EPA/600/R-16/380

United States
Environmental Protection
Agency

EPA Report on the Analysis of the Transport and Fate of Metals Released from the Gold King Mine in the Animas and San Juan Rivers

Response to Peer Review Comments

Office of Research and Development
National Exposure Research Laboratory

EPA/600/R-16/380

EPA Report on the Analysis of the Transport and Fate of Metals Released from the Gold King Mine in the Animas and San Juan Rivers

Response to Peer Review Comments

U.S. Environmental Protection Agency
Office of Research and Development
Washington, D.C.

January 2017

Disclaimer

This document has been reviewed in accordance with U.S. Environmental Protection Agency policy and approved for publication. Mention of trade names or commercial products does not constitute endorsement of recommendation for use.

Table of Contents

I. INTRODUCTION

EPA's Office of Research and Development, National Exposure Research Laboratory has been conducting an analysis of the release of acid mine drainage from the Gold King Mine on August 5, 2015 and its transport and fate within the Animas and San Juan Rivers. This project's objectives are to provide analysis of water quality following the release of acid mine drainage in the Animas and San Juan Rivers in a timely manner in order to 1) generate a comprehensive picture of the plume at the river system level; 2) help inform future monitoring efforts; and 3) to predict potential secondary effects that could occur from materials that may remain stored within the system. The project focuses on assessing metals contamination in the rivers following the release of metals from the mine and during the movement of the plume and in the first several months following the release. A quality assurance project plan was developed for the work in this project.

The modeling included in the final report was reviewed with a mid-course panel peer consultation in February 2016. Versar, Inc. (Versar), an independent contractor, assembled five scientific experts with expertise in the following areas: (1) geochemistry, (2) fate and transport (water/sediment), (3) water quality analysis simulation (WASP) modeling, (4) groundwater modeling, (5) geospatial analysis (EnviroAtlas modeling), and (6) bioaccumulation. The reviewers met in Athens, Georgia for three days to evaluate the scientific integrity of EPA's analysis and characterization of the fate, transport, and potential impacts of acid mine drainage (AMD) release in the Animas and San Juan Rivers. EPA scientists presented their analysis and findings to the reviewers, and afterwards each reviewer provided his individual written response to a set of charge questions. The five reviewers were screened by Versar for scientific qualifications and any conflicts of interest. The peer consultation followed procedures specified in EPA's Peer Review Handbook, 4th edition. EPA prepared a written response to the peer consultation. (EPA/600/R-16/113, June 2016)

The modeling was updated based on feedback provided during the peer consultation, and developed into a final report. The final report was categorized as Influential Scientific Information (ISI), as defined by the OMB Peer Review Bulletin, and the report underwent a thorough peer review. It was also listed on EPA's publicly available Peer Review Agenda. Peer review of this document occurred via a letter review that was independently managed by Versar. Four reviewers were chosen to review the report and respond to charge questions pertaining to hydrology, geochemistry, fate and transport, and potential impacts from the Gold King Mine release. The same set of scientific expertise was represented as had occurred in the mid-course peer consultation. Versar conducted a thorough COI screening for each reviewer and met with EPA staff to discuss any actual or potential COI. The peer reviewers each provided written responses to a set of charge questions. Following receipt of the Peer Review Report, the final report was revised to reflect suggested changes and clarifications from the peer reviewers. The four selected reviewers are listed below.

Charles Fitts, Ph.D., Fitts Geosolutions, Scarborough, Maine
Glenn C. Miller, Ph.D., University of Nevada, Reno, Nevada
Ronald L. Schmiermund, Ph.D., Economic & Environmental Geochemistry, Inc., Lakewood, Colorado
Mark A. Williamson, Ph.D., Geochemical Solutions, LLC, Loveland, Colorado

This EPA document contains the EPA response to the independent peer reviewers' comments provided in Versar's summary peer review report. In some cases, the reviewers offered comments or opinions that were outside the scope of the charge questions and this scientific project. This document responds only to comments directed to the technical aspects related to the ORD analysis.

Comments were prepared by the EPA Office of Research and Development Gold King Mine Project Team:

Kate Sullivan, Ph.D.
Mike Cyterski, Ph.D.
Christopher Knightes, Ph.D.
Stephen Kraemer, Ph.D.
John Washington, Ph.D.
Lourdes Prieto
Brian Avants (Oak Ridge Institute for Science and Education, Contractor to EPA)

II. CHARGE TO REVIEWERS:

Part 1. Overall Project and Analysis

Question 1. Were project objectives clearly identified and did analyses address the objectives? Please explain.

Question 2. Given the data that were available to the researchers, were assumptions about data inclusion and use appropriate? How so?

Question 3. Does the analysis provide meaningful results and scientifically defensible conclusions regarding GKM plume movement and characteristics? Please explain.

Part 2. Fate and Transport

Question 4. Does the research appropriately characterize the metals concentrations and load produced from the Gold King Mine spill? Please explain.

Question 5. Were empirical methods and modeling that were used to assess plume water quality characteristics appropriately applied and interpreted given available data? Please explain.

Question 6. Were empirical methods and modeling that were used to assess deposition and bed sediments appropriately applied and interpreted given available data? Please explain.

Question 7. Were the data statistically analyzed and visualized properly in regards to metal concentrations in the surface water in the post-plume period in the Animas and San Juan Rivers? Please explain.

Question 8. Were the data analyzed and visualized properly in regards to sediment metal concentrations in the streambed in the post-plume period in the Animas and San Juan Rivers? Please explain.

Question 9. Were the geochemical principles to characterize transport and fate of acid mine drainage regarding neutralization, precipitation and mineral saturation appropriately applied and interpreted? Please explain.

Question 10. Were exposure analyses based on GKM concentration results appropriately applied and interpreted? Please explain.

Question 11. Was the potential for groundwater uptake from the Gold King Mine appropriately applied and interpreted? Please explain.

Part 3. Application of Software-Based Analytical Models

Question 12. Does the final report appropriately and adequately respond to the mid-project external peer review comments regarding the development and application of the WASP model? Please explain.

Question 13. Does the final report appropriately and adequately respond to the mid-project external peer review comments regarding the development and application of groundwater modeling? Please explain.

Question 14. Does the final report appropriately and adequately respond to the mid-project external peer review comments regarding the development and application of bioaccumulation modeling? Please explain.

III. GENERAL IMPRESSIONS

General Impressions		
Reviewer Name	**Reviewer Comment**	**EPA Response**
Charles Fitts	1) I focused on Chapters 1-6, 9, 10, and Appendix D. I have long lists of specific observations for these sections, and only write up larger items under the charge questions. I probably spent too much time in the weeds editing, but after such a close examination it was fairly easy to collect the main larger points. This report is much improved over the interim presentations we saw in February. The presentation is generally clear, although it could use rounds of proof-reading to catch typos and grammar issues. Some sections could be trimmed and clarified as I note. In many places the words chosen to describe model results are those you would use to describe reality; it is important to always include modifiers that make it clear that you are talking about simulated values, not real values. I don't find any major flaws in the overall conclusions	1) The final report and appendices have been extensively edited.
	2) I continue to think that the WASP modeling results are far less accurate and useful than the empirical model results. Omitting WASP modeling entirely would improve the strength of the report and save you a lot of tough explaining about mis-matched masses and numerical dispersion. I would look at the major project objectives and honestly assess in what areas, if any, the WASP modeling was critical to meeting the objectives	2) The EPA/ORD team appreciates the reviewers comments about use of the WASP model. We have carefully reviewed both the Empirical and WASP models for input and assumptions. As a result, we have recalibrated and improved agreement in estimates of mass and peak concentrations, as described in Appendix B. WASP settling velocities were parameterized by using the improved estimates of plume mass, and sensitivity around the settling velocity was

General Impressions		
Reviewer Name	**Reviewer Comment**	**EPA Response**
		performed. As discussed by several reviewers, WASP characterizes the plume with a greater amount of dispersion, and therefore, higher concentrations for longer periods than the Empirical Model, thus generally estimating greater mass. We have added text that all water quality models like WASP incorporate dispersion as a typically observed process that affects the movement of materials in waterbodies. In the report, we put greater emphasis on the rather unusual behavior of the plume in that it moved with a tight core of high concentration over a long distance with little apparent dispersion. In the final report, we have improved integration of the modeling results and emphasized the data-based analyses. We continue to use the WASP model as it was particularly useful for timing and movement of the plume, especially in the San Juan River.
	3) The groundwater modeling comes to reasonable conclusions, although one could come to similar conclusions in a lot less work and fewer words by just presenting local scale models of a couple of key wells. The groundwater modeling could be more realistic if it tested a broader range of alluvium K values, rather than sticking to huge-scale regional values. The regional models included vast far-field areas of bedrock with un-calibrated head values, which are a distraction and not important when you look at the key well capture zones.	3) The EPA/ORD team acknowledges the reviewer's comments regarding the groundwater modeling. The regional modeling is presented to provide boundary and initial conditions for the expanded use of local scale models. New local scale simulations using AnAqSim have been added to the final report with a broader range of alluvium K values to document simulations about the influence of explicit buried stream channels and aquifer heterogeneity and anisotropy.

General Impressions		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	4) The Gold King release in August of 2015 received extensive coverage in the media and was visually vivid in the yellow color that it gave the Animas River. The Draft Document reviews the chemical and potential biological effects of the spill and examines how this spill compares with several decades of discharge of acid mine drainage into the Animas and San Juan Rivers. After reading this report, it confirmed to me that while the spill was a serious event, the long term drainage in the Animas region is much more problematic, and this point is revealed throughout the document. The Draft is generally well written and is technically sound. As is the case with many documents of this detail, the Executive Summary will be read the most extensively, and I have some suggestions that should be considered. There is a question on the evenness of the document, in that some chapters go into great statistical detail, while others are generally more descriptive. The excellent mid-review comments from a group of experts were very useful and mostly followed, although in some cases, (e.g., Chapter 8) the basis for some of the figures is a bit unclear. In general, the figures are good, although several are difficult to read due to some of the print in the graphs is fuzzy (e.g., Fig. 8-12) or difficult to interpret (Fig. 8-2), since log plots are sometimes difficult to follow. The pictures were uniformly helpful, and showed both the vivid yellow color, but also the areas of slower flow where the iron precipitates settled. The quality of the analysis is very good, and will be useful in a variety of settings, since it brings together a large variety of disciplines to understand how receiving waters are affected by acid mine drainage,	4) The final report has been extensively edited and figure legibility has been improved.

General Impressions		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	both as a catastrophic failure, but also from continual smaller drainage. I reviewed the extensive comments of the mid-project external peer review group. They were privy to a different set of documents than I had, which consisted of the draft report, tables and figures, the appendices and the response to the mid-project external peer review group. As such, I cannot comment extensively on whether the final report appropriately and adequately responded to the earlier review. However, I did read the comments and the EPA responses and felt that the final draft report was consistent with those comments, and I can only assume that the response was adequate. I do have some specific comments, however, under charge questions, 12-14.	
	5) <u>Accuracy of Information Presented</u> – As a component of the overall *information* presented, I will consider *data* accuracy. Appendix F (QA/QC Control for laboratory analytical data) was not provided but an in-depth review of the QA/QC was outside this review, anyway. I assumed that formal QA/QC criteria were met, but was not able to determine other aspects of data quality (e.g., relationship of total to dissolved metals, ion balance, conductivity/concentration relationships etc.). Such determinations would be facilitated by inclusion of a data summary spreadsheet. Water quality data was compromised by coarse filtration practices and calls into question conclusions related to iron and aluminum chemistry.	5) A quality assurance appendix has been added to the final report. The appendix contains quality analyses of data to the extent allowable from the information that was provided by the data owners. This includes assessment of the relationship of total to dissolved metals and ion balance as suggested by the reviewer.

General Impressions		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	6) Descriptions of sediment collection, processing (e.g., sieving) and analysis (including digestion) are apparently absent in the report and engenders questions about the applicability, if not accuracy of sediment compositional data. This is important because a comparison of empirical sediment quality to WASP-predicted sediment quality seems to be the best (only?) method of validating the model.	6) The discussion of laboratory methods, especially for sediments has been increased in the final report. Specific methodology is provided in Appendix A.
	7) Hydrologic data (specifically flow data) derived from USGS gaging stations is critical to the WASP modeling and apparently suffers from problems familiar to the EPA team. The fact that steps taken to 'correct' at least one inconsistency (acknowledged by EPA in a separate communication) but not discussed at all in the report, and that other similar inconsistencies appear to this reviewer to exist, raises questions about data accuracy and application.	7) The hydrology data obtained for modeling the Gold King plume is also discussed in more detail in Appendix C, including the uncertainties in the record at a number of the USGS gages during this period and filtration practices. The flow records have an unusual amount of short-term temporal oscillations at multiple sites, creating uncertainty in tracking the relatively small plume volume through the rivers. Nevertheless, we believe that the uncertain record does not strongly affect the estimates of mass movement. The observed flow during the plume does not affect WASP in that once the model is initialized, it generates flow rather than utilizing the observed record at each site.
	8) <u>Clarity of Presentation</u> – I acknowledge that the product being reviewed is a draft, but the editorial problems are extensive to the point that they often compromise the reader's ability to understand the point being made, at least in a timely way. Often the figure and table explanations were sufficiently flawed as to prevent understanding the table or figure. I began succinctly	8) The final report has been extensively edited. The distinction between acidity and pH has been carefully documented, and corrected as applicable. Regarding metals, we state in chapter 3 and several subsequent chapters, that we exclude the major cations of Ca, Mg, Na and K from the

General Impressions		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	listing editorial comments as I came to them, but soon realized that there were too many. There are also problems with consistency and accuracy of words being used. For example, "acidity" is locally misused to describe pH, and "metals" is often used without an adequate qualifier.	metals of concern. Absent other qualifiers at specific locations in the report, our definition of metals includes the remaining analytes minus anions, nutrients and field parameters, consistent with definition for these analytes.
	9) I had trouble initially assimilating the intended purpose/necessity of recreating the plumes as a basis for fitting/calibrating the WASP model. In my experience, heavy reliance on computer models, especially in sensitive (probably defensive), arguments destined to be digested by the public, necessitates great care and transparency. The appearance of a 'black box' can be fatal and that's how the WASP model came across to me, at least initially. I believe the empirical <u>data</u> should be presented and tabulated first, with as much reliance on graphics as possible, followed by the empirical <u>model</u> with its justification, and finally by the WASP model with clear objectives stated.	9) The empirical model and WASP developed are described in greater detail in the final report to ensure transparency of the methods. The presentation generally follows the reviewers suggested organization, where the data is displayed and then how it was incorporated into the Empirical Model. The details of WASP model calibration remain in the Appendix but the display of model results with observations has been increased.
	10) I think the entire report would benefit from additional and shorter, more focused, sub-headings (sections) accompanied by hierarchal numbering. The current layout makes it difficult to keep track of the subject and context of a given section	10) More sections and sub-sections with numbering have been added to the final report.
	11) <u>Soundness of Conclusions</u> – A sound conclusion requires a valid interpretation of valid (accurate) data. Given that questions remain about the foundational data, it is impossible to declare the conclusions completely sound. However, if the data used for the analysis can be	11) A quality assurance appendix has been added where data comparability and quality is examined to the extent practicable with the available information provided with the data. Several checks on data such as ion balance, and duplicate

