[27] i.e., $(CV[Lr]^2+1) = (CV[L]^2+1)(CV[R]^2+1)(CV[C_L]^2+1)\mathbf{S}_{lnL,lnR}{}^2\mathbf{S}_{lnF,lnCL}{}^2$

for $E[R]_{Ps}$ with B.3.3.3 Eq 1 for CV[R] as a function of E[R] provides consistent estimates of the remediation factor required to achieve the loading capacity and AWQC. The same holds for any value of the target load ratio as may be used to establish interim benchmarks having Lr>1.

Approximate Scaling of* $\mathbf{E[R]_{Ps}}$ *Between Stream Monitoring Locations

The required remediation factor to achieve a target Lr, $E[R]_{Ps}$, at locations not having loading capacities (i.e., $E[C_L]$ and $CV[C_L]$) can be estimated from $E[R]_{Ps}$ for locations that do. In this case, the reference location, designated by "1," has $E[R_1]_{Ps}$ based on $E[C_{L1}]$ and $CV[C_{L1}]$ and the second location, designated "2," is where $E[R_2]_{Ps}$ is to be estimated without explicit estimation of $E[C_{L2}]$ and $CV[C_{L2}]$. The target Lr is the same for both stations: $Lr_1 = Lr_2$. The desired relationship starts with expanding the ratio of Lr_2/Lr_1, where Lr=RC/AWQC; that is:

$$[5] \quad Lr_2/Lr_1 = \{R_2C_2/AWQC_2\} / \{R_1C_1/AWQC_1\} = (R_2/R_1)(C_2/C_1)(AWQC_1/AWQC_2) = 1$$

Both R_2 and R_1 are remediation factors required to achieve a specific target true value of the load ratio (e.g., Lr=1, 3, 7, 10 or 20). Solving for R_2/R_1 yields (for $Lr_2=Lr_1$):

$$[6] \quad R_2/R_1 = (C_1/C_2)(AWQC_2/AWQC_1) = (C_1/AWQC_1)(AWQC_2/C_2)$$

Taking expected values:

$$[7] \quad E[R_2/R_1] = E[(C_1/AWQC_1)(AWQC_2/C_2)]$$

Expansion of the expected values on both sides of Eq 7 is needed. Starting with the left-hand side, the ratio R_2/R_1 will be lognormally distributed with expected value:

$$[8] \quad E[R_2/R_1] = E[R_2](1+ CV[R_1]^2)\mathbf{S}_{R1,R2} / E[R_1]$$

where:

$\mathbf{S}_{R1,R2}$ = $\exp\{-p\{\ln(CV[R_1]^2+1)\ln(CV[R_2]^2+1)\}^{1/2}\}$

p = Correlation coefficient between the natural log of R_1 and the natural log of R_2, $p_{lnR1,lnR2}$

For $Lr_2=Lr_1$, R_2 and R_1 will be highly correlated such that $p_{lnR1,lnR2}=1$ to a close approximation. If $E[R_2]$ and $E[R_1]$ are close enough that $CV[R_1]=CV[R_2]$, then the quantity $(1+CV[R_1]^2)\mathbf{S}_{R1,R2}$ will be approximately 1.0, and Eq 8 will reduce to the following:

[9] $E[R_2/R_1] = E[R_2]/E[R_1]$

Eq 9 says that the expected value of the ratio R_2/R_1 is equal to the ratio of the expected values; the equality is exact for $p_{lnR1,lnR2}=1$ and $CV[R_1]=CV[R_2]$ and approximate otherwise. Examination of $E[R]_{Ps}$ results for SF271, CC288, and NM305 for Ps values between 0.01 and 0.90 and Lr values between 1 and 20 show $p_{lnR1,lnR2}>0.99$. In general, CV[R] increases as E[R] decreases (see Section B.3.3.3), so that the difference in $CV[R_1]$ and $CV[R_2]$ will increases with the difference in $E[R_1]$ and $E[R_2]$, causing the approximation $E[R_2/R_1]=E[R_2]/E[R_1]$ to deviate somewhat from exactness at very low and very high R_2/R_1 ratios. The deviations are unlikely to be of practical significant if both locations require low values of $E[R]_{Ps}$, which is the general rule throughout the upper basin. For practical purposes, Eq 9 is therefore considered generally valid in the upper basin.

Next, the expected value of the right-hand side of Eq 7 is developed as follows.

[10] $E[(C_1/AWQC_1)(AWQC_2/C_2)]=$

$= \{E[C_1/AWQC_1]*E[AWQC_2/C_2]\}\mathbf{S}_{1,2}$
$= \{E[C_1](1+CV[AWQC_1]^2)\mathbf{S}_{1,1}/E[AWQC_1]*E[AWQC_2](1+CV[C_2]^2)\mathbf{S}_{2,2}/E[C_2]\}\mathbf{S}_{1,2}$

where:

$\mathbf{S}_{1,2}$ = $\exp\{p_{1,2}\{\ln(CV[C_1/AWQC_1]^2+1)\ln(CV[AWQC_2/C_2]^2+1)\}^{1/2}\}$
$\mathbf{S}_{1,1}$ = $\exp\{-p_{1,1}\{\ln(CV[C_1]^2+1)\ln(CV[AWQC_1]^2+1)\}^{1/2}\}$
$\mathbf{S}_{2,2}$ = $\exp\{-p_{2,2}\{\ln(CV[C_2]^2+1)\ln(CV[AWQC_2]^2+1)\}^{1/2}\}$
$p_{1,2}$ = Correlation coefficient between the natural log of $(C_1/AWQC_1)$ and the natural log of $(AWQC_2/C_2)$

$p_{1,1}$ = Correlation coefficient between the natural log of C_1 and the natural log of $AWQC_1$

$p_{2,2}$ = Correlation coefficient between the natural log of C_2 and the natural log of $AWQC_2$

Following Eq 7 and substituting the right-hand sides of Eqs 9 and 10 results in the following relationship, where R_2 and R_1 are remediation factors required to achieve a specific target true value of the load ratio Lr.

[11] $E[R_2]/E[R_1] =$

$$= \{E[C_1](1+CV[AWQC_1]^2)^{\sigma_{1,1}}/E[AWQC_1]\}\{E[AWQC_2](1+CV[C_2]^2)^{\sigma_{2,2}}/E[C_2]\}^{\sigma_{1,2}}$$

Replacing $E[R_2]$ and $E[R_1]$ with $E[R_2]_{Ps}$ and $E[R_1]_{Ps}$ (recognizing that $E[E[R]_{Ps}] = E[R]_{Ps}$) gives:

[12] $E[R_2]_{Ps}/E[R_1]_{Ps} =$

$$= \{E[C_1](1+CV[AWQC_1]^2)^{\sigma_{1,1}}/E[AWQC_1]\}\{E[AWQC_2](1+CV[C_2]^2)^{\sigma_{2,2}}/E[C_2]\}^{\sigma_{1,2}}$$

Eq 12 can be simplified to a more practical form because metal concentrations and AWQC(H)[28] are generally highly temporally and spatially correlated, particularly for zinc in the upper basin. These correlations are discussed as follows:

- First, temporal correlations between metal concentrations and AWQC at the same location (over time) are positive and generally very high. In particular, using the same sample data as used for the TMDL (USEPA 2000c), sample correlation coefficients[29] for ln[dZn] and lnAWQC(H) were, respectively, 0.911, 0.949, and 0.965 for NM305, CC288, and SF271 (NM305 is at the mouth of Ninemile Creek; CC288 is at the moth of Canyon Creek; and SF271 in on the SFCDR at Pinehurst). This means that, over time, high (and low) zinc concentrations coexist with high (and low) hardness, and thus high (and low) AWQC(H).