General Impressions		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	demonstrated to be valid, accurate and applicable, then valid interpretations and sound conclusions are possible. I believe the logic of the interpretations and deductive conclusions to be appropriate to the nature of the investigation but are dependent, in part, on resolution of data issues discussed above.	comparisons suggest that data is sound within reasonable error bounds.
Mark Williamson	12) At the outset it must be said that the text of this report is in relatively sad shape. There are numerous misspellings, incomplete sentences and outright errors. Too many to catalog in this review. Occasionally these items made it guesswork as to what the study's authors intended to say, thus potentially misinterpreting the opinions and findings.	12) The final report has been extensively edited.
	13) Editorial matters aside, the report appears to me to be an appropriate and useful effort to understand what can be understood about the impacts of the Gold King Mine (GKM) discharge given the available data (to date). In many respects I would characterize the study/report as a scoping study that seeks to constrain various potential impacts, identified as objectives of the study. It has limitations relative to solid conclusions. However, as noted throughout my comments, perhaps a bit more effort to identify, quantify, and qualify error would offer the interpretative constraints that I feel the study deserves. The report represents a considerable effort and contribution to understanding the Gold King Mine release.	13) The various analyses presented in the report are constrained by the availability of some types of data and benefitted from the availability of a considerable amount of metals sampling data. The reconstruction of the plume definitely has uncertainties whether using the observed data or utilizing models. We have attempted to explore the critical study questions to the fullest extent allowed by the available data, but there are key limitations with lack of key or well-timed data for critical aspects of analysis that merit interpretative constraints. We increased discussion of uncertainty and error in the individual analyses such concentration estimates and mass balance calculations; data limitations challenged our ability to quantify. We rely on corroboration of multiple lines of evidence when drawing synthesis and findings.

General Impressions		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	14) It is easy to be critical, with the benefit of hindsight, of a study seeking to respond to extraordinary circumstances. But the work represented by this report is an appropriate and welcome analysis. My comments below are offered in the spirit of improving clarity and constraining over interpretation.	14) We appreciate the reviewers comments and have paid attention to identifying, quantifying and qualifying error and uncertainty in these analyses.

IV. RESPONSE TO CHARGE QUESTIONS

Part 1: Overall Project and Analysis

Question 1: ***Were project objectives clearly identified and did analyses address the objectives? Please explain.***		
Reviewer Name	**Reviewer Comment**	**EPA Response**
Charles Fitts	15) Yes, I think objectives were clearly stated, and I think generally these objectives were addressed.	15) No response required.
Glenn Miller	16) The objectives were clearly defined and addressed well by the analyses, and apparently clarified in part due to the comments from the mid-project review. Chapter 2 specifically discusses what the concerns of this spill were and how they were to be addressed. A major difficulty in this analysis is due to the problem of overlaying the impacts of a major acidic spill into receiving waters that have already been contaminated by decades long drainage from a large number of smaller sources of acidic drainage. Another objects is to assess the resulting exposure of that contamination to humans and aquatic biota. When the spill occurred, I followed the news accounts of the Gold King Mine release in August of 2015 rather closely and had the same questions that were addressed in the objectives, and sought to understand the impacts of that spill, which were largely answered, and answered well in the document.	16) The EPA/ORD team acknowledges the reviewer's comments regarding the challenge of distinguishing the effects of a specific release in an area with historical acidic releases.

Question 1: ***Were project objectives clearly identified and did analyses address the objectives? Please explain.***		
Reviewer Name	**Reviewer Comment**	**EPA Response**
Ronald Schmiermund	17) I think the goals and objectives were adequately identified in Chapter 2 but could benefit from additional explanation and justification. For example: • Why quantify (and characterize) the release? Answer – to provide boundary conditions for modeling … • Why quantify fate and transport…..? Answer – to test the validity and completeness of the empirical observations, test the understanding of the river system in response to the GKM blowout and to determine where metals are likely to have been retained in the system …	17) The EPA/ORD team has added additional discussion to the final report as suggested by the reviewer to clarify the purpose of the analyses and approaches used to assess multiple aspects of the release.
	18) It seems that each objective was addressed via extensive data analysis, although the analysis is not always clearly or extensively presented. (See original submittal from R. Schmiermund for additional specific comments).	18) The EPA/ORD team has worked to ensure that analyses are presented as extensively as necessary in the main report or the appendices. The final report has been extensively reviewed and edited to ensure analyses are presented clearly.

Question 1: ***Were project objectives clearly identified and did analyses address the objectives? Please explain.***		
Reviewer Name	**Reviewer Comment**	**EPA Response**
Mark Williamson	19) Yes, the objectives of the study were very clearly identified. The objectives speak directly to concerns related to public and environmental health as well as scientific clarification and understanding.	19) No response required.
	20) While the objectives were clearly stated, and the methodologies employed were reasonable, the study was ultimately limited. This limitation is directly tied to a lack of objective-critical data, despite the abundance of data related to the mine discharge in general. The most significant data limitation relates to characterization of the discharge itself and the lack of data for the actual chemical composition of the mine pool that was released, and the characterization of the pulse passing from Cement Creek (which included erosional debris in addition to mine pool water). This lack limited the characterization of the source, and therefore constrains the subsequent downstream analysis. This situation could have, in concept, been avoided. However, under the trying, stressful and (I presume) unexpected circumstances, mobilizing to fill these data gaps were challenging and difficult to fill. Many data required filling through estimation methods and assumptions. Although there is not really much that can be done about this after the fact, it places limits on the error associated with conclusions reached in the study.	20) The EPA/ORD team agrees that key data were not available that would have enabled more complete analysis of the geochemistry of the plume and deposits afterward. We agree that characterization of the pulse of material passing through Cement Creek was one of the major uncertainties of the study. However, we note that we did have sufficient information to characterize the mass and concentration of the plume beginning in downstream reaches where sampling allowed reasonable interpretation of what occurred. This does not eliminate the uncertainty but it narrows the likely magnitude of potential error within the river system.

Question 2: ***Given the data that were available to the researchers, were assumptions about data inclusion and use appropriate? How so?***		
Reviewer Name	**Reviewer Comment**	**EPA Response**
Charles Fitts	21) Yes. Any discrepancies or details are minor and included below.	21) No response required.
	22) Given the 500+ km length of the affected drainage, from the Gold King Mine until to Lake Powell, the data that was collected was impressive, and, using hydrologic data from previous studies, the analyses were valid and well-supported. As is the case in any modelling study, assumptions need to be made in order to constrain the models to what is a reasonable interpretation of the data. In this case the analyses were based on known geochemistry of solute oxidation and precipitation of the particle bound metals. There did not appear to be any assumptions that were outside the realm of reasonableness, and the modeling efforts were largely consistent with the observed geochemistry and transport processes. The modeling results supported the empirical data, which was sometimes constrained by missing the peak plume concentrations, and the variability of analytical results that were received.	22) The EPA/ORD team acknowledges the reviewer's comments regarding an appropriate use of modeling and data.
Ronald Schmiermund	23) It is my impression that virtually all the available data were included, although it is difficult to test that impression. The mid-project peer review (Dr. Nordstrom) notes that some chemical analyses appear to be compromised due to dissolved metals exceeding total metals. The analytical data was not examined at that level for this review, but suggests screening should be done or, if already done, noted. Flow data from at least one (seemingly critical) USGS gage is suspect and was acknowledged to be so via a supplementary inquiry by this reviewer (see Assumption 1 below). A detailed and	23) We have performed several measures of data quality in Appendix F, including cation-anion balance, dissolved/total ratios, and independent methods of calculating TDS. We performed dissolved/total evaluations on the major elements. While there were some ratios exceeding unity, they generally did not fall outside of conventional lab tolerances.

Question 2: ***Given the data that were available to the researchers, were assumptions about data inclusion and use appropriate? How so?***		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	seemingly thorough reconciliation was performed and adjustments made, but were not discussed or noted. This sort of omission leads to other questions. Assuming that the data is valid, the uses of the data appear to be appropriate.	
	24) Specific Comments: The amount of data gathered is clearly impressive as was the apparent degree of consistency in collection and analytical techniques given the large number of participants. The lack of earlier water quality data at the closest Cement Ck. monitoring station and the distance to that station from the GKM were unfortunate, but still remarkable in their completeness. Similarly, the lack of water quality data at the GKM portal following the blowout was disappointing but likely explained by the conditions and accessibility. However, within the Cement Ck. watershed these data gaps necessitated a number of assumptions related reconstructing the plume.	24) The final report discusses uncertainties in the flow records available during the plume period from some USGS gages. The flow records have an unusual amount of short-term temporal oscillations at multiple sites, creating uncertainty in tracking the relatively small plume volume through the rivers.
	25) Assumption 1: The volume of the GKM "plume" (water + dissolved and suspended material derived from the GKM) flowing down Cement Ck. is assumed to be equal to the 'wave' volume or the cumulative volumetric discharge over the period of the wave's passage above base flow as reported by USGS for the 09358550 stream gage. This appears to be a valid assumption. However, inspection of the published USGS Q data for the 'wave' that reported to the Animas R. gage (09359020) downstream of Silverton	25) We do partially correct the Silverton gage for the short duration of the plume, and have explained this in Chapter 4. Nevertheless, the EPA/ORD team feels that the uncertainty in the flow record does not strongly affect the estimates of mass from the plume.

Question 2: ***Given the data that were available to the researchers, were assumptions about data inclusion and use appropriate? How so?***		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	about 15 minutes later is less than half of the wave volume in Cement Ck. – they should be approximately equal. On the surface, this discrepancy creates a major problem with respect to uncertainty about the actual volume of the GKM discharge and associated concentrations. Upon request from this reviewer, a detailed explanation provided by EPA exposed complexities in the 09359020 USGS gage data and published Q values (gage data is no longer available on the USGS website) and presented a revised estimate of the 'wave' volume at 09359020 that is approximately equal to the 'wave' volume in Cement Ck.	
	26) This revised agreement is satisfying, to be sure, but the USGS data is available to anyone and should cause the same concern for any reader. Furthermore, the fact that the arguably erroneous reported volume for 09359020 is equal to flow volumes downstream is suspicious. That is, if approximately 3 million gallons is, in fact, correct for 09359020 downstream of Silverton and the next downstream gage at Tacoma (09359500) reports approximately 1.5 million gallons, where did the balance go? There may be an explanation, but this situation is illustrative of the need for greater and more detailed explanations to accompany other assumptions, presumably in an appendix.	26) We note that the volume of the release was determined at the Cement Creek gage. The flow record at this site was very good and the reported discharge was the same as what would be determined from the adjusted rating curve published with the station. Thus, we feel the initial plume volume estimate is reliable. The uncertainty in the flow records affects confidence in tracking the plume volume through the river system as it passes downstream locations and after considerable additional water has been added to the system. A more detailed explanation has been added to Chapter 4. Relatively small changes in flow are more difficult to detect in the lower Animas and San Juan Rivers with confidence. However, reasonable estimates of mass can be determined.

Question 2: ***Given the data that were available to the researchers, were assumptions about data inclusion and use appropriate? How so?***		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	27) Assumption(s) 2 related to reconstructing the dissolved metal GKM plume: Assumptions about time-invariant mine discharge quality may be necessitated by lack of data, but are probably incorrect. A volume of 3E6 gallons translates into a great extent of flooding of the GKM tunnels and composition of the mine pool is unlikely to be homogenous. A justification/discussion of the assumption is required.	27) The reviewer's suggestion associated with the implications of time-invariant mine discharge quality are acknowledged but there was no data to inform alternatives. The final report acknowledges this.
	28) Doubling the estimated GKM discharge concentration (sentences 1048 and 1049) to account for a "first flush" seems numerically arbitrary – please justify	28) The EPA/ORD team acknowledges and describes in the final report that the choice for characterizing the first flush in the estimate of the Gold King mass at Cement Creek is rather arbitrary but informed by our technical judgement in Chapter 3. However, this phenomenon has been documented in the literature for estimating the initial flushes from soils and streams during storms, albeit for a different mode of action than occurred in the plume. The approach was discussed by the mid-project peer reviewers in light of the lack of better information.
	29) The equations given for calculating the GKM discharge quality (line 1047) makes the implied assumption that the content of the wave is a homogenous mixture of background water and GKM effluent combined in proportion to their relative input volumes at any point in time. This may or may not be completely true for the peak of the wave given the likely density of the GKM slurry that may allow the leading edge of the wave to behave	29) Similarly, the estimate for the peak concentration of particulates that were carried at the peak of the flood through Cement Creek is rationalized, but in the end cannot be verified with data. There was no evidence that a debris-flow like phenomenon reached as far downstream as the gage in Cement Creek, although channel disturbance was evident in the North Fork of Cement Creek. However, we do

Question 2: ***Given the data that were available to the researchers, were assumptions about data inclusion and use appropriate? How so?***		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	like an autonomous debris flow with limited mixing with stream water.	know that once in the Animas River the plume moved through the river system not as a homogenous well mixed unit, but that it maintained a core of segregated materials, possibly representing a material density.
	30) Assumption 3: Reconstruction of the suspended metal plume involves a different assumption (and model) relative to the dissolved metal plume. The need for a different assumption and associated model requires addition explanation to be credible	30) We used somewhat different analyses and associated processes for the dissolved plume than the suspended plume; both are adjusted relative to flow; a small adjustment is made for the dissolved background material. This is explained in more detail in the report.
Mark Williamson	31) Given the circumstances, all data related to the discharge from the Gold King Mine (GKM) are valuable and have a place in the type of analysis presented. All data would, to my mind, be included with provision for deletion upon subsequent analysis that demonstrates the extent to which they are suspect, or outliers. The use of data followed relatively conventional analysis techniques and, thus, seems to be appropriate. However, as noted above, with a compromised quantification of the source (to the Animas River), appropriate technique for analysis does not necessarily immediately confer accuracy, precision or reliability to the study conclusions.	31) The EPA/ORD team agrees with the reviewer and used all data unless it was suspect. There were a few instances where data had sufficient uncertainty that was rejected or its use was limited.

Question 2: ***Given the data that were available to the researchers, were assumptions about data inclusion and use appropriate? How so?***		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	32) I was not able to discreetly review all data to assess overall quality. I assume there are instances where such concerns are real (for example, dissolved constituent analysis reported as larger than dissolved).	32) A quality assurance appendix (F) has been added to the final report where data comparability and quality is examined to the extent practicable with the available information provided with the data. Several checks on data such as ion balance, and duplicate comparisons suggest that data is sound within reasonable error bounds.

Question 3: ***Does the analysis provide meaningful results and scientifically defensible conclusions regarding GKM plume movement and characteristics? Please explain.***		
Reviewer Name	**Reviewer Comment**	**EPA Response**
Charles Fitts	33) I would answer yes for the empirical model and no for the WASP model, as discussed below under Charge Question 5.	33) The EPA/ORD team appreciates the reviewer's comments about use of the WASP model. We have carefully reviewed both the Empirical and WASP models for input and assumptions. As a result, we have revisited and refined the empirical estimates of mass and peak concentrations, and refined the settling velocities in the WASP model, resulting in improved agreement between the empirical and WASP model approaches. In the report, we put greater emphasis on the rather unusual behavior of the plume in that it moved with a tight core of high concentration over a long distance with little apparent dispersion. In the final report we have improved integration of the modeling results and emphasized the data-based analyses. We continue to use the WASP model as it was particularly useful for movement of the plume and other aspects of the dynamics of the plume that the spatially limited empirical modeling could not do, especially in the San Juan River.
	34) I firmly believe that the analysis does provide meaningful results and is scientifically defensible. Under any circumstances, the release of 3 million gallons of highly contaminated water through a reactive waste rock dump was catastrophic and the visual impacts were seen by a very large number of people. Yellow acid mine water is not acceptable to anyone, and a large portion of the U.S.	34) No response required.

Question 3: ***Does the analysis provide meaningful results and scientifically defensible conclusions regarding GKM plume movement and characteristics? Please explain.***		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	was deeply concerned. However, the analysis provided in the document describes very well that the Animas drainage been highly contaminated for a very long time, and in fact, the release of 3 million gallons of water from the mine represented only a few days of normal drainage from the myriad of mines located in this stream basin. The task of the scientists who performed the analysis was to determine the additional burden on the receiving water and biota, and any excess exposures that might be forthcoming in the future. The analysis was meaningful and helpful for understanding the issues with acidic drainage and the incredible difficultly in management of those wastes.	
Ronald Schmiermund	35) The results being sought would surely be considered meaningful (i.e., concentrations relative to guidances, the magnitude of the metal reservoir in sediments, potential for release from sediments, etc.). Scientific defensibility is more difficult. Regarding scientific defensibility, it must be noted that use of complex models such as WASP always makes assessing defensibility challenging and the rationale for invoking WASP could be made clearer in this situation. Even the so-called 'empirical' model is complicated and could benefit from a clear explanation of its objective (presumably to fill in missing field observations and to create a synthetic data set suitable for comparison with another (WASP) model). Taken together, the approach has the appearance of validating a model with another	35) We have increased the explanation of basis for the empirical model in the final report. The purpose of the empirical model is to maximize the use of observed data to reconstruct the concentrations and mass carried in the plume. This allowed quantification of potential exposure to adverse concentrations of metals and also allowed the material to be tracked in the system. We do not use the Empirical and WASP models to validate each other but offer them as two largely independent estimates. The WASP model is required to represent the same observed empirical total mass at certain points in order to minimize drift in predictions. In between the check points, the WASP model relies on physical processes for

Question 3: ***Does the analysis provide meaningful results and scientifically defensible conclusions regarding GKM plume movement and characteristics? Please explain.***		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	model and begins to look like a house of cards. Fig. 4-13A does not inspire a lot of confidence, especially given that it represents the first downstream observation point.	predictions of concentrations. We have clarified this in the final report. The empirical model is not as much a model as it is a carefully reconstructed estimate of the plume based on observation. In that sense, the empirical model offers a relevant interpretation of observations to which WASP can be compared, at least in terms of total mass at the observation points. We have carefully reviewed both the Empirical and WASP models for input and assumptions. As a result, we have revisited and refined the empirical estimates of mass and peak concentrations, and refined the settling velocities in the WASP model, resulting in improved agreement between the empirical and WASP model approaches. However, they still provide independent estimates of movement, concentrations, and mass through the rivers.
	36) Specific Comments: A helpful approach to the report might be to first acknowledge the empirical data gaps (which has been done adequately), then describe the need to combine the available data into a single 'best fit' synthetic data set to fill in the holes, describe the methods used to do so, present the synthetic data set, and finally justify the need	36) The EPA/ORD team appreciates the suggestions of the reviewer. The presentation and description of the models has been clarified and expanded in the final report. Plume timing is a major contribution of the WASP model that assists in development of the Empirical Model.

Question 3: ***Does the analysis provide meaningful results and scientifically defensible conclusions regarding GKM plume movement and characteristics? Please explain.***		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	for WASP. I assume the latter is to allow for a contaminant mass balance	
	37) When explaining the WASP model the first effort should be to validate it against the actual and synthetic ('empirical') data base, starting from the large scale (e.g., plume timing from source to Lake Meade), then move to the smaller scale (e.g., matching plume shape, peak concentrations etc.). This is done in Figs. 6-19 and 6-21 for sediments, but should be more prominently presented. • Conclusion 1 (line 3811) – The basis and credibility of the release characterization should be made clear (i.e., inferred from post blowout data, assumptions about time invariance and data collected in Cement Ck. at Silverton) • Conclusion 2 (line 3826) - Acid neutralization upon mixing with Cement Creek (line 3847) is cited for inducing precipitation of iron and aluminum oxy-hydroxides from clear, low-pH water. Indeed, quiescent flow from a large diameter pipe in 2009 shows clear water and photos of the mine pool post blowout is described as clear (Fig. 3-7). However, other photos suggest water with abundant suspended iron oxyhydroxide exiting the portal before and after the blowout. Add field observations to clarify.	37) Additional discussion has been added to the final report clarifying the development and comparison of the Empirical and WASP models. • Discussion of assumptions has been increased in the final report. The report provides 1 year of data from the mine post release that characterizes some of the time variance discussed by the reviewer. • The geochemical analyses suggested that a large portion of the precipitation of iron and aluminum oxy-hydroxides probably occurred in the Animas River. Photographs of the mine show clear water prior to the release, but during the release there was a mixture of yellow colored water and mud in the photos. We have clarified photos and added some additional description of field observations.

Question 3: ***Does the analysis provide meaningful results and scientifically defensible conclusions regarding GKM plume movement and characteristics? Please explain.***		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	• Dr. Nordstrom (mid-project review, Question 11, Comment 89 and 91) discussed the value of carbonate phase saturation index calculations as a means of elucidating the interaction of Cement Ck and Animas R. waters. He also recommends additional mixing calculations. This reviewer attempted to follow-up on that suggestion only to find that results of the empirical modeling (i.e., synthetic peak compositions) were not included in the report. I recommend that some empirically modeled peak compositions be presented.	• Empirical peak concentrations have been added to the final report in Appendix E.
	38) Conclusions related to the mass balance could be better stated with consistent percentages and a figure. It would also be helpful if various conclusions related to increases relative to background or ambient conditions could be put into context with some statistics (e.g., x% greater than the background mean).	38) The final report improves clarity on mass balance numbers. The report does not generally provide many values as percentage greater than background for the plume, instead describing changes as orders of magnitude.

Question 3: ***Does the analysis provide meaningful results and scientifically defensible conclusions regarding GKM plume movement and characteristics? Please explain.***		
Reviewer Name	**Reviewer Comment**	**EPA Response**
Mark Williamson	39) As noted above, many important data related to the study objectives were either not collected, or necessarily estimated. Thus, the extent to which the study analysis is meaningful and/or scientifically defensible must be judged with respect to the error associated with estimates and conclusions. Obviously, simply following an appropriate methodology does not assure meaningful-ness and defensibility in the presence of incomplete data. That said, the analysis does provide value and perspective while also providing a solid basis for continued monitoring and interpretation to refine initial conclusions and findings. A fuller description and discussion of errors and their impact on finding might prove helpful. Absent a rigorous propagation of errors, perhaps there is value in a comparison of findings for minimum and maximum constraints. Such approaches can separate findings that are strongly supported from those that remain speculative.	39) The final report contains more discussion of errors and uncertainties.

Part 2: Fate and Transport

Question 4: ***Does the research appropriately characterize the metals concentrations and load produced from the Gold King Mine spill? Please explain.***		
Reviewer Name	**Reviewer Comment**	**EPA Response**
Charles Fitts	40) The mass of dissolved metals released from the GKM is based on mine water chemistry about 10 days after the release, when the mine was open, not closed. Is there any way to estimate what the likely differences would have been between pre-release (closed mine) and post-release (open mine) concentrations? Were there any analyses of GKM seepage water before the release? I'm not a geochemistry expert, but perhaps using the Aug 15+ equilibrium pH and DO compared to the pre-release GKM effluent pH and DO (I assume there are such data) and equilibrium modeling could yield estimates of the pre-release mine water chemistry. Even if this sort of analysis/discussion is qualitative, it would be helpful. See the last sentence of the caption for Fig. 3-8. The concentrations in the 7 Aug sample are significantly higher than later mine water samples; is the difference mostly in colloid/particulate? Could the 7 Aug chemistry be closer to earlier concentrations? Why were these higher 7 Aug concentrations not weighed more than later concentrations? The text, tables, and figures in the *Metals Released From the Mine* section should always clearly state whether concentrations are dissolved, colloid/particulate, or total. In many places, this wasn't clear. My impression is that the concentrations discussed in this section were mostly dissolved, but that some samples were total.	40) The EPA/ORD team has not identified any data from the mine pre-release. There has been abundant data collected since the water treatment facility was constructed post event (presented in the report). The initial samples post release were higher in colloidal/particulate metals. Increases particulates has been noted in the literature after these kind of events with a return toward equilibrium. We calculated that the water turnover was probably 5 days pre-release, and we noted that an equilibrium appeared to be established in that timeframe in the sampling. The concentrations assigned to the release were 99.5% dissolved. The final report has been edited to clarify whether concentrations are dissolved, colloid/particulate, or total.

Question 4: ***Does the research appropriately characterize the metals concentrations and load produced from the Gold King Mine spill? Please explain.***		
Reviewer Name	**Reviewer Comment**	**EPA Response**
Glenn Miller	41) The data set that was generated by many groups (federal, state, local and tribal) was large, and given the constraints of conducting sampling at precisely the correct times to catch the maximum concentrations of metals in the plume, the data collected was used effectively in the models to estimate the fate of the contaminants as they traveled from a highly acidic origin to regions of the drainage where the pH rose and the metals (particularly iron and aluminum) effectively precipitated with other metals. While the total load of metals released into Cement Creek will never be known with great certainty, the sampled water and analyses conducted on the various streams allowed a reasonable estimate to be made. Additionally, the water quality measurements provided in the storm event that occurred shortly after the spill and the spring runoff all provide additional data to support the estimates of how the spill affected the receiving waters all the way to Lake Powell.	41) The EPA/ORD team acknowledges that some datasets may have been incomplete in what was measured, but also when it was measured. Computer simulation models allowed some interpolation and extrapolation from the existing data, and the testing of hypotheses.
Ronald Schmiermund	42) Not entirely. Characteristics of the Level 7 portal effluent, and the derived ‘slurry’ containing eroded waste dump material, should be considered ‘inferred characteristics’, given the lack of empirical data collected from the site itself. We lack pre-blowout water quality at the portal, actual blowout water, confirmation of the time-invariant effluent quality assumption and estimated volumes of eroded waste dump. The approach to the dissolved component is unsatisfying, but probably the best that can be done.	42) The EPA/ORD team agrees with the reviewer that there are many critical uncertainties with information regarding the release from the mine and delivered from Cement Creek to the Animas River and that these are inferred characteristics. The final report discusses these uncertainties.

Question 4: ***Does the research appropriately characterize the metals concentrations and load produced from the Gold King Mine spill? Please explain.***		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	43) Specific Comments: Line 924 suggests that pre-blowout samples could not be collected due to the GKM tunnel being sealed. This may be misleading given photos that show water was being released during and prior to construction activities, and appears to have been actively flowing in a corrugated ditch prior to the blowout.	43) The mine has a complicated history of blockage and opening as documented in the Bureau of Reclamation Report. The EPA/ORD team was not able to discover any pre-event samples or data to assist us in estimating the effluent. Therefore, we relied on post-event data. We changed the language in the referenced sentence.
Mark Williamson	44) The characterization of the release form the GKM is problematic, and will remain so. There is a lack of water samples (and analysis) from the released mine pool (initial water released) and characterization of the early time and bulk discharge from Cement Creek. It is possible to *constrain* the metals concentrations and the discharge from Cement Creek. Given the empirical nature of characterization such as associated with the GKM, one either has the right samples, or not. In the present case, not so much. The researchers were required to make estimates, which is fine and appropriate. Their approach is one that I would probably use. But the results may not be appropriate, in the sense of not being of the highest quality and scientifically less defensible for the conclusions to be reached later in the study. It simply increases the width of the error bars that need to be discussed relative to the conclusions reached.	44) The EPA/ORD team agrees that there is uncertainty about the initial plume construction in Cement Creek in particular, given lack of sampling. However, sampling from 64 kilometers of the source through the rest of the Animas River did provide good quantification of the plume concentrations and mass, thus constraining the errors. The largest uncertainty due to lack of data is how much mass may have actually been deposited within this river reach.
	45) I would anticipate that initially the GKM discharged water with high concentrations of metals (and other constituents), which is largely consistent with the study. I	45) The final report includes more discussion about uncertainties with the initial plume estimates. The report provides data on the effluent for the year

Question 4: ***Does the research appropriately characterize the metals concentrations and load produced from the Gold King Mine spill? Please explain.***		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	would also, however, expect that a rather large mass of sludge to be discharged as well. This would contribute to the chemical mass attributable to the GKM (as distinct from that derived from erosion of waste rock, tailings and other debris in Cement Creek). In time, the mine pool might have returned to pre-spill conditions (as assumed), but it seems unlikely given the introduction of oxygen and the exposure of material previously submerged by water. My experience has been that once opened, old mine workings' discharge is routinely higher at the outset, and diminishes to a new steady state. Although the geochemical evaluation (Appendix D) claims to have made "conservative" estimates, the issue is still problematic and the uncertainty should be better represented in later report discussion. I would probably propagate a maximum and minimum source (Cement Creek discharge) through the subsequent downstream assessment to bind the conclusions. These comments in no way represent a negative assessment of the work conducted as much as a call to highlight the uncertainty and acknowledge strongly that the discharged chemical mass cannot be known conclusively. To the extent the uncertainty does not compromise later conclusions, discuss that prospect in the report text.	after the event, but we are clear to say that there is no guarantee that the effluent concentrations are the same post-event as they were pre-event. However, at the reported leakage rates, there appeared to be a fairly rapid turnover rate within the mine.