[28] AWQC(H) means AWQC are functions of hardness H.
[29] The correlation coefficient $p_{ln[dZn],lnAWQC(H)}$ is estimated as Sqrt{r^2} from the linear regression lnAWQC(H) = m*ln[dZn] + b; the zinc concentrations are for samples corresponding to the hardness H and discharge Q data used for the TMDL.

- Second, metal concentrations and AWQC(H) at the same location are both highly (negatively) correlated with discharge Q. In particular, using the zinc TMDL sample data, sample correlation coefficients for lnQ and ln[dZn] were, respectively, -0.762, -0.881, and -0.927 for NM305, CC288, and SF271; the sample correlation coefficients for lnQ and lnAWQC(H) were, respectively, -0.748, -0.973, and -0.962.

- Third, discharge Q is highly spatially correlated *between* locations (the spatial correlations exist within time increments generally on the order of a few days during high flow events and perhaps weeks during low flow periods). In particular, the correlation coefficients of lnQ ($p_{lnQ1,lnQ2}$) between NM305, CC288, and SF271 exceeded 0.92 for daily discharge during water year 1999. The high Q-correlations mean that metal concentrations and AWQC(H) are generally highly temporally correlated between stations. Therefore, the ratios of $C_1/AWQC_1$ and $AWQC_2/C_2$ between locations are generally highly *negatively* correlated.

- Further, the coefficients of variation for concentration, CV[C], and AWQC(H), CV[AWQC], are of generally similar magnitude, at least for zinc in the upper basin. In particular, using the same sample data as used for the zinc TMDL: CV[Zn]=0.46 and CV[AWQC]=0.37 for NM305; CV[Zn]=0.54 and CV[AWQC]=0.42 for CC288; and CV[Zn]=0.47 and CV[AWQC]=0.42 for SF271.

The correlations and similarity of coefficients of variation means that the affected terms in Eq 12 will compensate and tend to cancel. The net effect is indicated by examination of $E[R]_{Ps}$ estimates for SF271, NM305, and CC288, as discussed next.

$E[R]_{Ps}$ estimates for SF271, NM305, and CC288 over the range of Lr from 1 to 20 and Ps from 0.01 to 0.90 showed that the ratios of $E[R_2]_{Ps}/E[R_1]_{Ps}$ between any of these locations could be closely approximated by the following reduced form of Eq 12; that is:

$$[13] \quad E[R_2]_{Ps}/E[R_1]_{Ps} = E[C_1]/E[C_2]$$

For these three locations, the difference between the actual ratios of $E[R_2]_{Ps}/E[R_1]_{Ps}$ and $E[C_1]/E[C_2]$, based on the zinc concentrations reported in the RI and FS (Table 1.5-13), was less than 10 percent for any value of Lr and Ps in the ranges estimated (1 to 20 and 0.01 to 0.90). The average value of the ratio of the ratios, $(E[R_2]_{Ps}/E[R_1]_{Ps})/(E[C_1]/E[C_2])$, was between 0.98 and 1.01. Simply put, using Eq 13, $E[R]_{Ps}$ for any one of these locations could be accurately

estimated from $E[R]_{Ps}$ for either of the other two locations. This indicates that, for these locations, the following terms in Eq 12 are approximately 1; that is:

$$\{(1+CV[AWQC_1]^2)\mathbf{S}_{1,1}/E[AWQC_1]\}\{E[AWQC_2](1+CV[C_2]^2)\mathbf{S}_{2,2}\}\mathbf{S}_{1,2} = 1, \text{ approximately}$$

In practical terms, these three locations (NM305 at the mouth of Ninemile Creek, CC288 at the moth of Canyon Creek, and SF271 on the SFCDR at Pinehurst) are considered generally representative of conditions in affected areas of the upper basin. The relationship summarized by Eq 13 is therefore considered generally applicable in the upper basin. This conclusion is of great practical value. It means that, using currently available data, $E[R]_{Ps}$, the remediation factor required to achieve target load ratios with specified probability, at any location can be estimated from $E[R]_{Ps}$ for any other location. That is, form Eq 13:

$$[14] \quad E[R_2]_{Ps} = E[R_1]_{Ps}\ (E[C_1]/E[C_2])$$

For zinc, Eq 14 says that the required remediation factor at location 2, $E[R_2]_{Ps}$, can be approximated as the required remediation factor at location 1, $E[R_1]_{Ps}$, multiplied by the ratio of the expected values of the zinc concentrations at these locations, $E[C_1]/E[C_2]$. For example, using existing data, $E[R]_{Ps}$ for monitoring locations along the East Fork of Ninemile Creek can be estimated from $E[R]_{Ps}$ for the mouth of Ninemile, NM305. Similarly, $E[R]_{Ps}$ for locations along the SFCDR could be estimated from $E[R]_{Ps}$ for SF271. And so on. For the upper basin, this approximation procedure is considered generally applicable and appropriate for practical purposes.

B.3.5 COMPOSITE REMEDIATION AND SCALING FACTORS FOR COMBINED SOURCES

This section first develops general background on composite remediation factors for combined sources. Subsequent subsections deal, first, with combined loadings for source areas downgradient of Pinehurst to Harrison, followed by development of scaling factors that can be used to make loading estimates that include the Bunker Hill Superfund Site (BHSS). Although these scaling factors support parametric estimates that are functions of BHSS loadings, estimates of BHSS loadings are outside the present scope of this technical memorandum and are not included here.

Composite remediation factors for time t=0, Rc_0, resulting from multiple sources (i=1,n) each having remediation factor R_{0i} and load L_i are defined as follows:

[1] $Rc_0 = {}_{i=1,n} R_{0i} L_i / {}_{i=1,n} L_i$

The composite remediation factor Rc_0 by Eq 1 is a weighted average of the R_{0i} for the n source areas, where the weights are L_i divided by the sum of L_i, $\mathbf{E}L_i$. The expected value of the composite remediation factor, $E[Rc_0]$, can be estimated as follows, where the sums are assumed approximately lognormal:

[2] $E[Rc_0] = E[\mathbf{E}_{i=1,n}R_{0i} L_i / \mathbf{E}_{i=1,n}L_i]$

$= E[\mathbf{E}R_{0i}L_i](1 + CV[\mathbf{E}L_i]^2)\mathbf{S}_{\ln\mathbf{E}R0iLi,\ln\mathbf{E}Li} / E[\mathbf{E}L_i]$

$= (\mathbf{E}E[R_{0i}]E[L_i]\mathbf{S}_i)(1 + CV[\mathbf{E}L_i]^2)\mathbf{S}_{\ln\mathbf{E}R0iLi,\ln\mathbf{E}Li} / \mathbf{E}E[L_i]$

where:

$\mathbf{S}_i$	=	$\exp\{-p\{\ln(CV[R_{0i}]^2+1)\ln(CV[L_i]^2+1)\}^{1/2}\}$
p	=	Correlation coefficient between the natural log of R_{0i} and the natural log of L_i
	=	$p_{\ln R0i,\ln Li}$
$\mathbf{S}_{\ln\mathbf{E}R0iLi,\ln\mathbf{E}Li}$	=	$\exp\{-p\{\ln(CV[\mathbf{E}R_{0i}L_i]^2+1)\ln(CV[\mathbf{E}L_i]^2+1)\}^{1/2}\}$
p	=	Correlation coefficient between the natural log of $\mathbf{E}R_{0i}L_i$ and the natural log of $\mathbf{E}L_i$
	=	$p_{\ln\mathbf{E}R0iLi,\ln\mathbf{E}Li}$

If the $CV[R_{0i}]$ are equal to zero or can otherwise be neglected to a first approximation, $E[Rc_0]$ simplifies to the following ($CV[R_{0i}]=0$ means $\mathbf{S}_i=1$ and $CV[\mathbf{E}R_{0i}L_i]=CV[\mathbf{E}L_i]$):

[3] $E[Rc_0] = \mathbf{E}E[R_{0i}]E[L_i] / \mathbf{E}E[L_i]$

If the coefficient of variation terms are not negligible, Eq 3 can be considered a general first-order approximation of $E[Rc_0]$. Eq 3 was used in PAT 1 to estimate a composite remediation factor for load contributions downgradient of Pinehurst, including loads from the lower basin, as discussed in the following.