Question 5: ***Were empirical methods and modeling that were used to assess plume water quality characteristics appropriately applied and interpreted given available data? Please explain.***		
Reviewer Name	**Reviewer Comment**	**EPA Response**
Charles Fitts	46) The empirical methods seem reasonable. The shape factor discussion needs to be clarified and so do some of the associated Figures (see my notes under Specific Observations). The WASP model simulates more longitudinal dispersion than the data indicate (Fig. 4-13). The WASP dispersion could be mostly numerical, given the several km length of each WASP segment, time stepping, and the assumption of thorough mixing within each segment at each time step. There should be more detailed discussion of the broader WASP-simulated plume compared to the empirical data-based plume, and there should be a discussion and analysis of numerical dispersion and whether that was the main cause of the excess dispersion in the WASP model.	46) Additional discussion and presentation of the "shape factor" has been added to the final report in Chapter 4. Additional work and discussion on numerical dispersion in the WASP model has been incorporated into Appendix B. We add discussion on how much numerical dispersion WASP adds, how introducing negative dispersion could affect results, and how increasing the minimum time step affects numerical dispersion. No longitudinal dispersion is added into the model. All dispersion in the model is numerical.
	47) The argument that the first observed yellowboy coincides nicely with the broader dispersion on the climbing limb of the WASP simulation (Fig. 4-12) is not a strong one. Given the intensity of the yellowboy in the Animas River, it could probably have been noticeable at a tiny fraction of the peak concentration, well out ahead of the empirical plume peak.	47) Separate from formal model calibration/validation, the report compares anecdotal observations of plume arrival with WASP predictions. This was not done to justify the models timing accuracy but simply to provide a visual reference. A quantitative assessment of WASP travel timing to peak is provided in the final report.
	48) Looking at Figure 4-13, at all stations except the first, the total mass in the WASP model plume is significantly larger than the total mass in the empirical model plume (mass is proportional to area under the curve). There seems to be an effort to match the peak concentrations,	48) We have clarified that all water quality models like WASP incorporate dispersion as a typically observed process that affects the movement of materials in waterbodies. In the report we put greater emphasis on the rather unusual behavior of

Question 5: ***Were empirical methods and modeling that were used to assess plume water quality characteristics appropriately applied and interpreted given available data? Please explain.***		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	which with the greater dispersion of the WASP model, means it is overstating particulate+dissolved mass in the plume by a significant amount. Since the upstream input mass appears correct (note reasonable match of WASP and empirical models in Fig. 4-13A), the model systematically underestimates the mass that settles from the water column to the river bed at downstream locations. This and the excessive dispersion are significant deviations from reality in the WASP model. If these issues can't be overcome in the WASP model, it may be best to drop the WASP modeling altogether or limit use of WASP to simulate plume travel in the San Juan River based on empirical inputs at Farmington. I think the empirical model is good a representation in the Animas River, and the most realistic way to estimate mass transfer to/from bed sediments in the reaches from one station to the next. I'm more comfortable with that analysis for the Animas River than with the WASP model which overstates dispersion and mass in the water column and understates the mass transferred to sediment. It will be difficult to defend the WASP model results, but not the empirical model results. The report essentially admits this on p. 44, lines 1434-1435 where it states *We believe that the Empirical Model reflects the passage of the core of the plume and bulk of metals better because it is tied to field observations…*	the plume in that it moved with a tight core of high concentration over a long distance with little apparent dispersion. The final report clarifies that the peak concentrations were not matched between the WASP and Empirical models for calibration. The mass of the plume estimated by the Empirical Model was used to determine particle settling velocities in WASP. In the final report we have improved integration of the modeling results and emphasized the data-based analyses. We continue to WASP as it is particularly useful for movement of the plume and other aspects of the dynamics of the plume that the spatially limited empirical modeling could not do, especially in the San Juan River.

Question 5: ***Were empirical methods and modeling that were used to assess plume water quality characteristics appropriately applied and interpreted given available data? Please explain.***		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	49) The trace metals – aluminum signature discussed on pages 53-54 seem to be a tool that could be used as empirical evidence of plume timing as it moved through the San Juan, but I did not see this employed in Chapter 4. Perhaps some of the Chapter 4 empirical model Figures for the San Juan River could show the timing of samples with the trace-aluminum anomalies indicative of the plume.	49) Using the Aluminum correlation to identify plume samples in the San Juan is a useful suggestion, and we used this approach in snowmelt calculations.
Glenn Miller	50) While I am not a modeler, the use of the WASP model was helpful in that it could be used to explain how the particulate mass acted in the rivers. The combination of the model and the empirical data resulted in picture that helped the reader to understand the dynamics of the spill, which were constrained by the analytical data produced, as well as the variable flow characteristics of the streams. It is entirely reasonable to assume that a high gradient stream with rapid movement will maintain a high suspended sediment load (and particulate from the spill), while a slower moving lower gradient stream will deposit greater amounts of suspended material in the stream sediment, which is largely what the model accomplished. The water quality clearly improved as the plume moved downstream, both in response to dilution, but also to particulate aggregation and deposition in the bottom sediments, where they will contribute to an existing elevated concentration from historic drainage.	50) No response required.
Ronald Schmiermund	51) Yes, I believe so. However, the explanations provided on pages 39 through 41 made it difficult to follow. After reading and re-reading p. 40 and bouncing between	51) The final report has additional discussion of the basis of the Empirical Model, including more discussion of the plume shape and verification

Question 5: ***Were empirical methods and modeling that were used to assess plume water quality characteristics appropriately applied and interpreted given available data? Please explain.***		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	figures, I got the essence of the approach, but the reader should not need to do that. Specific Comments: The explanation of plume shape seems especially weak.	against observed metals concentrations. The flow of the text in this section has been improved.
Mark Williamson	52) It is difficult to find fault with empirical methods for situations such as the GKM discharge. Things are happening quickly and there is little or no time for forethought. Also, as might be expected in the case of the GKM, there was more than one team collecting samples/data. Not all can be expected to use identical approaches, although one should expect them to be in reasonable agreement with each other and standard approaches.	52) No response required.
	53) Owing to the challenges of the situation, most monitoring locations did not capture data related to the peak of the GKM plume passage. This is unfortunate, but somewhat understandable. In light of the missing data, and the need to speak to the totality of the plume, it became unavoidable that some data would need to be estimated for those peak plume times when empirical data were not collected. I think that the modeling techniques used to infill these data gaps were basically appropriate. As elsewhere, this is another source of error, and I found that consideration of error (limitation of conclusions) was not amplified as much as perhaps it could be to constrain some of the conclusions reached. It seems as though a useful modeling opportunity was missed however. I	53) Additional discussion of error and uncertainty has been added to the final report. Geochemical mixing models were in fact depicted using Geochemist's Workbench with Animas alkalinity for the estimated "Peak Concentration" (Fig. C-9), "Plume + Cement Creek Background" (Fig. C-10) and for "Cement Creek Background" (Fig. C-12). Additional discussion has been added to Appendix C.

Question 5: ***Were empirical methods and modeling that were used to assess plume water quality characteristics appropriately applied and interpreted given available data? Please explain.***		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	would have been inclined to utilize PHREEQC or Geochemist's Workbench to conduct a few mixing simulations combining the estimated GKM discharge with Animas River water (from upstream of Silverton) to assess the outcome and compare to field observations. This is not a critical feature, perhaps only an opportunity missed. This could have taken the place of many geochemical calculations (discussed in Appendix D) to illustrate geochemical processes that account for field observations.	

Question 6: ***Were empirical methods and modeling that were used to assess deposition and bed sediments appropriately applied and interpreted given available data? Please explain.***		
Reviewer Name	**Reviewer Comment**	**EPA Response**
Charles Fitts	54) As noted above and admitted in the text, the WASP model shows too high a mass in the water column and too little mass transfer to the river bed in the reach between Silverton and Durango. Perhaps this is the result of tuning the WASP model to match peak concentrations combined with WASP's too-large dispersion. If the WASP model had been tuned to match total plume mass, peak concentrations would have been lower, but it would have had more appropriate mass transfer to the river bed and been closer to observed conditions and the empirical model. Since the WASP mass balance between water column and river bed is not correct, its results regarding	54) The WASP model and Empirical Models have been re-calibrated to reflect this reality.

Question 6: ***Were empirical methods and modeling that were used to assess deposition and bed sediments appropriately applied and interpreted given available data? Please explain.***		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	deposition/resuspension are difficult to defend and should fixed or not presented. The empirical model matches observed water column data which is as good as can be done.	
	55) The WASP model made the greatest underestimation of deposition in the rapid canyon area below Silverton, yet photos in Fig. 6-11 C, D and 6-14 (last one) show significant deposition in this reach. Either the WASP results or the empirical results are not correct for this reach; given that the empirical is data-based, it is probably the correct one.	55) Resuspension analyses for snowmelt (Figs. 6-27 to 6-30) have been largely dropped in the final report and replaced with analysis of observed data. We note, however, that observations were similar to WASP predictions.
	56) The discussion that accompanies Figs. 6-27 to 6-30 was hard to follow. I could not always understand the explanation of these analyses. Since much of this is based on WASP, which is not accurately representing settling vs suspension, I am leery of the conclusions. I think a much more compelling approach to the resuspension questions are the sampled concentrations during the spring 2016 snowmelt (Figs. 6-24 and 6-25), which are within historic ranges for the most part. That point was not made in the text.	56) We have carefully reviewed both the Empirical and WASP models for input and assumptions. As a result, we have revisited and refined the empirical estimates of mass and peak concentrations, and refined the settling velocities in the WASP model, resulting in improved agreement between the empirical and WASP model approaches, as explained in Appendix B.
Glenn Miller	57) See response to Question 5. Additionally, the geochemistry of the spill is largely controlled by the pH of the water, and the oxidation rates of iron, which convert soluble ferrous iron to insoluble ferric iron (as the pH is raised). Most of the metals in the drainage (copper, lead, zinc, aluminum, iron etc.) are governed by their solubility, which are reduced as the pH is raised, and also	57) No response required.

Question 6: ***Were empirical methods and modeling that were used to assess deposition and bed sediments appropriately applied and interpreted given available data? Please explain.***		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	the particulate sorption that promotes attachment to the particles.	
Ronald Schmiermund	58) Presumably the empirical model for sediments consisted of a mass balance based on 'colloidal/particulate' mass multiplied by wave volume, initialized as the calculated Cement Ck. 'colloidal/particulate' mass. If so, this appears to be appropriate. However, as pointed out in my response to Question 2 (Assumption 1) there are inconsistencies in the volume of the wave as one might calculate it from the published USGS gage record, which casts doubt on the model for bed sediments.	58) The mass of the plume was calculated by combining flow and reconstructed concentrations on 15 minute intervals during the time of the plume. This was as opposed to a "bulk volume." This calculation is not as sensitive to the flow records as a bulk calculation would be. Especially in that the plume volume was small compared to river volume at most of the river stations downstream from the Animas headwaters. This is described in Chapter 4.
	59) Specific Comments: A tabulation of settled-upon 'wave' volume at each gaging station would be useful along with an explanation of any adjustments made to the data.	59) A tabulation of the wave volume at each gaging station in the Animas River has been added to the final report.
Mark Williamson	60) Given the potential for underestimation of the GKM chemical mass discharge, and that about 50% of the estimated plume volume seems to disappear, estimates of metal removal, as a percentage of GKM discharge in particular, or Cement Creek in general, may be off. It seems appropriate to develop and offer some sense of the magnitude of uncertainty.	60) Cement Creek had a good hydrology record and an exceptionally clear signal of the mine passage. From a flow perspective, we believe the plume is well quantified at this location.
	61) The GKM discharge and lost plume volume notwithstanding, the discussion of uncertainty and the	61) We have carefully reviewed both the Empirical and WASP models for input and assumptions. As a

Question 6: ***Were empirical methods and modeling that were used to assess deposition and bed sediments appropriately applied and interpreted given available data? Please explain.***		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	empirical versus WASP model that is presented is a good contribution. I do wonder why the empirical model (field data) was not more influential in calibrating the WASP model. The differences between the two models is presented, but perhaps not sufficiently reconciled. The empirical model is more mapping and less model and seems it should/could be used to adjust the WASP calculations (although I am not familiar with WASP and its intricacies). Further, as a model like WASP would seem to be most beneficial in the San Juan River reach, efforts to calibrate it in a (relatively) more constrained reach of the Animas might be beneficial in interpretation of the San Juan?	result, we have revisited and refined the empirical estimates of mass and peak concentrations, and refined the settling velocities in the WASP model (based on the empirically estimated plume mass at different locations), resulting in improved agreement between the empirical and WASP model approaches. We have performed additional analyses on the WASP structure (extent of numerical dispersion compared to commonly observed dispersion in rivers, effect of increasing minimum time steps, and introduction of negative dispersion). We have expanded discussion on WASP and model output in Appendix B.
	62) For the San Juan River reach, I am curious why a simple mass balance mixing model was not investigated to assess the transport of GKM contributions. It is noted in the report that lead (Pb) was enhanced in the Animas River relative to San Juan. It follows then that normalization of other parameters relative to lead in a mixing model between the San Juan and the Animas might reveal some things about the transport of constituents from GKM. Perhaps it was tried and, having no real positive contribution, was not discussed in the report.	62) The final report now includes a discussion of mass transfer of a number of metals between the Animas and San Juan Rivers during the plume.

Question 7: ***Were the data statistically analyzed and visualized properly in regards to metal concentrations in the surface water in the post-plume period in the Animas and San Juan Rivers? Please explain.***		
Reviewer Name	**Reviewer Comment**	**EPA Response**
Charles Fitts	63) Not my area of expertise.	63) No response required.
Glenn Miller	64) I believe the data were appropriately analyzed using statistical methods. This point was examined carefully by the authors, with the desire to attempt to disentangle the load released from the GKM, compared to the sediment metals load that had been released over the previous many decades. While some of the figures in the last three chapters were difficult to follow, due to the difficultly in reading the figures (at least on my computer), it was apparent from the statistical treatment that while the contribution of the GKM is certainly not trivial, the loading from historical discharges forms a much larger sediment load. As described in the document, some increased release of lead and zinc can be ascribed to the GKM spill, although that concentration is likely to return to the base conditions that depend on the meteoric events, including storm runoff and spring melt. In summary, the statistical treatment of the loading appears to be valid and useful.	64) The EPA/ORD team attempted to use empirical lines of evidence to compare the loads associated with the acute GKM release in context of the long term more chronic release from regional acid mine drainage. The final report provides clearer figures.
Ronald Schmiermund	65) Because I'm not a geostatistician, I am cautious to comment on this issue. However, it seems that the word "statistical", which appears in the text 58 times, is sometimes used in a very general way and implies a greater degree of statistical analysis than was possible with the data available. Lines 2842-2848 describe the difficulties of applying statistical testing in this case and do not inspire a lot of confidence in the approach. Were	65) Limited statistical testing was done to avoid egregious error inflation. The EPA/ORD team applied both parametric and non-parametric tests and used concurrence between the two to ascribe more confidence to the results. This is discussed further in Chapter 9.

Question 7: ***Were the data statistically analyzed and visualized properly in regards to metal concentrations in the surface water in the post-plume period in the Animas and San Juan Rivers? Please explain.***		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	any other more transparent approaches considered? (e.g., normalizing concentrations to flow, presenting analyte ratios (e.g., normalization to a conservative analyte like sulfate), etc.).	
	66) Table 8.6 addresses some pre- and post-event dissolved and total metal concentration. Please provide date ranges for pre- and post-event sampling. Were the criteria for log normality met? Explain the colors as supporting or rejecting the null hypothesis.	66) Dates are indicated in the table (caption) or text in the final report.
	67) Regarding statistics applied to sediments: Table 8-5 seems to be the critical table for supporting one conclusion about bed sediments (lines 3946-7) and should be more prominently presented. The statement that "Concentrations were logged…" implies log-normal distributions – did they meet the criterion for normality? – this would justify the two different tests listed. Identify "SE", presumably 'standard error'. The caption is inconsistent with the text (p. 84) where snowmelt samples are described as being collected between mid-April and mid-June 2016 – which are the 'pre-event' and 'fall 2015' samples?	67) We did limit comparisons to concentrations taken at approximately the same flow levels (see Figure 9-7). We also used an approach (detailed in Chapter 5, see Figure 9-42) where concentrations of metals were examined in relation to the amount of aluminum present as a method for detection of GKM influence. We present the p-values of each test, but avoid interpreting the significance of those results (i.e., rejecting the null hypothesis) given the data limitations, as described in the text. Additional discussion has been added to this section. Statistical tables and the main text of this chapter have been adjusted to reflect reviewer comments.

Question 7: ***Were the data statistically analyzed and visualized properly in regards to metal concentrations in the surface water in the post-plume period in the Animas and San Juan Rivers? Please explain.***		
Reviewer Name	**Reviewer Comment**	**EPA Response**
Mark Williamson	68) Generally, I find no particular concerns with the presentation of metal concentrations post-plume. However, I do find figures 8-2 and 8-3 a bit less useful than they might be if they illustrated samples that were taken pre- and post-plume.	68) Pre and post plume samples have been identified on the figures in the final report.

Question 8: ***Were the data analyzed and visualized properly in regards to sediment metal concentrations in the streambed in the post-plume period in the Animas and San Juan Rivers? Please explain.***		
Reviewer Name	**Reviewer Comment**	**EPA Response**
Charles Fitts	69) Not my area of expertise.	69) No response required.
Glenn Miller	70) See Question 7. As indicated above, the GKM discharge and sediment loading do add to the overall loading in the sediments, although it is a relatively small component, based on the statistical treatment presented in the document.	70) No response required.
Ronald Schmiermund	71) Much effort has obviously been expended in presenting data in graphic form. Unfortunately, work remains to be done to clean-up and clarify many graphs and associated captions.	71) The final report has been extensively edited.
	72) Specific Comments: Fig. 6-15 – The geochemical modeling used to generate the precipitate masses should be accompanied (in an appendix) by a complete list of the input parameters (in addition to the thermodynamic constants involved that	72) The input parameters for the simulations are listed in Appendix C, Table C-11, which has been expanded based on the peer reviewer's comments. The stable minerals shown in Figure 6-15 were evaluated using actual data and saturation indices

Question 8: ***Were the data analyzed and visualized properly in regards to sediment metal concentrations in the streambed in the post-plume period in the Animas and San Juan Rivers? Please explain.***		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	should appear elsewhere) so that the results could be checked. Were the more stable phases (right section) determined by re-equilibrating the precipitated phases with 'fresh' Animas R. water?	for the hydroxide minerals, and this is explained in the report text.
	73) Fig. 6-16 – It should be stated in the caption that multiple samples were collected in the same spot?, in the same interval of river?, over what period of time? And the 'n' should be provided. As Dr. Nordstrom suggested, multiple plots for each element of importance would be informative.	73) The caption for Figure 6-16 has been revised to provide the requested information. Multiple plots for individual metals in both the Animas and San Juan Rivers have also been added to the final report
	74) Fig. 6-17 – I assume that the orange line results from WASP modeling (please label). It seems to me that this type plot is one test of the WASP model's accuracy and should contain more information on empirical observations. The "A", "B" etc. labels should have lines to the plot indicating the exact point or river interval being discussed in the caption. What is 'Total Sediment Concentration'?	74) Figure 6-17 has been removed from the final document.
	75) Fig. 6-18 – This figure combined with Fig. 6-17 seems to me to contain the critical 'take-aways' for the sediment studies. They are, however, not very satisfying. First, be consistent in the concentration units used between the two figures. Fig. 6-17 would be better if presented for individual elements, or, Fig. 6-18 would benefit from superposition of the WASP model for individual elements (captured in Fig. 6-19). Please provide date ranges for the various data sources. Box-and-whisker plots for the post-	75) Figures have been added for individual elements and with improved presentation of observations in the final report.

Question 8: ***Were the data analyzed and visualized properly in regards to sediment metal concentrations in the streambed in the post-plume period in the Animas and San Juan Rivers? Please explain.***		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	release data might be useful if the horizontal scale was expanded. This plot seems to me to be compelling data to support a return to background water quality, at least in some reaches and should be emphasized.	
	76) Fig. 6-19 – This is the most important Figure for sediments and should be the basis for conclusions. Why was the plot not extended to the San Juan? Identify the open circles as was done in Fig. 6-18. The USGS gage data shown in Fig. 6-19 does not agree with Fig. 6-18 (e.g., no station shown at AK≈20, 60 and 70 on Fig. 6-18).	76) The plot has been extended to the San Juan in the final report. The figures have been edited.
Mark Williamson	77) As with charge question #8 above, I find no particular concerns with the presentation of metal concentrations post-plume.	77) No response required.

Question 9: ***Were the geochemical principles to characterize transport and fate of acid mine drainage regarding neutralization, precipitation and mineral saturation appropriately applied and interpreted? Please explain.***		
Reviewer Name	**Reviewer Comment**	**EPA Response**
Charles Fitts	78) Not my area of expertise.	78) No response required.
Glenn Miller	79) I have added some comments in the specific comment section in this regard. However, in general, the geochemical treatment of the spill and how the chemistry changes over time is examined correctly. Basically, the very acidic water that came from the mine water running over a reactive waste rock dump is neutralized as it is diluted and neutralized with alkaline water downstream in the Animas River and ultimately in the San Juan river, the iron is oxidized to ferric iron and both aluminum and iron precipitate readily either as various aluminum and iron precipitates, or binds to other particles that aggregate and precipitate in the sediments, particularly as the energy of the water is reduced when it traverses regions with low elevation loss. The models used the geochemistry appropriately, and the results tend to describe the outcome of the spill contaminants with scientific rigor.	79) The EPA/ORD team acknowledges the high level view that the GKM release experienced natural dilution and neutralization as it flowed from Cement Creek to the Animas River, continuing into the San Juan River.
Ronald Schmiermund	80) The application of geochemical principals is discussed in Appendix C. I have no issue with principals, but do question the data and modeling used implement those principals. Specific Comments: • Obviously a great deal of the geochemistry is about, and dependent upon, iron and aluminum, however the analytical data for both, but especially aluminum, are	80) Much of the discussion in the Spring 2016 Peer review on 0.45 μm filtering was spurred by a flaw in the draft interpretation presented. In that earlier interpretation, non-detects for Fe and Al were mistakenly included having the value of their limit of detection. For these nondetect samples, this mistake caused apparent gross over-saturation with respect to amorphous $Fe(OH)_3$, saturation indices of 5 and 6 for example. With the nondetects now correctly reported as nondetect, calculated

Question 9: ***Were the geochemical principles to characterize transport and fate of acid mine drainage regarding neutralization, precipitation and mineral saturation appropriately applied and interpreted? Please explain.***		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	compromised by the coarse (0.45µm) filtration. This issue is alluded to in lines 708-721 but seemingly ignored in the interpretation of the geochemical modeling. Why go to the trouble of producing reaction models (Figs. C-9, C-10 and C-12) when the input data is likely compromised?	saturation indices for $Fe(OH)_3$ and $Al(OH)_3$ all fell in a reasonable range, mostly centering at about SI~0.5 (Figs. C-13 and C-14). While Nordstrom has shown 0.45 µm filtering to be a potential problem for fresh precipitates, the appearance of Figures C13 and C14 suggests that use of 0.45 µm filters yielded analytical data that comports reasonably with thermodynamic calculations. Given the reasonable results for these data depicted in Figs. C-13 and C-14, proceeding with the calculations is justified, consistent with original recommendation from Dr. Nordstrom.
	• Much attention is given to the neutralization processes in the Animas River that result in the formation of initially suspended and later precipitated iron oxy-hydroxides. No doubt this takes place. However, some photos clearly record bright orange water exiting the GKM portal prior to and during the initial minutes after the blowout (other show clear water). How does this affect the reconstruction of the GKM blowout chemistry?	• Much of the bright orange color recorded of the discharge during the early release might well be soil suspended by the vigorous, turbulent flow. Part of it might also be mine waters oxidized near the dammed mine opening as atmospheric O_2 diffused into the mine, but the extent of Fe oxidation in the minepool seems likely to be small overall for a combination of reasons we describe in the report under "Characterization of the release volume and chemistry." Unfortunately, no sample analyses were discovered of the pre-release minepool to confirm a best inference. Regardless, the assumption that the Fe was dissolved Fe^{2+} is conservative in that Fe oxidation and precipitation (to generate the bright orange color) are acid-producing reactions, according

Question 9: ***Were the geochemical principles to characterize transport and fate of acid mine drainage regarding neutralization, precipitation and mineral saturation appropriately applied and interpreted? Please explain.***		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	• The use of geochemical modeling such as Geochemists Workbench is utterly and totally dependent on the thermodynamic data base. My experience has been that, unlike the actual modeling program, thermodynamic databases are not well vetted, not maintained, not updated, and frequently modified by users without proper documentation. Merely citing a source such as Geochemists Workbench (Bethke, 1998) is not adequate. Without providing the database, or at a minimum a list of all relevant/critical species considered with their corresponding log K values, the results are not credible and very likely cannot be reproduced or meaningfully critiqued by someone else. Table C-4 is useful and should be expanded to incorporate the necessary data I mention. "Suppressed Minerals" probably requires explanation for those not familiar with Geochemist's Workbench.	to Reactions 1 through 4 of the manuscript. So by assuming all the Fe was Fe^{2+}, the estimated acidity for the mine release was maximized. • Documentation of the thermodynamic database was improved, including the database version as well as the thermodynamic data that was added to the database as well as the references for these data. The reasoning behind "Suppressed Minerals" was added.
	81) Reference is made to log Ks for calcite and dolomite (Parizek et al., 1971), which is old data and should be replaced by more recent citation (e.g., Nordstrom & Munoz, 1994). The signs for calcite and dolomite log Ks (App. 2 of App. C) are reversed and should be updated to +9.67 and +19.76, respectively for calcite and disordered	81) The inadvertent listing of the pK values as opposed to the log K values was corrected (the signs were corrected). The Parizek et al. data are sufficient for these purposes. There are still newer data than Nordstrom and Munoz as well that fall within the incremental difference of Parizek and Nordstrom.

Question 9: ***Were the geochemical principles to characterize transport and fate of acid mine drainage regarding neutralization, precipitation and mineral saturation appropriately applied and interpreted? Please explain.***		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	dolomite. I was pleased to see the updating of the conventional assumption for atmospheric log CO_2 fugacity to -3.4 from -3.5.	But again, minor refinements to thermo data, which will always be an ongoing process, offer negligible improvements to the calculations, given all the other uncertainties imposed on this effort by the unplanned-response circumstances.
Mark Williamson	82) The geochemical principles used in the study were very straightforward and standard. Calculations made to assess mineral saturation were helpful, but not surprising. The presentation read as calculations made to confirm the standard and expected. It is appropriate to make them for the sake of completeness. As noted above, it seems as though there would have been value in conducting a geochemical modeling simulation to mix upper Animas River water with the characterized discharge from Cement Creek. Such an exercise would essential provide expectations for the mixing phenomenon and potentially inform the characterization of Cement Creek as the calculations point to requirements for Cement Creek discharge, that unfortunately could not be measured in the heat of the moment following the GKM release, to account for observed effects in the Animas River. This follows from my perspective that very often the things one must do to acceptably model/represent field observations informs as to the particulars of the event.	82) Geochemical mixing models are depicted with Animas alkalinity for the estimated “Peak Concentration” (Fig. C-9), “Plume + Cement Creek Background” (Fig. C-10) and for “Cement Creek Background” (Fig. C-12).

Question 10: ***Were exposure analyses based on GKM concentration results appropriately applied and interpreted? Please explain.***		
Reviewer Name	**Reviewer Comment**	**EPA Response**
Charles Fitts	83) Not my area of expertise.	83) No response required.
Glenn Miller	84) The exposure analysis was done satisfactorily, and shows that the impacts were transient, and unlikely result in a non-trivial increase in exposure to humans, and to a significant impact on acute exposure to biota in the affected surface water. Using a variety of water criteria (aquatic, irritation, drinking water, etc.) the document showed that the standards were exceeded only in a transient manner, primarily in the Animas River. However, a comment is made in the document that the impact on reproductive success was not determined, and the only criteria that were used were acute toxicity. Even in this case most of the exceedances were less that the 96-hour toxicity assessments. Thus, with the exception of possible impacts on reproductive success, the comparisons of the criteria concentrations were fully applied appropriately.	84) The bioaccumulation of contaminants in fish modeling has not been included in the final report. The report relies on water quality criteria screening to identify potential exposure of human and aquatic life.
Ronald Schmiermund	85) No comment	85) No response required.
Mark Williamson	86) I do not consider myself particularly well qualified regarding exposure analyses. However, I feel that the considerable uncertainty in chemical constituent concentrations required for the analysis, due to modeling plume peaks and Cement Creek discharge needs to be discussed. Given the uncertainties, it seems that the exposure analyses may only be generally applicable. The BASS analysis may be the most applicable tool, but that does not mean it is suitable. Given the transient nature of	86) The fish bioaccumulation modeling has not been included in the final report. The report relies on water quality criteria screening to identify potential exposure of human and aquatic life.