B.3.5.1 Combined Loads: Pinehurst to Harrison

Metal loads at Harrison (LC60) are the sum of loads at Pinehurst (SF271) plus the increment of loading from Pinehurst to Harrison. This section documents the equations used to estimate post-remediation metal (zinc) loads in the basin downgradient of Pinehurst to Harrison. The loads were the sum of loads from the following four source areas that included the lower basin (CSM Unit 3) and the North Fork Coeur d'Alene River (NFCDR); that is:

1: SFCDR from SF271 to the confluence with the NFCDR
2: NFCDR (NF50)
3: Lower Basin—CDR from the confluence to Cataldo (LC50)
4: Lower Basin—CDR from Cataldo to Harrison (LC60)

A composite remediation factor at time t=0, R_{PH0}, representing the sum of these four source areas was calculated from R_0 point estimates and the expected value of pre-remediation loading L for each of the four areas. The expected value of the composite remediation factor, $E[R_{PH0}]$, for each alternative was approximated as a weighted average as follows:

$$[1] \quad E[R_{PH0}] = {}_{i=1,4}R_{0i}\, E[L_i] \,/\, {}_{i=1,4}E[L_i]$$

$$= {}_{i=1,4}R_{0i}\, E[L_i] \,/\, E[L_{PH}]$$

where:

R_{0i} = R_0 point estimate for area i, i=1,2,3,4
= $E[R_{0i}]$ approximately
$E[L_i]$ = E[L] for area i, i=1,2,3,4 (estimated from historical data)
$E[L_{PH}]$ = Loading from Pinehurst to Harrison
= **E**$E[L_i]$

The R_{0i} point estimates were equated with expected values following the discussion of Section B.3.3.3. This approximation was considered particularly justified because significant remediation is not included under any of the alternatives for the NFCDR (i=2) and the lower basin from the Confluence to Cataldo (i=3). The true R_0 for the NFCDR and the lower basin from the Confluence to Cataldo would thus be 1.0, making E[R]=1 with CV[R]=0 for these areas, which represent the majority of zinc load downgradient of SF271.

$E[R_{PH}(t)]$ was estimated as $E[R_{PH0}]E[e^{-\$t}]_{PH}$, where $E[e^{-\$t}]_{PH}$ is the expected value of the source depletion factor for the loads downgradient of Pinehurst to Harrison. $E[e^{-\$t}]_{PH}$ represents the combined effect of source depletion in the four source areas. However, data for zinc source volumes and concentrations needed to estimate TEM', and thus decay rates **$**, are not available for either the NFCDR or the upper CDR (from the confluence to Cataldo). In the absence of such data, $E[e^{-\$t}]_{PH}$ was approximated as the arithmetic average of $E[e^{-\$t}]$ estimated for the upper basin above Pinehurst and $E[e^{-\$t}]$ estimated for the lower CDR from Cataldo to Harrison, where there is data to estimate TEM', and thus **$**.

$CV[R_{PH}(t)]$ was estimated using the equations in Section B.3.3.3 that relate $CV[R(t)]$ to $E[R(t)]$. Loads were estimated as the probabilistic product of $R_{PH}(t)$ and L_{PH}, the pre-remediation loading from Pinehurst to Harrison (L_{PH}= **E**$E[L_i]$) using the equations of Section B.3.2.

B.3.5.2 Combined Loads That Include the BHSS

The post-remediation estimates presented in this technical memorandum do not include loadings from the BHSS. However, remediation factors, loads, and load ratios that *do* include BHSS loading can be approximated from results that *do not* include BHSS loading. Expressions are developed in the following for composite remediation factors R(t)" and scaling factors for loading F(t)" and load ratios Lr(t)", where the double prime " represents the combined effect of loads with the BHSS. Although these expressions support parametric estimates that are functions of BHSS loadings, the actual BHSS loadings are outside the present scope of this technical memorandum and are not included here.

Remediation Factors

The composite remediation factor R(t)" representing the combined effect of remediation inside and outside of the BHSS is defined as follows:

$$[1] \quad R(t)'' = \{R(t)L + R(t)_{BHSS}L_{BHSS}\} / \{L + L_{BHSS}\}$$
$$= \mathbf{E}RL / \mathbf{E}L$$

where:

$R(t)$	=	Remediation factor for sources other than BHSS
L	=	Pre-remediation (historic) load from sources other than BHSS
$R(t)_{BHSS}$	=	Remediation factor for the BHSS
L_{BHSS}	=	Historical loading from the BHSS
ERL	=	$R(t)L + R(t)_{BHSS}L_{BHSS}$
EL	=	$L + L_{BHSS}$

R(t)" is specific to monitoring locations and is, therefore, different for Pinehurst (SF271) and Harrison (LC60). The expected value of the composite remediation factor, E[R"], can be estimated as follows, where the time dependence is implicit; that is:

[2] $E[R''] = E[\mathbf{E}RL / \mathbf{E}L]$

$= \mathbf{E}E[RL](1+CV[\mathbf{E}L]^2)\mathbf{S}_{\ln\mathbf{E}RL,\ln\mathbf{E}L} / \mathbf{E}E[L]$

$= (E[R]E[L]\mathbf{S} + E[R_{BHSS}]E[L_{BHSS}]\mathbf{S}_{BHSS})(1+CV[\mathbf{E}L]^2)\mathbf{S}_{\ln\mathbf{E}RL,\ln\mathbf{E}L} / (E[L]+E[L_{BHSS}])$

where:

$\mathbf{S}$	=	$\exp\{-p\{\ln(CV[R]^2+1)\ln(CV[L]^2+1)\}^{1/2}\}$
p	=	Correlation coefficient between the natural log of R and the natural log of L
	=	$p_{\ln R,\ln L}$
$\mathbf{S}_{BHSS}$	=	$\exp\{-p\{\ln(CV[R_{BHSS}]^2+1)\ln(CV[L_{BHSS}]^2+1)\}^{1/2}\}$
p	=	Correlation coefficient between the natural log of R_{BHSS} and the natural log of L_{BHSS}
	=	$p_{\ln R\text{-}BHSS,\ln L\text{-}BHSS}$
$\mathbf{S}_{\ln\mathbf{E}RL,\ln\mathbf{E}L}$	=	$\exp\{-p\{\ln(CV[\mathbf{E}RL]^2+1)\ln(CV[\mathbf{E}L]^2+1)\}^{1/2}\}$
p	=	Correlation coefficient between the natural log of $\mathbf{E}RL$ and the natural log of $\mathbf{E}L$
	=	$p_{\ln\mathbf{E}RL,\ln\mathbf{E}L}$

A first-order approximation of E[R"] can be obtained by neglecting the coefficient of variation terms. In this case, Eq 3 simplifies to the following, which is the weighted average of E[R] and $E[R_{BHSS}]$, where the weights are the respective loads divided by the sum of the loads; that is:

[4] $E[R''] = (E[R]E[L] + E[R_{BHSS}]E[L_{BHSS}]) / (E[L]+E[L_{BHSS}])$

Loads and Load Ratios

Similarly, the combined load F(t)" from sources other than the BHSS, F(t), and load from the BHSS, $F(t)_{BHSS}$, can be estimated as follows, starting with the basic relationship:

[5] $F(t)'' = F(t) + F(t)_{BHSS}$

The ratio F(t)"/F(t) can be used with the loading F(t) and load ratio Lr(t) results presented in Section 3 to estimate the combined loading F(t)" and load ratios Lr(t)" that include the BHSS. The F(t)" estimates can be obtained by multiplying F(t) by the ratio F(t)"/F(t); similarly, the Lr(t)" estimates can be obtained by multiplying Lr(t) by the ratio F(t)"/F(t). Ratios are specific to monitoring locations and are therefore different for Pinehurst (SF271) and Harrison (LC60). The ratios are developed as follows, starting with Eq 5 and substituting various equivalent relationships:

[6a] $F(t)''/F(t) = 1 + F(t)_{BHSS} / F(t)$

[6b] $F(t)''/F(t) = 1 + R(t)_{BHSS} L_{BHSS} / R(t) L$

[6c] $F(t)''/F(t) = 1 + R_{0\text{-}BHSS} e^{-\$t}{}_{BHSS} L_{BHSS} / R_0 e^{-\$t} L$

The expected value of the ratio given by Eq 6a can be estimated as follows, where the time dependence is implicit; that is:

[7] $E[F''/F] = E[1 + F_{BHSS}/F]$

$= 1 + E[F_{BHSS}](1+CV[F]^2)\mathbf{S}/E[F]$

where:

$\mathbf{S} = \exp\{-p\{\ln(CV[F_{BHSS}]^2+1)\ln(CV[F]^2+1)\}^{1/2}\}$

p = Correlation coefficient between the natural log of F_{BHSS} and the natural log of F

$= p_{lnF\text{-}BHSS,lnF}$

If it can assumed that the F_{BHSS} and F are highly correlated, which they probably would be (such that $(1+CV[F]^2)\mathbf{S}$ is approximately equal to 1), then E[F"/F] can be closely approximated by the following:

[8] $E[F''/F] = 1 + E[F_{BHSS}]/E[F]$

Similarly, *if* $e^{-\$t}_{BHSS}$ and $e^{-\$t}$ are approximately equal, then Eq 6c simplifies to the following:

[9] $F(t)''/F(t) = 1 + R_{0\text{-}BHSS}L_{BHSS}/R_0L$

And the expected value is:

[10] $E[F(t)''/F(t)] = E[1 + R_{0\text{-}BHSS}L_{BHSS}/R_0L]$

$= 1 + E[R_{0\text{-}BHSS}L_{BHSS}/R_0L]$

$= 1 + E[R_{0\text{-}BHSS}L_{BHSS}](1+CV[R_0L]^2)\mathbf{S}/E[R_0L]$

$= 1 + E[R_{0\text{-}BHSS}]E[L_{BHSS}]\mathbf{S}_{BHSS}(1+CV[R_0L]^2)\mathbf{S}'/E[R_0]E[L]\mathbf{S}$

where:

$\mathbf{S}' = \exp\{-p\{\ln(CV[R_{0\text{-}BHSS}L_{BHSS}]^2+1)\ln(CV[R_0L]^2+1)\}^{1/2}\}$

p = Correlation coefficient between the natural log of $R_{0\text{-}BHSS}L_{BHSS}$ and the natural log of R_0L

$= p_{lnRL\text{-}BHSS,lnRL}$

$\mathbf{S}_{BHSS} = \exp\{-p\{\ln(CV[R_{0\text{-}BHSS}]^2+1)\ln(CV[L_{BHSS}]^2+1)\}^{1/2}\}$

p = Correlation coefficient between the natural log of R_{BHSS} and the natural log of L_{BHSS}

$= p_{lnR0\text{-}BHSS,lnL\text{-}BHSS}$

$\mathbf{S} = \exp\{-p\{\ln(CV[R_0]^2+1)\ln(CV[L]^2+1)\}^{1/2}\}$

p = Correlation coefficient between the natural log of R_0 and the natural log of L

$= p_{lnR0,lnL}$

If it can be assumed that $R_{0\text{-}BHSS}L_{BHSS}$ and R_0L are highly correlated, which they probably would be (such that $(1+CV[R_0L]^2)\mathbf{S}'$ is approximately 1), and that $\mathbf{S}_{BHSS}$ and $\mathbf{S}$ are approximately equal, then Eq 10 simplifies to the following:

[11] $E[F(t)''/F(t)] = 1 + E[R_{0\text{-}BHSS}]E[L_{BHSS}]/E[R_0]E[L]$

A similar analysis would show that E[F(t)"/F(t)] by Eq 6b could be approximated as follows:

[12] $E[F(t)''/F(t)] = 1 + E[R(t)_{BHSS}]E[L_{BHSS}]/E[R(t)]E[L]$

To summarize (where, for consistency, time t has been made explicit in each expression), where all equalities are approximations:

[13a] $E[F(t)''/F(t)] = 1 + E[F(t)_{BHSS}]/E[F(t)]$

[13b] $E[F(t)''/F(t)] = 1 + E[R(t)_{BHSS}]E[L_{BHSS}]/E[R(t)]E[L]$

[13c] $E[F(t)''/F(t)] = 1 + E[R_{0\text{-}BHSS}]E[L_{BHSS}]/E[R_0]E[L]$, if $e^{-\$t}_{BHSS}=e^{-\$t}$ (approximately)

The expected value of E[F(t)"/F(t)] can be expanded as follows:

[14] $E[F(t)''/F(t)] = E[F(t)''](1 + CV[F(t)]^2)\mathbf{S} / E[F(t)]$

where:

$\mathbf{S}$ = $\exp\{-p\{\ln(CV[F''(t)]^2+1)\ln(CV[F(t)]^2+1)\}^{1/2}\}$

p = Correlation coefficient between the natural log of F"(t) and the natural log of F(t)

= $p_{\ln F''(t),F(t)}$

Assuming that F"(t) and F(t) are highly correlated, which is likely, $(1 + CV[F(t)]^2)\mathbf{S}$ is approximately 1.0 and Eq 14 simplifies to the following:

[15] $E[F(t)''/F(t)] = E[F(t)''] / E[F(t)]$

Substituting the right hand side of Eq 15 into the left hand side of Eq 13 yields:

[16a] $E[F(t)'']/E[F(t)] = 1 + E[F(t)_{BHSS}]/E[F(t)]$

[16b] $E[F(t)'']/E[F(t)] = 1 + E[R(t)_{BHSS}]E[L_{BHSS}]/E[R(t)]E[L]$

[16c] $E[F(t)"]/E[F(t)] = 1 + E[R_{0\text{-}BHSS}]E[L_{BHSS}]/E[R_0]E[L]$

To reiterate, the ratio F(t)"/F(t) can be used with the loading F(t) and load ratio Lr(t) results presented in this technical memorandum to estimate the combined loading F(t)" and load ratios Lr(t)" that include the BHSS. The F(t)" estimates can be obtained by multiplying F(t) by the ratio F(t)"/F(t); similarly, the Lr(t)" estimates can be obtained by multiplying Lr(t) by the ratio F(t)"/F(t). Ratios are specific to monitoring locations and are therefore different for Pinehurst (SF271) and Harrison (LC60). The BHSS Inclusion Factors (BIFs) in Figures 4-26 and 4-27 of the main text are based on Eqs 16 a,b,c; that is:

[17] $E[F(t)"]/E[F(t)] = BIF$

and, therefore,

[18] $E[F(t)"] = BIF \cdot E[F(t)]$

Assuming high temporal correlation between loadings with and without the BHSS, the following scaling relationships can be used as first approximations, based on generalization of the results up to this point; that is:

[19] $F(t)_{Pn}" = BIF \cdot F(t)_{Pn}$

[20] $E[Lr(t)"] = BIF \cdot E[Lr(t)]$

[21] $Lr(t)_{Pn}" = BIF \cdot Lr(t)_{Pn}$

In other words, estimates with the BHSS are equal to estimates without the BHSS times the BIF scaling factors. To reiterate, the scaling factors (BIF) are approximate and require independent estimates of the remediation factor appropriate for the BHSS.