Question 10: ***Were exposure analyses based on GKM concentration results appropriately applied and interpreted? Please explain.***		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	the GKM plume, I wonder how applicable results from a model like BASS that are (in my limited experience with exposure analyses) often dependent on reference data derived from long-term exposure.	

Question 11: ***Was the potential for groundwater uptake from the Gold King Mine appropriately applied and interpreted? Please explain.***		
Reviewer Name	**Reviewer Comment**	**EPA Response**
Charles Fitts	87) Chapter 9 and Appendix D could be trimmed substantially and they shouldn't have so much duplication. In the descriptions of the AEM, FDM, Gflow and Modflow, only the barest essentials need to be written and the reader can simply be referred to sources for more detail.	87) The groundwater assessment chapter was trimmed by removing the preview of the observed elevated metals at the 35m66km community well and improving the presentation with the other metals at the end of the chapter. Appendix D was tightened by focusing on the essentials of the various modeling approaches used in the analysis, and by removing the discussion of results leaving the full presentation of results for the chapter.
	88) The Appendix D presentation lacked a table listing the calibrated model properties in the GFlow models. For both the rock and the alluvium, list base and top elevations, recharge rate, Ks, porosity, etc. For the Hermosa models, list base elevations in the different alluvium domains and show a map-view of those domains. I think that the recharge rate was made the same in the rock and the alluvium, but that can't be gleaned from Appendix D.	88) Tables were added to the Appendix D presenting the fundamental model parameters for the regional GFLOW models of the lower Animas and Mid Animas River floodplains. These regional scale models were considered essential since they provided the boundary and initial conditions for the MODFLOW and AnAqSim local scale models.

Question 11: ***Was the potential for groundwater uptake from the Gold King Mine appropriately applied and interpreted? Please explain.***		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	89) A similar comment for the Modflow models: please add a table listing the Kh, Kv, thicknesses, etc. for the layers in the model, recharge rates, and details about how the wells were represented – what layers, etc.	89) The MODFLOW and AnAqSim models inherited the model parameters from the GFLOW model, and properties contained in the Tables of Appendix D.
	90) The rock heads in Fig. D-16 are as much as 600 ft lower than the rock heads in Fig. D-19. Certainly actual heads don't change that much in a few months. Since there is nothing to calibrate to out in the rock except one well quite close to the alluvium, it is distracting to extend the model out that far. It would be better to just do a local scale model of the alluvium near the critical wells, imposing heads and gradients from irrigation ditches and observed tributary connections or well water levels. The key questions revolve around flow patterns near wells located close to the river, and the answers shouldn't hinge on guesses about what is happening in rock miles away. I would just do a 3D model covering a small area (see excerpt of Fig. D-21 below), with a range of assumptions about alluvium Ks, pumping rates, etc. 	90) Calibration statistics for the model predicted hydraulic heads in the rock areas are no longer reported since these are far-field targets of minimal impact on the near-field. A local-scale MODFLOW model is presented to represent the full three-dimensional flow influences.

Question 11: ***Was the potential for groundwater uptake from the Gold King Mine appropriately applied and interpreted? Please explain.***		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	91) Since the irrigation ditches and river are head-specified boundaries, what happens beyond them has very little impact on the simulated well capture zone.	91) The irrigation ditches are known to be ephemeral and an important influence in the local domain. Local inhomogeneities such as buried river channels are also of importance in the local domain. Both geologic conceptual influences are explored with local scale models.
	92) The GFlow models presented use a single K value for the alluvium based on a very large-scale model calibration that assumes a uniform K in the entire alluvium. In reality, these are quite heterogeneous braided stream deposits, and it would be more informative to test a series of small-scale models at a couple of the wells of concern, using a range of K values common in these deposits to estimate the likely range of travel times from stream to well. That the Gflow and Modflow models of the mid Animas community well give similar travel times only confirms that both models used the same aquifer properties and imposed similar gradients and discharges. I would recommend just presenting 3D Modflow models at scales like the model shown in D-36 and D-37, using well and irrigation ditch water levels to constrain boundary heads, and vary alluvium properties and well discharges in reasonable ranges to give a range of travel time and capture zone results.	92) At the suggestion of the peer reviewer, we decided to refine the local scale models with explicit representation of aquifer heterogeneity rather than expand the range of the parameter values for the simplified averaged representation of hydraulic conductivity. Based on field observation, the macro-scale aquifer heterogeneity is dominated by braided steam channel deposits; the sediment type is predominantly gravel with an expected fairly consistent hydraulic conductivity. The influence of layering and silt is represented as anisotropy in the hydraulic conductivity parameter.
Glenn Miller	93) The question of groundwater uptake was an important issue, and one that was a real concern. However, the large portion of the drainage, particularly in the Animas River basin, is a gaining stretch, meaning that underground water does flow to the river, and would not allow delivery	93) The EPA/ORD team evaluated the situations where the Animas River might lose water and associated dissolved constituents to pumping wells in the flood plain aquifer. We examined the well water

Question 11: ***Was the potential for groundwater uptake from the Gold King Mine appropriately applied and interpreted? Please explain.***		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	of contaminated water to wells near the river. In certain instances, however, a large withdrawal of water could reverse this trend, where a localized cone of depression could pull water towards the well. This potential impact was addressed satisfactorily, and there was no data that conclusively showed an increase in contaminant load, but also could not completely exclude the possibility that some contaminant transport could have occurred. This issue was considered appropriately.	quality data for evidence of river plume-to-pumping-well communication.
Ronald Schmiermund	94) No comment	94) No response required.
Mark Williamson	95) The groundwater analysis contains much uncertainty due the overall lack of field characterization (as noted in the report). Pathways or barriers may easily be more site, and time dependent than can be established at the scale studied. Nonetheless less, the analysis is helpful to establish perspective, but may not be particularly definitive. The assessment seems reasonable for uptake from the GKM, at least for the basic, overall system. However, the geochemical constraints and challenges related to modeling trace element constituents can be expected to hamper the reliability of these model calculations. Sorption on sediments, potential redox and pH changes can all affect the actual chemical constituent, as distinct from particle tracking (conservative chemical movement) often used in groundwater studies.	95) The EPA/ORD team approaches the complexity of the geochemical processes involved in metals transport through heterogeneous saturated porous aquifers. The step-wise and progressive approach is proposed to start with attempts to understand water balance, then to build upon with conservative advective transport, and leave the evaluation of reactive trace element transport to potential future studies.

Part 3: Application of Software-Based Analytical Models

Question 12: ***Does the final report appropriately and adequately respond to the mid-project external peer review comments regarding the development and application of the WASP model? Please explain.***		
Reviewer Name	**Reviewer Comment**	**EPA Response**
Charles Fitts	96) I quickly scanned the mid-project report and think that most of the points raised have been addressed. However, with the clearer presentation of the WASP modeling methods and results in the present report, new issues around mass and dispersion have come to light as discussed in Charge Question 6.	96) No response required.
Glenn Miller	97) While I am not an expert on the WASP model, the results of the modeling effort appear to support the empirical results. Recognizing that it is much easier to make a model consistent with an actual spill, after it has occurred compared to when the modeling is done prior to the spill, the model, through my reading was helpful for explaining the time varying concentrations of metals observed in the sampling.	97) The EPA/ORD team agrees that WASP modeling assisted the analyses of the fate and transport of metals in the 550 kilometers of river affected by the spill, and was valuable in filling in gaps that the spatially limited Empirical Model could not address as well.
	98) I also believe that it is worth noting that the overall goal of this work was to understand how the spill affected the water quality in the receiving water, and to determine potential impacts immediately following the spill, as well as predicting of additional impacts would occur. In my opinion, the report has done this, and rather well.	98) No response needed; we appreciate the confirmation of our approach.
	99) The high degree of uncertainty that existed immediately after the spill has been largely continued. As discussed in the report, the amount of contaminant load from the mine water was a rather small contribution to the total load that was released to Cement Creek that made its way to the Animas River. The much larger portion of contaminant	99) No response needed; we appreciate the confirmation of our approach.

Question 12: ***Does the final report appropriately and adequately respond to the mid-project external peer review comments regarding the development and application of the WASP model? Please explain.***		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	load came from the result of the acidic mine water when it washed over the very reactive/oxidized rock immediately below the release point. It remains unclear of how much the acidity of the mine water affected the waste rock contribution. Would 3 million gallons of distilled water running over the same waste rock have resulted in a similar contaminant load? But there is no question that a very large amount of contaminants made the trip to the Animas River, and the WASP model, at least to this reviewer, rationalizes what happened to that contaminant load, and that is helpful for understanding what impact the spill has had. The use of the conductivity measurements, as suggested by the mid-project reviewers was a very useful contribution, since it generally pinpoints the plume dynamics, since it is not great leap of faith to assume that the high conductivity water should closely mimic the metals and particulate load.	
	100) Dr. Nordstrum suggested that reporting sulfate measurements would have been helpful, and I certainly agree. Other than mentioning it a few times, and indicating the total load in the release, I did not observe reports of sulfate concentrations in the report. Sulfate measurements can be highly useful, since it can be used for indicating dilution of fresher water. While not completely conserved due to gypsum precipitation and dissolution, at concentrations between <1000-1400 mg/L,	100) There was little sulfate data collected during the Gold King plume to assist in the way suggested by the reviewer. Some sulfate data has been collected post plume.

Question 12: ***Does the final report appropriately and adequately respond to the mid-project external peer review comments regarding the development and application of the WASP model? Please explain.***		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	it can be used as a tracer, if used with the proper constraints. While it may not be feasible to complete an analysis of the sulfate in the short time available, I looked for a discussion of sulfate, but did not see any.	
	101) Overall, however, given the constraints of sampling immediately after the spill, and not knowing exactly how the plume changed over time, I found the discussion and the conclusions very helpful, and feel that the response to the mid-review was adequate and improved the quality of the report.	101) No response required
Ronald Schmiermund	102) I note that the mid-project peer review included a three-day meeting of the peer review team and EPA scientists. This is presumed to have allowed a more detail and different type of review of the project than accorded this review. Only the comments of Dr. Nordstrom (the geochemist) will be reviewed here. Specific Comments: Dr. Nordstrom mid-project review: a) Question 1 - I did not find that the current presentation was structured in a way that felt natural to me (see my response to Charge Question 3) and would build confidence in the reader that they were following the study correctly. I spent a lot of time backtracking to understand the context.	102) See below. a) The order of presentation of material has been revised in the final report and more subsections have been added.

Question 12: ***Does the final report appropriately and adequately respond to the mid-project external peer review comments regarding the development and application of the WASP model? Please explain.***		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	b) Question 3 - I'm not sure I agree about merging the two sections, but I do feel that the relationships between empirical and WASP needs clarification (see my response to Charge Question 3). More importantly, I advocate more sub-sections.	b) The relationship between the empirical and WASP modeling has been clarified as part of the reorganization with additional discussion.
	c) Question 4 – I find no sensitivity analysis in the final report. Although I don't know what product was available to the mid-project review, it seems that the detailed analysis continues to be lacking or unclear in some areas. The treatment of individual metals may still not be what was requested by the mid-project reviewer.	c) Sensitivity analysis was done with WASP settling velocity as described in Appendix B. Individual metals have been presented throughout the report. However, there are many metals and complete treatment of all of them would be infeasible in the report. The project findings emphasize lead, copper, zinc, arsenic, and cadmium as suggested by the mid- project peer reviewers.
	d) Question 5 – I completely agree that the lack of direct data for the actual GKM effluent is a very significant deficit and that the methods used to estimate the GKM effluent quality are questionable in some respects and remain inadequately explained.	d) The explanation of the EPA/ORD approach to estimate the GKM effluent quality is thoroughly presented in Section 3.3.2. Although there is a lack of data for the mine effluent prior to release, we note that measured post-release mine adit effluent dissolved concentrations (Aug 7, Aug 11, Aug 15, Sept 21) are largely similar to immediate post release-data (with some exceptions). Turnover in the mine is understood to be rapid during the pre-release period and reasonably similar

Question 12: ***Does the final report appropriately and adequately respond to the mid-project external peer review comments regarding the development and application of the WASP model? Please explain.***		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	e) Question 6 – The deficits in the analytical data obviously must remain, but I don't see an effort to address them and exclude problematic data. Filtration procedures are now explained and the limitations acknowledged. However, the empirical and modeled estimations and conclusions do not appear to take into account coarse (0.45 µm) filtration. The lack of a summary table of analyses makes evaluation of the analytical data difficult. An accompanying CD with data presented in a consistent way would be valuable.	to what it was post-release. Therefore, we posit that our representative estimated mine adit concentrations can be expected to be similar to the actual. More uncertainty comes from the lack of data to quantify the plume peak in Cement Creek where there is no post-event data to assist reconstruction of this unique event. e) Quality assurance analysis has been added to the final report in Appendix F. Several checks on data such as ion balance, and duplicate comparisons suggest that data is sound within reasonable error bounds. The appendix includes an assessment of the accuracy and completeness of the data by three methods: i) cation/anion balance, ii) sum of analytes, TDS and specific conductance, and iii) dissolved/total major analytes. These evaluations do not fully resolve any issues with standard filtration issues and the empirical and WASP analyses take data at face value and do not attempt to guess or arbitrarily correct data.

Question 12: ***Does the final report appropriately and adequately respond to the mid-project external peer review comments regarding the development and application of the WASP model? Please explain.***		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	f) Question 7 – "Clay" only appears 3 times in the final draft, so I don't think this recommendation has been adequately addressed. g) Question 10 (comment 89) appears to have been addressed in Figs. 5-9 and 5-10, but the recommended additional work has not. A mixing/titration simulation in which pH and SI_{CAL} are calculated could be compared to observations.	f) The EPA/ORD team acknowledges that the erosion of fine clays might provide more surfaces for metal sorption and partitioning from dissolved to solid phase. Additional statements regarding the potential role of clays has been added to the discussion of the plume load in Cement Creek. g) Appendix C, the Geochemical Assessment includes both mixing/titration simulations (Figs C9-C12) as well as calculations evaluating saturation with calcite (Fig C8) and hydroxide minerals (Figs C13 and C14). These simulations and calculations using data agree quite well as described in Appendix C.
Mark Williamson	103) Although I am not familiar with WASP, it appears that the study made most reasonable attempts to address mid-project review comments. The one mid-project review comment regarding calibration seems to still require thought. The discrepancy between WASP and the empirical model does receive comment in the report (why the authors feel a difference exists) but as I noted above, using the empirical model (field mapping) to try to calibrate and reconcile seems to be a reasonable goal, unless there is some clear reason why that cannot happen.	103) We have carefully reviewed both the Empirical and WASP models for input and assumptions. As a result, we have revisited and refined the empirical estimates of mass and peak concentrations, and refined the settling velocities in the WASP model, resulting in improved agreement between the empirical and WASP model approaches.