B.3.6 LOADING DUE TO MULTIPLE SOURCES AND "WHAT-IF" SCENARIO ANALYSES

The section presents the basic equations for estimating the expected value and coefficient of variation of metal loadings resulting from multiple individual sources. These equations were

used in PAT1 to add the loadings from the lower basin to the upper basin loading estimates. These equations are used more generally in what are called "what-if" or scenario analyses, as implemented in PAT2.

What-if analyses may be used to estimate the net cleanup effects at selected monitoring locations resulting from cleanup of upstream sources. Cleanup of upstream sources are characterized in terms of the E[R] and CV[R] for the specific sources, where the values of E[R] and CV[R] may be hypothetical (as in "what if"), estimated (by formal analysis), or a combination of both. The resulting downstream cleanup effects are characterized in two ways, as:

- Resultant E[R] and CV[R]
- Resultant probabilities of meeting loading capacities and AWQC ($Lr \leq 1$) or multiples of loading capacities and AWQC ($Lr > 1$)

In general, the sources may be tributaries or stream reaches, which may or may not be contiguous. The individual sources may be aggregated (i.e., source additions) or disaggregated (i.e., source subtractions). Aggregation and disaggregation may also be combined.

Aggregation is used to sum sources contributing to a given monitoring point or stream reach. The most important application for post-remediation loading is the summation of estimated loadings from the upper and lower basin, as implemented in PAT1. In the what-if analysis implemented in PAT2, the major tributaries and segments of the SFCDR are added to estimate loadings in the SFCDR at Pinehurst and, by adding the loading from the lower basin, the CDR at Harrison.[30]

Disaggregation is used to subtract sources and estimate loadings along reaches between sampling points. An important example is the loading from the lower basin. This is estimated as the CDR loading at Harrison minus the SFCDR loading at Pinehurst minus the NFCDR loading at Enaville minus the estimated loading from Pinehurst to the SFCDR-NFCDR confluence. The loading to the SFCDR from Wallace to Elizabeth Park is the SFCDR loading at Elizabeth Park minus the loading from Canyon Creek minus the loading from Ninemile Creek minus the loading from the upper SFCDR. The loading to the SFCDR from the BHSS (implemented in PAT2 only) is estimated as the SFCDR loading at Pinehurst minus the SFCDR loading at Elizabeth Park minus the loading from Pine Creek.

The loading for any specific location or reach "x" is symbolized F_x. Although the symbolism "F," for post-remediation conditions, is used in the following equations, the same equations apply to either pre-remediation, L, or post-remediation, F, conditions. There is thus no loss of

[30] All loading estimates in PAT1 exclude contributions from the BHSS.

generality to pre-remediations conditions. Again, depending on context and estimation needs, the x locations may be discrete sampling/monitoring points or stream reaches, segments, or the like.

F_x is the cumulative metal loading due to n identified contributing source aggregations (additions) or disaggregations (subtractions). The individual sources contributing to F_x are symbolized F_i and represent the loading from source i. I_f's add to F_x for source additions and subtract from F_x for source subtractions (or sink additions, sinks being negative sources). F_x is summed over all contributing sources, 1 through n, the total number of sources. Post-remediation conditions are tacitly functions of time, F(t). Formally, the relationship is as follows:

[1] $F_x = {}_{i=1,n}\, a_i F_i$

The sign of the F_i's (positive for source additions and negative for source subtractions) in Eq 1, and subsequent equations, enter through the coefficient a_i, where a_i=+1 for source additions and a_i=–1 for source subtractions.

The expected value and coefficient of variation of F_x, $E[F_x]$ and $CV[F_x]$, are estimated from the mathematical relationships defining the sums of random variables, as documented in Appendix A. Because the coefficient of variation $CV[F_x]$ depends on the correlation between pairs of sources, $p_{Fi,Fj}$ (where j is a dummy variable), it is useful to define $CV[F_x]$ in terms of the variance, $V[F_x]$; that is:

[2] $E[F_x] = {}_{i=1,n}\, a_i E[F_i]$

[3] $CV[F_x] = V[F_x]^{½} / E[F_x]$

$V[F_x] = {}_{i=1,n}(CV[F_i]E[F_i])^2 + 2 \quad {}_{i=1,n} \quad {}_{j=i+1} p_{Fi,Fj} a_i a_j E[F_i]E[F_j]CV[F_i]CV[F_j]$

where:

$F_i = R_i \mathbb{C} L_i$
$E[F_i] = E[R_i]\, E[L_i]\, \mathbf{S}$
$CV[F_i] = \{(CV[R_i]^2+1)(CV[L_i]^2+1)\mathbf{S}^2 - 1\}^{1/2}$
$\mathbf{S} = \exp\{p_{lnR,lnL}\{\ln(CV[R_i]^2+1)\ln(CV[L_i]^2+1)\}^{1/2}\}$
a_i, a_j = +1 where source loadings are added
= –1 where source loadings are subtracted

The F_x and F_i nomenclature used in Eqs 1, 2, and 3 depends on context and is meant to be flexible. For example, for one source, n=1, and F_x=F_i. An F_x made up of one or more sources can itself contribute to another location or reach and thus be an "F_i" in that context. For example, Canyon Creek is an "F_x" for loads at the mouth of the creek and an "F_i" for loads at Pinehurst. Similar rationale holds for R_i and L_i.

B.3.6.1 Special Cases

This section documents two special cases of Eqs 1 through 3 that are of particular use. These cases are the sum of two sources and the difference of two sources.

Sum of Two Sources

For the special case of estimating the combined loading from two sources, as used in PAT1 for adding the lower basin loading to the upper basin loading, Eqs 1, 2, and 3 simplify to the following[31]:

[4] $$F_x = F_1 + F_2$$

[5] $$E[F_x] = E[F_1] + E[F_2]$$

[6] $$CV[F_x] = V[F_x]^{½} / E[F_x]$$

$$V[F_x] = (CV[F_1]E[F_1])^2 + (CV[F_2]E[F_2])^2 + 2p_{F1,F2}E[F_1]E[F_2]CV[F_1]CV[F_2]$$
$$= V[F_x]_{p=0} + (V[F_x]_{p=1} - V[F_x]_{p=0})\, p_{F1,F2}$$

where:

$$V[F_x]_{p=1} = (CV[F_1]E[F_1] + CV[F_2]E[F_2])^2$$
$$V[F_x]_{p=0} = (CV[F_1]E[F_1])^2 + (CV[F_2]E[F_2])^2$$

For the sum of two sources the minimum variance and coefficient of variation occurs when the two sources are perfectly negatively correlated, such that $p_{F1,F2}$= –1. The maximums occur when the two sources are perfectly positively correlated, such that $p_{F1,F2}$=1. To the extent there is correlation between two sources, it is generally more likely to be positive rather than negative.

[31] In this case, $a_1 = a_2 = +1$, such that $a_1a_2 = +1$.

Difference of Two Sources

For the special case of estimating the loading in a given reach F_x between an upgradient (U) and downgradient (D) sampling station, Eqs 1, 2, and 3 simplify to the following[32]:

[7] $F_x = F_D - F_U$

[8] $E[F_x] = E[F_D] - E[F_U]$

[9] $CV[F_x] = V[F_x]^{½} / E[F_x]$

$$V[F_x] = (CV[F_D]E[F_D])^2 + (CV[F_U]E[F_U])^2 - 2p_{FD,FU}E[F_D]E[F_U]CV[F_D]CV[F_U]$$
$$= V[F_R]_{p=0} - (V[F_R]_{p=-1} - V[F_R]_{p=0})\, p_{FD,FU}$$

where:

$$V[F_x]_{p=-1} = (CV[F_D]E[F_D] + CV[F_U]E[F_U])^2$$
$$V[F_x]_{p=0} = (CV[F_D]E[F_D])^2 + (CV[F_U]E[F_U])^2$$

In contrast to the sum of two sources, the minimum variance and coefficient of variation for the difference of two sources occurs when the two sources are perfectly positively correlated, such that $p_{FD,FU}=1$. The maximums occur when the two sources are perfectly negatively correlated, such that $p_{FD,FU}= -1$. In general, upgradient and downgradient stations on the same stream will have a significant positive correlation. The correlation will increases as the distance between the stations decreases, and becomes 1.0 as the distance goes to zero. For the case of virtually perfect correlation, such that $p_{FD,FU}$ is 1.0, $CV[F_x]$ is a minimum, and Eq 9 reduces to the following:

[10] $CV[F_x] = \{(CV[F_D]E[F_D])^2 + (CV[F_U]E[F_U])^2 - 2CV[F_D]E[F_D]CV[F_U]E[F_U]\}^{½} / E[F_R]$

B.3.6.2 Remediation Factor Estimated from Pre- and Post-remediation Loadings

This section documents the equations used to estimate aggregate values of E[R] and CV[R] for downstream monitoring locations that are the sum of two or more reaches. These equations were used in PAT1 to estimate the aggregate values of E[R] and CV[R] for each remedial alternative resulting from the combined remediation of the upper and lower basins. These estimates were based on the pre-remediation loads and post-remediation loads in the CDR at Harrison (assuming no load contribution from the BHSS). In PAT2, these equations were used to estimate the

[32] In this case, $a_D = +1$ and $a_U = -1$, such that $a_U a_D = -1$.

aggregate values of E[R] and CV[R] at both the SFCDR at Pinehurst and the CDR at Harrison due to remediation in upstream tributaries and reaches. The formal relationships used were as follows:

[11] $R = F / L$

[12] $E[R] = E[F](1+CV[L]^2)\mathbf{S} / E[L]$

[13] $CV[R] = \{(CV[F]^2+1)(CV[L]^2+1)\mathbf{S}^2 - 1\}^{1/2}$

$\mathbf{S} = \exp\{-p\{\ln(CV[F]^2+1)\ln(CV[L]^2+1)\}^{1/2}\}$

p = Correlation coefficient between the natural log of F and the natural log of L

$= p_{lnL,lnF}$

For F and L lognormally distributed, R is also lognormally distributed and Eqs 12 and 13 are mathematically exact. The correlation coefficient $p_{lnL,lnF}$ was estimated by back-calculating the value required to assure that any estimate of the true value of R (from R=F/L) would not exceed 1.0 (R>1.0 being physically impossible for practical purposes) with negligible probability, " ("=0.01 was used in PAT1 and the more restrictive "=0.001 was used in PAT2 to assure that E[R]=1.0 for the case of no remediation in any upstream area). The estimate of $p_{lnL,lnF}$ that makes P[R>1]= " is made as follows:

[14] $p_{lnL,lnF} = (A - B^2)/2C^{1/2}$

where:

$A = \ln\{(CV[F]^2+1)*(CV[L]^2+1)\}$

$B = \ln\{E[L]/E[F]\{(CV[F]^2+1)/(CV[L]^2+1)\}^{1/2}\} / u_{1-"}$

$C = \ln\{CV[F]^2+1\}*\ln\{CV[L]^2+1\}$

$u_{1-"}$ = the normal standard variate for a probability of 1-"

This section completes the documentation for the analysis of the remedial alternatives as implemented in PAT1. The following section completes the documentation of the what-if analysis as implemented in PAT2.

B.3.6.3 What-if Scenario Analyses

For what-if analyses implemented in PAT2, Eqs 2 and 3 were used with loads at selected monitoring locations to aggregate and disaggregate sources as needed to estimate loads for remedial action areas (RAAs) of interest. Estimates of $E[L_i]$ and $CV[L_i]$ as needed for RAAs of interest used the estimates of E[L] and CV[L] documented in Appendix C Section C.3, which were based on analysis of the historical data presented in the RI and summarized in the FS, for the following monitoring locations:

- Canyon Creek (CC288)
- Ninemile Creek (NM305)
- Upper SFCDR (SF228)
- Pine Creek (PC305)[33]
- SFCDR at Elizabeth Park (SF268)
- SFCDR at Pinehurst (SF271)
- NFCDR (NF50)
- CDR at Cataldo (LC50)
- CDR at Harrison (LC60)

The following RAAs were defined for what-if analyses. Each RAA could have its own hypothetical (or estimated) values of E[R] and CV[R], from which E[F] and CV[F] could be estimated.[34]

- Canyon Creek (CC288) RAA

- Ninemile Creek (NM305) RAA

- Upper SFCDR (SF228) RAA

- Pine Creek (PC305) RAA

- SFCDR Wallace to Elizabeth Park RAA
  Load estimated as Loading{SF268 – (SF228 + CC288 + NM305)}

- BHSS (Elizabeth Park to Pinehurst) RAA
  Load estimated as Loading{SF271– (SF268 + PC305)}

[33] Two apparent outliers (10 Jan and 22 Mar 1995) were removed from the Pine Creek load data.
[34] Suppressing the "i and x" subscripts on R and F to avoid confusion without loss of meaning.

- NFCDR (NF50) RRA

- CDR Pinehurst to Cataldo RAA
  Load estimated as Loading{LC50 – (SF271 + NF50)}[35]

- CDR Cataldo to Harrison RAA
  Load estimated as the differential from Cataldo to Harrison

- Upper Basin RAA – Post-remediation loading measured at Pinehurst (SF271)
  Estimated as remedial actions (if any) at the following RAAs: Canyon Creek, Ninemile Creek, Pine Creek, Upper SFCDR, SFCDR Wallace to Elizabeth Park, and the BHSS. The upper basin RAA without the BHSS is estimated by setting E[R]=CV[R]=0 for the BHSS RAA.

- Basin-Wide RAA – Post-remediation loading measured at Harrison (LC60)
  Estimated as remedial actions (if any) at all upper and lower basin RAAs. The basin-wide RAA without the BHSS is estimated by setting E[R]=CV[R]=0 for the BHSS RAA.

- Lower Basin RAA – Post-remediation loading measured at Harrison (LC60)
  Estimated from the basin-wide RAA by setting E[R]=CV[R]=0 for all upper basin RAAs and the NFCDR RAA.

Based on the expectation that loads from relatively nearby stations will be highly correlated, a correlation coefficient of 0.9 was assumed between stations for estimating the pre-remediation loads for the following RAAs:

- SFCDR Wallace to Elizabeth Park RAA
  Load estimated as Loading{SF268 – (SF228 + CC288 + NM305)}

- BHSS (Elizabeth Park to Pinehurst) RAA
  Load estimated as Loading{SF271– (SF268 + PC305)}[36]

- CDR Pinehurst to Cataldo RAA
  Load estimated as Loading{LC50 – (SF271 + NF50)}[37]

[35] Note this includes that portion of the SFCDR from SF271 to the SFCDR-NFCDR confluence, which is included in the lower basin for the PAT2 analyses.
[36] Load dominated by SF271 and SF268.
[37] Load dominated by SF271 and LC50, although very high CV[L] for NF50.