Question 13: ***Does the final report appropriately and adequately respond to the mid-project external peer review comments regarding the development and application of groundwater modeling? Please explain.***		
Reviewer Name	**Reviewer Comment**	**EPA Response**
Charles Fitts	104) The current report now includes local-scale 3D Modflow models of two wells, and demonstrates that if the Modflow and Gflow models are similarly constrained, they will give similar estimates of capture zones and travel times. The mid-project review suggested running several model realizations to test reasonable ranges of input values. Some of that was done, but I think the K ranges tested were not as wide as they should have been, given the heterogeneous nature of alluvium deposits in braided stream environments.	104) Local scale simulation were added focused on the 75m71km community well using the AnAqSim model and various realizations of buried river channel deposits represented as high permeability zones. The decision was made to explore the local-scale realizations of aquifer heterogeneity rather than to expand the range of the averaged "lumped" regional-scale hydraulic conductivity (K) under the sensitivity analysis.
Glenn Miller	105) Yes, the mid-project review was reasonably critical of the hydrologic modeling effort, particularly related to horizontal versus vertical water movement constraints, and use of the models. I found the final report reasonable and helpful. The complexity of the hydrologic system with a large number of wells required a large amount of data that may or may not have been available. Coupled with the results of analytical results from the wells, there was not, at the least, large amounts of contamination from the spill. However, providing data on the conserved anions (including sulfate in this case) would have provided some additional data on whether migration from the river was observed. In general, however, the mid-project comments appear to have been taken seriously by the report authors, and the groundwater models modified to extract as much predictive information as possible.	105) An extensive data set was not found on sulfate water quality in the river before, during, and after the plume, nor in floodplain wells. Observations of a relatively conservative tracer would have been valuable in confirming the model results. The groundwater modeling used the empirical signal observed at the 35m66km community well based on zinc and other dissolved metals concentrations for our hypothesis testing.

Question 13: ***Does the final report appropriately and adequately respond to the mid-project external peer review comments regarding the development and application of groundwater modeling? Please explain.***		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	While the authors cannot exclude the potential that one of the municipal wells had drawn water from the Animas River, the analytical data indicating that even if it had, the zinc concentrations were sufficiently low (by an order of magnitude) that violations of the secondary standard for zinc had not be observed. Thus, with a reasonable certainty, the chances of the river being in direct communication with drinking water and other municipal wells appears to not occur, at the least, to a large extent.	
Ronald Schmiermund	106) No comment	106) No response required.
Mark Williamson	107) For the most part, comments seem to be addressed. However, the scale of the model domains, and the field data to support them produce uncertainty. The discussed issue of gaining versus losing reaches and the site specific temporal link to this makes the assessment generally uncertain, but helpful. Modelers can, and will discuss endlessly the subtleties of models. The present study seems to have responded to review comments satisfactorily to provide the initial assessment that it seems to be, pending more detailed and discreet assessment as need is identified.	107) The EPA/ORD team has presented a groundwater modeling approach that does not over-sell our ability to make predictions and assign numerical uncertainty to those predictions, especially given the lack of site specific field data. The approach was to bound the possible given the observations we did have.

Question 14: ***Does the final report appropriately and adequately respond to the mid-project external peer review comments regarding the development and application of bioaccumulation modeling? Please explain.***		
Reviewer Name	**Reviewer Comment**	**EPA Response**
Charles Fitts	108) Not my area of expertise.	108) No response required.
Glenn Miller	109) The mid-project reviewers spent a fair amount of time on this question, and indeed it appears that the authors of the document took these concerns to heart. The use of the lack of an observable fish kill was criticized as not being sufficiently conservative. However, I would tend to agree with the authors of the study that the transient nature of the exposure was unlikely to cause a major exposure of aquatic species, including the invertebrates. However, the draft report does examine the potential for bioaccumulation of several metals, and the treatment of this issue is thorough. One might even argue that the data were a bit over interpreted, since the exposure was transient and depuration of the metals was reasonably rapid.	109) The fish bioaccumulation modeling has not been included in the final report. The report relies on water quality criteria screening to identify potential exposure of human and aquatic life.
Ronald Schmiermund	110) No comment	110) No response required.
Mark Williamson	111) I am no bioaccumulation expert, and I sense there is much to debate and question. The report does seem to make an effort to satisfactorily respond to mid-project review.	111) The fish bioaccumulation modeling has not been included in the final report. The report relies on water quality criteria screening to identify potential exposure of human and aquatic life.

V. SPECIFIC OBSERVATIONS

Several peer reviewers noted that editorial quality of the document needed improvement and offered editorial advice. The final report has been extensively edited and individual suggestions may or may not have been used in that editorial review. Within the specific observation section, other content-related suggestions were provided. We have isolated those into a table and responded to them.

Specific Observations on Main Document, Figures and Tables				
Reviewer Name	**Page**	**Line**	**Comment or Question**	**EPA Response**
Charles Fitts	Table 2-8		Add a *river km* column and sort in ascending order of this.	River kilometer has been added to the table.
Charles Fitts	31	899-906	The average velocity over river km 0-12 doesn't need to equal average velocity at 12 km (channel shape and slope vary). The discussion of the comparison seems to assume they should be equal.	Stream velocity will vary through the 12.5 km reach. The statement is clarified to mean an average velocity for the distance.
Charles Fitts	Fig. 3-9		The last phrase *so turnover is fairly high (about 4 days)* probably warrants a bit more explanation and the text rather than the caption is the place to do it. I assume there were calculations of the mine pool volume and average residence time?	A brief discussion of turnover has been added to the text.
Charles Fitts	32	967	*was about 0.5 g mercury released* not *was no mercury released*. See Fig. 3-11	The statement about mercury has been clarified.
Charles Fitts	33	1002-1008	In this paragraph, add modifiers to make it clear that we are talking about colloid/particulate concentrations – just saying *metal(s) concentrations* leaves ambiguity. Also, it would strengthen the case for this approach to show how	The statement has been clarified.

Specific Observations on Main Document, Figures and Tables				
Reviewer Name	**Page**	**Line**	**Comment or Question**	**EPA Response**
			stable the ratio of Q/c_(colloid/particulate) was in the measurements from 16:00 on (see my comments in previous review). A graph of that ratio would be a helpful Figure.	
Charles Fitts	34	1041-1052	The discussion and equations explain how the dissolved estimates of Fig. 3-16B were arrived at, but do not explain how the total estimates of Fig. 3-16A were arrived at (I assume that was discussed in along with the *flow factor* in earlier paragraphs). Please clarify the origin of 3-16 A better.	Additional discussion of how total estimates were developed has been added to the final report.
Charles Fitts	38	1165-1170	There is no discussion of why the plume volume dropped from 3 million gallons at Cement Cr. to 1.2 million gallons at downstream stations. Did the missing water go into bank storage? This should be examined and discussed.	Additional discussion of hydrology and plume volume has been added to the final report in Chapter 4. The EPA team concluded that the loss was more related to measurement than to actual physical loss.
Charles Fitts	Fig. 4-4		The *normalized shape factor* is not discussed in the text or caption. Please explain it somewhere. It doesn't appear to be a best fit, as it overestimates dispersion.	Additional discussion of normalized shape factors, including how they were derived, how they relate to metals concentrations and how they were applied in the empirical modeling is included in the final report.
Charles Fitts	Fig. 4-5		*Average normalized shape factor* curve in legend, but not in graph. If in graph, it needs to be explained in caption and text.	Additional discussion of the normalized shape factor has been included in the final report.

Specific Observations on Main Document, Figures and Tables				
Reviewer Name	**Page**	**Line**	**Comment or Question**	**EPA Response**
Charles Fitts	40	1262	*45 minutes later than the nearest sample in time*? Please clarify	Additional discussion has been added on establishing plume peaks and the timing of sampling in the final report.
Charles Fitts	Fig. 4-7		Explain basis of 50% Figure in last sentence of caption – is it based on the ratio of discharges at Cement Cr confluence?	The Cement Creek concentrations were diluted by initial flow volume mixing for this analysis. This has been explained in report.
Charles Fitts	Fig. 4-12		Caption: last three sentences draw conclusions that plume may have been present, but not visible on the ascending and descending limbs of the plume. Given the significant excess dispersion (probably numerical) shown in the WASP model results compared to Empirical results in Fig. 4-13, the real explanation could be the WASP model predicting significant mass earlier than it should have due to numerical dispersion. The caption of Fig. 4-11 also indicates that the real leading edge of the plume was sharp, probably sharper than the WASP-simulated leading and trailing edges.	The early predictions of plume arrival in the figure are consistent with dispersion with the modeling predicting an early arrival time. Discussion of this is clarified in the final report.
Charles Fitts	44	1420-1421	*Fig. 4-13*, not *Fig. 4-12. Shaping factor* is not shown in most panels of Fig. 4-13, but empirical model colloid/particulate concentrations are, along with other data that helped guide the shape of the empirical model plume. For, D it would be better to give actual data with a right-hand scale	Additional discussion of normalized shape factors, including how they were derived, how they relate to metals concentrations and how they were applied in the empirical modeling is included in the final report.

Specific Observations on Main Document, Figures and Tables				
Reviewer Name	**Page**	**Line**	**Comment or Question**	**EPA Response**
			(conductance or whatever it is, rather than the undefined *Sonde shape factor*).	
Charles Fitts	45	1474-1475	*the Empirical Model centered at the suggested peak from GK WASP*. The empirical model did not to this – it just interpolated linearly between measured values. The *simulated sonde* shapes shown in Fig. 4-16 probably did use the WASP peak to position them in time. Drop *GK* from *GK WASP* to be consistent.	The empirical model has been modified since the review to simulate a plume using the shape factor centered at the WASP peak flow (rather than just interpolating between observations).
Charles Fitts	46	1517-1521	I find this paragraph confusing. Clarify what is meant by *for data providers* and the last two sentences.	Data providers is the general term we use to collectively refer to the agencies, tribes, and NGO's etc. that produced data that was accessed for the study. The term is introduced in the data methods.
Charles Fitts	Fig. 5-1		Make clear if these simulated concentrations are based on empirical or WASP model. Dark blue and yellow dots not shown in legend for part B. Why present the red line (San Juan = distilled water) since it is so unreasonable? Correct *XXXX* in caption.	The figures have been clarified that they were produced by the WASP model.
Charles Fitts	56 and Fig. 6-3	1913-1921	This discussion misses an important point. The higher total water column mass in the WASP model is because WASP is not transferring enough mass to the streambed. Both empirical and WASP models start with the same input	The EPA/ORD team appreciates the reviewers comments about use of the WASP model. We have carefully reviewed both the Empirical and WASP models for input and assumptions. As a result, we have revisited and refined the empirical estimates of mass and peak concentrations, and refined the settling

Specific Observations on Main Document, Figures and Tables				
Reviewer Name	**Page**	**Line**	**Comment or Question**	**EPA Response**
			mass, so the only way the WASP water column mass can be higher at downstream locations is if it transfers less to the streambed. That it has unrealistic high dispersion would only spread the mass out, it wouldn't change the mass. The empirical results are correct in that they match observations. The WASP results are incorrect in that they do not match observations. As I say in my answer to Charge Question 5, perhaps the WASP model should be dropped in total or at least for the Animas section. I don't see that it informs much, except possibly peak timing in the San Juan. The text says the truth is somewhere between the empirical and WASP results. I think a more accurate statement is that the truth lies close to the empirical results, which are well-calibrated to observations.	velocities in the WASP model, resulting in improved agreement between the empirical and WASP model approaches. As discussed by several reviewers, WASP characterizes the plume with a greater amount of dispersion, and therefore, higher concentrations for longer periods than the Empirical Model, thus generally estimating greater mass. We note that all water quality models like WASP incorporate dispersion as a typically observed process that affects the movement of materials in waterbodies. In the report we put greater emphasis on the rather unusual behavior of the plume in that it moved with a tight core of high concentration over a long distance with little apparent dispersion. In the final report we have improved integration of the modeling results and emphasized the data-based analyses. We continue to use the WASP model as it was particularly useful for movement and timing of the plume, especially in the San Juan River.
Charles Fitts	Fig. 6-19		Blue x and green triangle data not in charts, orange line is red, should say *total metals in sediment*, not *total particulate metals. Simulated concentrations are highest with a large settling of metals upon entering the Animas River at RK 12, but observed concentrations are highest in the mid-Animas, RK 60-110.* The caption keeps referring to *concentrations* where it should be referring to *WASP-simulated*	The figure has been clarified. The WASP model added local definition of deposit areas more clearly than the empirical model was able to do. Deposition of metals in sediments predicted by WASP, even in the relatively narrow zones such as in the vicinity of 120 to 130 km, was confirmed by field observations. This discussion has been added to the final report.

Specific Observations on Main Document, Figures and Tables				
Reviewer Name	**Page**	**Line**	**Comment or Question**	**EPA Response**
			concentrations. If it were up to me, I would not present these WASP-simulated sediment concentrations, but use empirical model results instead or just show measured concentrations. Need symbol in legend for post-peak measured sediment concentrations.	
Charles Fitts	61	2142-2143	*The metals in the plume and the background sediment mass were the same* – from the preceding sentences it seems the background mass is much higher than the plume mass – please clarify.	The background sediment mass was greater than the plume mass. This has been clarified in the final report.
Charles Fitts	Fig. 6-27		I don't see the point of a simulation that puts all deposited material into the water column as an initial condition, especially for low-flow conditions. Caption should end with *mg/L*.	The simulations of resuspension of deposited material have been replaced in favor of observed snowmelt-related data that has been acquired since the draft report.
Charles Fitts	100	3610-3613	This section should discuss the observed spike in zinc in relation to the simulation results. Why was the observed spike so much shorter – perhaps a higher K in reality at this location?	The section now includes a discussion of a number of additional factors that might account for an earlier arrival of dissolved zinc at the 35m66km community well, including: (1) three-dimensional streamlines; (2) aquifer anisotropy; (3) buried high-k river channels; and (4) pulsed pumping. More detailed modeling is included in Appendix D.
Charles Fitts	102	3685	Wouldn't it be more accurate to use the empirical model for plume timing, since is not confounded by numerical dispersion like WASP is and is just based on measurements?	The WASP model was used to interpolate plume characteristics between the point observations. This is clarified in the final report.