The pre-remediation correlations between sources for the upper basin and basin-wide RAAs were estimated by calibration assuming constant correlation between all contributing sources. Sources for these RAAs are all additive (such that $a_i = a_j = +1$ for all i and j in Eq 3 using L in place of F). Therefore, according to Eq 3, a constant correlation between all sources (i.e., constant $p_{Li,Lj}$ for all i and j) would allow estimates for $CV[L_x]$ to be made from $V[L_x]$ formulated as follows:

[15] $$CV[L_x] = V[L_x]^{1/2} / E[L_x]$$

$$V[L_x] = \Sigma_{i=1,n}(CV[L_i]E[L_i])^2 + 2p_{Li,Lj}\,\Sigma_{i=1,n}\Sigma_{j=i+1}E[L_i]E[L_j]CV[L_i]CV[L_j]$$
$$= V[L_x]_{p=0} + (V[L_x]_{p=1} - V[L_x]_{p=0})\,p_{Li,Lj}$$

where:

$V[L_x]_{p=1}$ = $V[L_x]$ for $p_{Li,Lj}=1$ for all L_i, L_j (the upper-bound estimate of $V[L_x]$)
= $(\Sigma_{i=1,n}CV[L_i]E[L_i])^2$

$V[L_x]_{p=0}$ = $V[L_x]$ for $p_{Li,Lj}=0$ for all L_i, L_j
= $\Sigma_{i=1,n}(CV[L_i]E[L_i])^2$

Using Eq 15, $p_{Li,Lj}$ was adjusted so that $CV[L_x]$ (by Eq 15) matched CV[L] estimated from the historical data, thus calibrating the estimated and measured coefficients of variation (note that the expected values are "self-calibrating," as implicit in Eq 2). In other words, $p_{Li,Lj}$ for the lower basin RAA was chosen so that $CV[L_x]$ computed by Eq 15 equaled CV[L] for Pinehurst (SF271) based on historical data. In the same way, $p_{Li,Lj}$ for the basin-wide RAA was chosen so that $CV[L_x]$ computed by Eq 15 equaled CV[L] for Harrison based on historical data. Since there was no superior alternative, the post-remediation correlations were assumed equal to the pre-remediation correlations; that is: $p_{Fi,Fj} = p_{Li,Lj}$, which appears to be a reasonable approximation.

Probabilities of Meeting TMDL Loading Capacities and AWQC

As part of the what-if analysis, estimates of the probabilities of meeting TMDL loading capacities and AWQC, and multiples of the loading capacities and AWQC, at specific downstream monitoring locations were made using the results of E[F] and CV[F] from the preceding equations in this section and the equations in Section B.3.4.

B.3.7 TIME INTEGRALS

Time integrals of metal loading and AWQC ratios are two potentially useful remedial performance metrics, which have not been discussed. For zinc, or any given metal, the time integral of loading F(t) over a given time interval at a particular location represents the accumulated zinc mass M passing that location during the given time interval. For a time interval from time t=0 to time t=T, the accumulated zinc mass M(T) and its expected value E[M(T)] can be estimated as follows, where the expression for E[M(T)] is approximate and assumes no correlation between F_0 and **$**.

[1] $$M(T) = {}_{0 \text{ to } T} F(t)\, dt$$

$$= {}_{0 \text{ to } T} F_0\, e^{-\$t}\, dt$$

$$= F_0(1 - e^{-\$T})/\$$$

[2] $$E[M(T)] = E[F_0(1 - e^{-\$T})/\$]$$

$$= E[F_0](1 - e^{-E[\$]T})(1 + CV[\$]^2)/E[\$]$$

Similarly for load ratio or AWQC ratio, the time integral of Lr(t) over a given time interval at a particular location represents the accumulated AWQC exceedances at that location during the given time interval. For a time interval from time t=0 to time t=T, the accumulated AWQC exceedances A(T) and its expected E[A(T)] can be reasonably approximated as follows.

[3] $$A(T) = {}_{0 \text{ to } T} Lr(t)\, dt$$

$$= {}_{0 \text{ to } T} Lr_0\, e^{-\$t}\, dt$$

$$= Lr_0(1 - e^{-\$T})/\$$$

[4] $$E[A(T)] = E[Lr_0(1 - e^{-\$T})/\$]$$

$$= E[Lr_0](1 - e^{-E[\$]T})(1 + CV[\$]^2)/E[\$]$$

Eqs 2 and 4 approximate $E[(1 - e^{-\$T})]$ as $(1 - e^{-E[\$]T})$.

B.4.0 CONCLUSIONS

This completes the documentation of the equations and associated approximations used in the analysis of the alternatives and the what-if analysis, as implemented in PAT1 and PAT2 at this time. Additional analysis inputs are documented in Appendix C. Principal results are presented and discussed in Section 4 of the main text, which also discusses the limitations of the analysis, and is included by reference in this appendix.

B.5.0 REFERENCES

Carslaw, H.S. and J.C. Jaeger. 1989. *Conduction of Heat in Solids.* 2d ed. Oxford University Press, Great Britain.

Crank, J. 1989. *The Mathematics of Diffusion*, 2d ed. Oxford University Press, Great Britain.

Natrella, M.G. 1966. *Experimental Statistics*. National Bureau of Standards Handbook 91. USGPO, Washington, D.C.

Ott, W.R. 1995. *Environmental Statistics and Data Analysis*, CRC Lewis Publishers, New York, N.Y.

Paulson, A.J., R. Balderrama, and E. Zahl. 1996. *Treatment of Fluvially Deposited Streamside Mine Waste: Material from Canyon Creek, Idaho*, U.S. Dept. of Energy, Spokane Research Center, Report of Investigation 9633

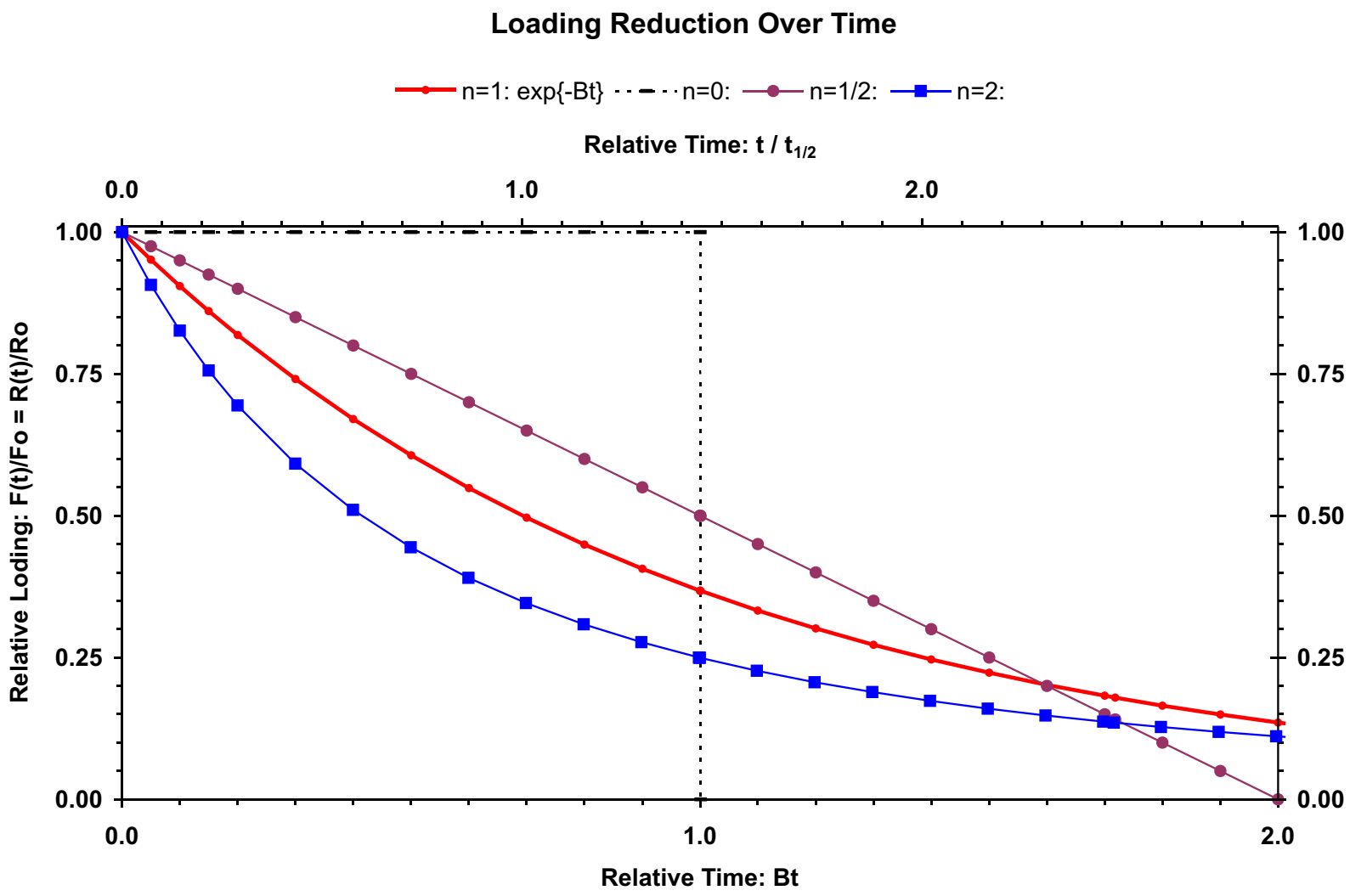
U.S. Environmental Protection Agency (USEPA). 2001. *Estimation of Background Concentrations in Soil, Sediment, and Surface Water in the Coeur d'Alene River and Spokane River Basins. Technical Memorandum* (Revision 2). Prepared by URSG and CH2M HILL for USEPA Region 10, under ARCS Contract No. 68-W9-0054. Seattle, WA. June 2001.

——— 2000a. *Coeur d'Alene Basin RI/FS, Remedial Investigation Report.* Draft (Revision 1). Prepared by URSG and CH2M HILL for USEPA Region 10, under ARCS Contract No. 68-W9-0054. Seattle, WA. October 2000.

———. 2000b. *Coeur d'Alene Basin RI/FS, Feasibility Study Report.* Draft (Revision 1). Prepared by URSG and CH2M HILL for USEPA Region 10, under ARCS Contract No. 68-W9-0054. Seattle, WA. December 2000.

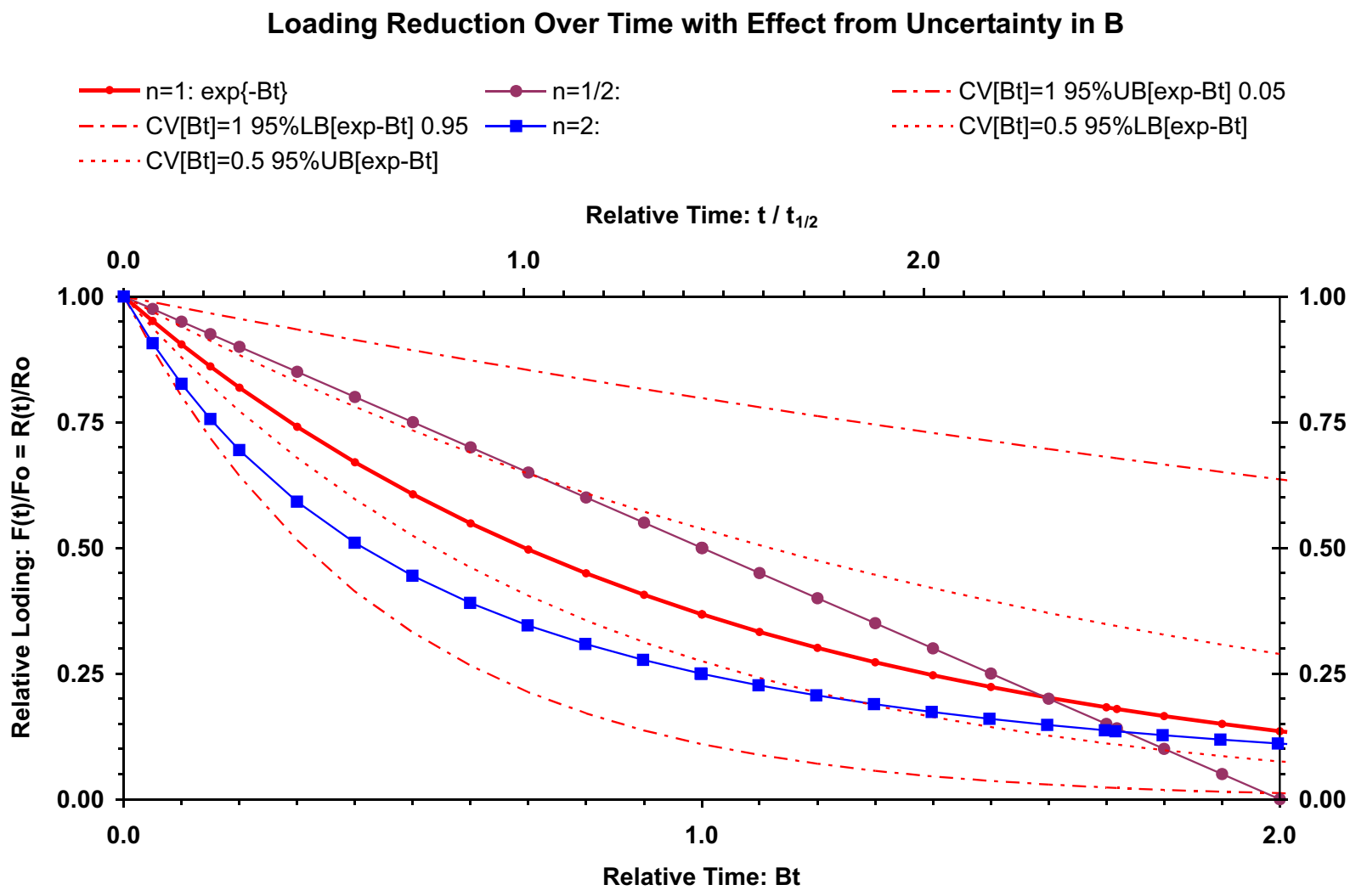
———. 2000c. *Total Maximum Daily Load for Dissolved Cadmium, Dissolved Lead, and Dissolved Zinc in Surface Waters of the Coeur d'Alene River Basin.* August 2000.

———. 1999. *National Recommended Water Quality Criteria—Correction.* EPA 822-Z-99-001. April 1999.



EPA REGION 10	Coeur D'Alene Basin RI/FS A PREDICATIVE ANALYSIS FOR POST-REMEDIATION METAL LOADING Appendix B Doc Control: 33753909-7326	Ref: 027-RI-CO-102 Coeur D'Alene Basin RI/FS Doc Control: 4162500.6538.05a Probabilistic Series 09/13/01

Figure B-1



EPA REGION 10	Coeur D'Alene Basin RI/FS A PREDICATIVE ANALYSIS FOR POST-REMEDIATION METAL LOADING Appendix B Doc Control: 33753909-7326	Ref: 027-RI-CO-102 Coeur D'Alene Basin RI/FS Doc Control: 4162500.6538.05a Probabilistic Series 09/13/01

Figure B-2