Specific Observations on Main Document, Figures and Tables				
Reviewer Name	**Page**	**Line**	**Comment or Question**	**EPA Response**
Charles Fitts	103	3704	*that might influence dissolved solute velocity and dispersion.* In a braided stream channel environment like this, there are several orders of magnitude variation in K between the most permeable channel gravels and the least permeable abandoned channel silts. It is quite possible somewhere between the river and well there is K that is an order of magnitude larger than the modeled K, and an arrival time of 8 days could easily occur. These modeled breakthough times are very crude estimates, since there is no K data for the well/river vicinity. I think odds are high that the 8/14 anomalies are due to the GKM plume, since that is the most likely cause. The discussion should be expanded to include these points.	We fully explored the influence of buried stream channel heterogeneity in AnAqSim local scale simulations detailed in Appendix D. The influence of preferential flow pathways on water supply wells are discussed in Chapter 8.
Charles Fitts	104	3756-3757	I would de-emphasize the point about the arrival time not matching modeled breakthrough. As noted above, travel times could easily be much shorter than simulated, given the heterogeneity of such an environment.	We de-emphasized the fact that the model predicted later solute breakthrough times at the 35m66km community well. As described above, we discuss a number of additional factors that might account for an earlier arrival of dissolved zinc at the 35m66km community well, including: (1) three-dimensional streamlines; (2) aquifer anisotropy; (3) buried high-k river channels; and (4) pulsed pumping. More detailed modeling is included in Appendix D.
Charles Fitts	107	3862-3865	This discussion neglects the empirical derived deposition in the canyon reach (see Fig. 6-9a), which is more data-	Additional discussion has been added to summary chapter on the importance of deposition in the canyon reach.

Specific Observations on Main Document, Figures and Tables				
Reviewer Name	**Page**	**Line**	**Comment or Question**	**EPA Response**
			based than the WASP deposition pattern.	
Glenn Miller	9	8	Change "Only 2800 kg of metals.. " to "Of the 490,000 kg of metals released to the Animas River from the spill, only 2,800 kg actually came from the mine water; the rest came from the water washing waste rock located immediately outside the mine" Rational: the term "only 2,800 kg of metals tends to minimize the issue (although it is correct, a new reader will think that the Document is trying to minimize the impact)	The intent is not to minimize the impact. It is clear however, that far more material came from outside the mine.
Glenn Miller	9	16	The method of calculating the weight of the metals should be defined a bit. Does this include only the metals? Does it include sulfuric acid?	Mass of the metals and the sulfates and other constituents measured at the mine are provided. Sulfuric acid was not measured and is not explicitly estimated.
Glenn Miller	9	17	I suggest starting this sentence out with a brief description of the decades of release of contaminants from historic mining. Indeed, the Gold King release is small relative to even a month of normal release, and that point is very important. However, as is the case above, the writing should not be seen as minimizing the effect of the spill.	We acknowledge the reviewer's observation, but believe the paragraph emphasizes the uniqueness and importance of the event.
Glenn Miller		1088	The term "mine waste" is correct. "ore" is an economic material, and since it was deposited outside the mine, it is waste.	The term "mine waste" is used to refer to the waste pile in the final report, not "ore".

Specific Observations on Main Document, Figures and Tables				
Reviewer Name	**Page**	**Line**	**Comment or Question**	**EPA Response**
Glenn Miller	49	1618	This should be “neutralizing”, not oxidizing	These clarifications have been added to the final report
Glenn Miller	Fig. 3-11		The title should be “major anions and cations” It can show that sulfate is the major anion, but the legend is unclear as written. For “C”, where is the “major metals” figure?	These clarifications have been added to the final report.
Glenn Miller	Fig. 4-4		Total metals. The metals included should be spelled out. Total metals minus cations is not clear, since many of the major metals are cations.	The final report defines what is meant by the term in this study, which includes calcium, potassium, sodium, and magnesium. These points are clarified in the text of the final report.
Ronald Schmiermund	General		Given the importance of “dissolved” versus “colloidal/particulate”, the word “metal(s)” should always be preceded by a modifier for clarity	Care is taken in the final report to make this distinction.
Ronald Schmiermund	14	246-247	Mine-waste rock has not been pulverized to remove sulfides – only applies to tailings	Distinction has been noted.
Ronald Schmiermund	15	273-283	Please clarify the differences (composition, stability and distribution) between “waste rock/dumps” and “tailings/tailings ponds or piles”	The final report refers generally to mine waste and the mine waste pile at the Gold King Mine. This has been clarified in the final report.
Ronald Schmiermund	21-22	Field & Lab Methods	These sections should make clear the extent to which the field and laboratory methods described were followed by each of the various collecting entities. Differences are alluded to on p. 23 and perhaps should be summarized in a	It was not possible to objectively determine the extent to which field and laboratory methods were following. Commercial testing laboratories carried out most analyses but field sampling would be at the discretion of the respective organizations. The project did perform some quality assurance data comparisons to the extent possible with the data. The final report

Specific Observations on Main Document, Figures and Tables				
Reviewer Name	**Page**	**Line**	**Comment or Question**	**EPA Response**
			table. There should be a reference to the SOPs for each entity.	contains links to quality assurance documentation of the respective organizations. Appendix F addresses quality assurance.
Ronald Schmiermund	22	507	This section and Table 2-4 contain a common but important omission that would call into question all sediment data unless resolved. No where do I find a specification of the sediment digestion method. Fortunately, Appendix A-8b does specify EPA Method 3050B as the digestion method used by EPA Regions 6 and 8. However, it should be included on p. 22 and in Table 2-4 in addition to the characteristics of that digestion (i.e., briefly describe as a 'partial' digestion and list the components of the sediment likely to be addressed and not addressed by the method and their respective relevance to this study.) The inconsistency in digestion methods, even among EPA regions, revealed in Table A-8b is potentially problematic. This demands a detailed explanation and a caveat of the data that was not obtained via the method chosen as the 'main' or 'preferred' data set for sediments (presumably those using 3050B).	Sediment digestion methods have been clarified in the final report. The data providers used similar methods, varying in minor ways depending on the lab analysis equipment used. EPA considers the data comparable.
Ronald Schmiermund	27	708-714	I approve of the acknowledgement that 0.45 µm is neither natural nor effective, but I think, having made the	Additional discussion of filtering consequences was added to the text.

Specific Observations on Main Document, Figures and Tables				
Reviewer Name	**Page**	**Line**	**Comment or Question**	**EPA Response**
			acknowledgement, that a reason for using that convention should be offered and an explanation of the consequences provided.	
Ronald Schmiermund	27	723	"Acidity" is misused here and elsewhere. In fact, "acidity" was not measured for this study and should be eliminated. The sentence should read "… metals released from and the low pH conditions resulting from the Gold King Mine blowout …". On page 32 *calculated* acidity is mentioned – are calculated acidities being referred to here?	Care is taken to the final report to apply the term "acidity" appropriately.
Ronald Schmiermund	27	724-725	The sentence implies that the only mechanism for subsequent metal mobilization is re-entrainment of settled solids and does not acknowledge desorption.	It has been clarified in the final report that the WASP model addresses mobilization by re-entrainment.
Ronald Schmiermund	27	730	For clarity modify the sentence "….. throughout the analysis: one based on contaminant concentration and one based on contaminant mass."	The distinction between concentration and mass is addressed in the final report.
Ronald Schmiermund	27	737	Should read "… concentration (expressed as mass of contaminant per unit volume of water or unit mass of sediment) …"	This has been edited in the final report.
Ronald Schmiermund	27	739-752	These paragraphs are difficult to follow and interpret, and might be taken by the public to be obfuscation. I personally	The document has been edited for clarity.

Specific Observations on Main Document, Figures and Tables				
Reviewer Name	**Page**	**Line**	**Comment or Question**	**EPA Response**
			find this sort of thing difficult to explain and don't presume to reword it. However, I encourage rethinking and restating the material. Line 746: which chapter is "this" chapter? Line 749: "Here we provide…" Where is "here"?	
Ronald Schmiermund	32	984	The near-surface mine waste being referred to here was likely not effected by "ore processing" as the ore was conveyed by tram line to the Gold King mill at Gladstone.	Some historic photos of the site show a large mine processing facility at the site at one time.
Ronald Schmiermund	Fig. 4-3B		This plot is incongruous. It implies that in the 3.8 km between the Cement Ck and Animas R gage 1.7 million gallons was lost to evaporation or some other withdrawal. Please clarify.	Discrepancies in flow volume and discussion of USGS flow records at the Animas below Silverton gage are discussed in the final report.
Ronald Schmiermund	Fig. 4-4		Be careful with the word "total". Does this imply (dissolved + colloidal/particulate) or something else. Does "total metals Less Cations" mean TDS less anions?	Care is taken in the final report to ensure clarity in summed totals and dissolved, and colloidal fractions.
Ronald Schmiermund	Fig. 4-5B & C		These plots are labeled "Plume Shape Factor" but appear to plot normalized peak height. Colors in B are different from those in A. No units appear in B & C	The figure has been edited.
Ronald Schmiermund	39	1198-1200	The conductivity as measured by the sondes does not necessarily confirm that metal concentrations were behaving	A comparison of sonde specific conductance to metals concentrations demonstrated that they behave

Specific Observations on Main Document, Figures and Tables				
Reviewer Name	**Page**	**Line**	**Comment or Question**	**EPA Response**
			consistently between sampling points, only that the combined effect of sulfate and other major ion concentrations behaved consistently. Using Fig. 4-4 to justify the coincidence of metals and conductivity is inconclusive since conductivity is not plotted.	consistently with the parameters measured onboard the sondes.
Ronald Schmiermund	Fig. 4-9		Identify the illustration as a "segment". There should be an analogous illustration for dissolved metals. A table identifying the required variables to solve the equations for the continuous batch reactor would be informative.	The clarification has been added to the figure in the final report. More detail on the WASP model is provided in Appendix B. A table is presented for the variable used and the sources for the necessary data. The appendix discusses how each process is modeled, the parameters used, and how these parameters are derived. In the appendix, the parameters per segment are presented.
Ronald Schmiermund	41	1309	Site 09358550 is 0.72+ miles upstream of the confluence – hardly "just upstream"	Distance noted in text.
Ronald Schmiermund	Fig. C-9		I'm quite familiar with Geochemist's Workbench, but I can't follow the figure caption	The caption was edited for clarity.
Ronald Schmiermund	Fig. C-10		Explain differences between plots	An explanation of the differences between the plots has been added.
Ronald Schmiermund	Figs. C-9, C-10, C-12		There should be a reference to the thermodynamic database used and a list of log Ks for important solids plus a list of all relevant species considered. Plots of precipitated masses are easier to	We referenced the thermodynamic database and added important log K values as summarized in Table C4. We list masses of minerals in Figure C11. Concentrations are depicted on the y axis. We judge that adding concentrations will over-complicate already complicated figures.

Specific Observations on Main Document, Figures and Tables				
Reviewer Name	**Page**	**Line**	**Comment or Question**	**EPA Response**
			interpret if paired with a plot of important aqueous concentrations.	
Ronald Schmiermund	Fig. C-13		Calculations related to aluminum phases are questionable given the 0.45 µm filtration. Nordstrom & Ball (1986), Nordstrom & May (1996). This may also apply to a lesser extent to Fe.	The limitations are now noted in the text.
Ronald Schmiermund	106	3840-3844	The terms “dilution of the flow” and “original strength” of the flow are ambiguous and should be clarified as they compromise understanding of the conclusion	The terms have been clarified in the final report.
Ronald Schmiermund	107	3853	‘acidity’ is an intensive, not an extensive quantity, quantity (no volume association)	The final report has been corrected.
Ronald Schmiermund	40	1244-1245	Explain “basin-scale relationship”	The term has been explained in the final report.
Mark Williamson			I would note that there are numerous editorial errors, blunders and omissions in the body text of this report. I cannot possibly capture them all. It is presumed that future editing by the report authors will capture and correct these.	The final report has been extensively edited.

Specific Observations on Appendices					
Reviewer Name	**Appendix**	**Page**	**Line**	**Comment or Question**	
Charles Fitts	D	4	81-90	The analysis of % contributed by groundwater flow is highly dependent on the length of the river reach between upstream and downstream gages, and the 10% value or the 21% values are not generally applicable. For example, if the reach between gages was longer or the location further up the drainage system, you would get a higher percentage.	The final report removes references to 10% of streamflow.
Charles Fitts	D	11		No need for Figure 8 – MODFLOW references are enough.	Figure D-8 is included given the educational objective of the appendix.
Charles Fitts	D	22	472-473	Domestic wells generally return most of their flow via a septic system, so the net is near zero. This is not true if a significant portion of water is used for irrigation, where water transfers to the atmosphere. You probably should reduce the simulated discharges of domestic wells.	We agree. Overestimate of well pumping will be conservative in overestimating potential for impact (more protective). This has been clarified in the final report.
Charles Fitts	D	Fig. D-20		Figure doesn't include lateral flows from rock into alluvium or irrigation flows lost to ET.	Explanation has been added to Figure caption in final report.
Charles Fitts	D	28	521-529	The water balance discussion is confusing. See point about Fig. D-20 above. Some of the irrigation diversion water returns to the river, but I think it is all assumed to exit to the atmosphere? See point above about domestic wells and septic systems. Please clarify the discussion.	This has been clarified in the final report.
Charles Fitts	D	31	558-559	Explain what you mean by *sanitation wells*.	The wells of the sanitation department were not used in the groundwater analysis. This has been clarified in the final report.

Specific Observations on Appendices					
Reviewer Name	**Appendix**	**Page**	**Line**	**Comment or Question**	
Charles Fitts	D	44	726	Numerical dispersion is an issue with solute transport, but not with particle-tracking. The earlier breakthough time is due to the 3D representation of the well – the 3D model has lower head at the well than the 2D model for the same discharge, so creates steeper horizontal gradients between the top of the well and the river. Shallow 3D pathlines have to endure less vertical resistance than deep 3D pathlines so they get farther.	The issue of 3D tracelines is added to the MODFLOW/MODPATH discussion.
Charles Fitts	D	47	774-777	The trials could have used a broader range of K (an order of magnitude), since the range in the deposits is much greater than the range in interpretations from one well's pumping tests. For example, there could be a cobble/gravel layer above the well screen between the river and well (not influencing a test much, but greatly influencing travel time).	The EPA/ORD team decided to broaden the representation of hydraulic conductivity by putting in explicitly the presence of buried subsurface river channel deposits (high K) rather than broadening the range of K in the sensitivity analysis.
Charles Fitts	D	49	828	The conclusion should add that at no well other than the RK66 one, were anomalous metals concentrations detected in the time soon after the release. That is a big take-home message.	We moved the groundwater conclusions to Chapter 8, and pointed out that only one of the several community wells was found to have an anomalous dissolved metals concentration soon after the release, but that the concentrations were well below drinking water action levels.

EPA
United States
Environmental Protection
Agency

PRESORTED STANDARD
POSTAGE & FEES PAID
EPA
PERMIT NO.G-35

Office of Research and Development (8101R)
Washington, DC 20460

Official Business
Penalty for Private Use
$300