Peer Review Summary Report

External Letter Peer Review of U.S. EPA's Report "Analysis of the Transport and Fate of Metals Released From the Gold King Mine Into the Animas and San Juan Rivers"

Peer Reviewers:
Charles Fitts, Ph.D.
Glenn C. Miller, Ph.D.
Ronald L. Schmiermund, Ph.D.
Mark A. Williamson, Ph.D.

Contract No. EP-C-12-045
Task Order 99

Prepared for:
U.S. Environmental Protection Agency
National Exposure Research Laboratory
109 TW Alexander Drive
Research Triangle Park, NC 27711

Prepared by:
Versar, Inc.
6850 Versar Center
Springfield, VA 22151

October 24, 2016

TABLE OF CONTENTS

I. INTRODUCTION

Versar, Inc. (Versar), an independent EPA contractor, coordinated an external letter peer review of the EPA Report "Analysis of the Fate and Transport of Metals Released from the Gold King Mine in the Anima and San Juan Rivers," which was prepared by EPA's Office of Research and Development (ORD), National Exposure Research Laboratory (NERL). The document is designated as Influential Scientific Information (ISI). For this peer review, four expert reviewers were selected to independently answer fourteen charge questions pertaining to hydrology, geochemistry, fate and transport, and potential impacts from the Gold King Mine release. The charge included questions on EPA's response to comments from a mid-project review of the project and conclusions, which was also contractor managed and held on February 23-25, 2016. This peer review summary report provides a summary of the external letter peer review comments and presents the reviewers' individual written comments.

I.1 Background on Gold King Mine Analysis of Fate and Transport

On August 5, 2015, EPA was conducting an investigation of the Gold King Mine (GKM) near Silverton, Colorado, to assess the on-going water release of acid mine drainage (AMD) from the mine, and to assess the feasibility of further mine remediation. While excavating near the mine entrance, pressurized water began leaking above the mine tunnel, eventually spilling about three million gallons of water stored behind the collapsed material into Cement Creek, a tributary of the Animas River.

Since that time, personnel from all parts of EPA have been assisting in response efforts. A portion of the response included ORD research to:

- Understand the geochemical drivers that mitigate spill effects within the rivers receiving the AMD;
- Characterize the GKM acid mine drainage (AMD) spill;
- Characterize transport and fate of AMD in Animas and San Juan Rivers, and;
- Estimate possible future water quality and biological impacts.

A team of ORD scientists with expertise in geochemistry, surface and groundwater hydrology, environmental engineering, water quality modeling, fish biology and bioaccumulation, statistics, and geographical information tools used the following EPA models and GIS tools to analyze the sampling data:

- Water Quality Analysis Simulation Program (WASP) to analyze the transport of metals through rivers;
- Bioaccumulation and Aquatic System Simulator (BASS) to determine the uptake of metals in fish during plume passage;
- Wellhead Analytic Element Mode (WhAEM) to look at groundwater transport and connection of wells to the river;
- EnviroAtlas for data gathering and geospatial analysis.

This project's objectives were to provide analysis of water quality following the release of acid mine drainage in the Animas and San Juan Rivers in a timely manner 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 during the plume and in the first month following the event. A quality assurance project plan was developed for the work in this project.

EPA sought an interim peer review of the project and conclusions through a panel peer review on February 23-25, 2016. The peer review was contractor managed and a final peer review report was delivered to EPA with reviewers' comments and suggestions for improvement. EPA has now developed a draft report, "Analysis of the Fate and Transport of Metals Released from the Gold King Mine in the Animas and San Juan Rivers" for this project to undergo peer review.

I.2 Peer Review Process

Versar was tasked by EPA with selecting and securing four scientific experts to conduct an external letter peer review of EPA's Report "Analysis of the Fate and Transport of Metals Released from the Gold King Mine in the Anima and San Juan Rivers." The peer review process provided a documented, independent, and critical review of the analysis. In recruiting these peer reviewers and coordinating the peer review, Versar was charged with evaluating the qualifications of peer review candidates, conducting a thorough conflict of interest (COI) screening process, independently selecting the four peer reviewers, distributing review materials, managing the written peer review period, and developing a final peer review report.

The peer review selection process was initiated by Versar to identify candidate reviewers with expertise in the following areas: (1) geochemistry, (2) fate and transport of metals – in water column, (3) fate and transport of metals – in sediment, (4) groundwater modeling, (5) familiarity with Water Quality Analysis Simulation (WASP) model, and (6) bioaccumulation. Versar's in-depth and multi-staged evaluation of qualifications was based on each candidate's curriculum vitae (CV), publications, professional accomplishments, and membership in professional societies. In total, Versar identified and contacted 19 candidate reviewers to determine their interest and availability to participate in this peer review.

In addition to the evaluation of candidates' expertise, Versar conducted a thorough conflict of interest (COI) screening of the candidate reviewers. Each candidate reviewer was required to complete a series of screening questions to help determine if they were involved with any work and/or organizations that might create a real or perceived COI. Following this screening process, a pool of seven peer reviewers were submitted to EPA. Because the review is classified as an ISI, Versar took the additional step, in accordance with the 4th Edition Peer Review Handbook, to hold a meeting with the EPA Contracts Office to discuss any actual or potential COI for the seven reviewers. Versar independently selected four reviewers to participate in the peer review and EPA provided consent on the four selected reviewers. The list of the four peer reviewers who participated in this peer review is provided below.

Following the selection process, Versar distributed to the reviewers the EPA report "Analysis of the Fate and Transport of Metals Released from the Gold King Mine in the Animas and San Juan Rivers," along with the fourteen charge questions (See Section II). Throughout the peer review period, Versar monitored the progress of the peer reviewers and responded to any reviewer questions. Versar received two technical questions which required a response from the EPA authors of the document. These questions and answers are provided in Section I.3 for transparency of the peer review process.

Peer Reviewers

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

I.3 Peer Review Monitoring

During the review, one reviewer posed two technical questions to Versar which required detailed responses from EPA. The exchange of information between the reviewer and EPA was facilitated by Versar, with no direct communication between the two parties. The questions and responses were distributed to all reviewers to consider in preparing their written comments. The reviewer questions and EPA responses are provided below.

Reviewer Question 1
In reference to Figure 4-3B in the report:
1) Why the difference between the two plume volumes?
2) Why was only the larger plume volume used when the slug volume remained nearly constant at a smaller volume between Silverton and Durango?

<u>EPA Response to Reviewer Question 1</u>

The plume volume of released water was determined by visible differences in water level at the stream gage in Cement Creek only. We were able to see it clearly initially in Cement Creek because of the small volume of flow in the stream (relatively). The peak of the release generated a wave of water within the Animas River that could be seen at USGS gages and this peak in water level remained visible as elevation of water level for some distance. However, the plume metals entered the river over a longer time than the peak of course (8 hours) and after the peak and later water did not influence the hydrograph of the Animas much. When we refer to the 1.2 million gallons, we're talking about this "core" at the peak that traveled rather coherently. Consider, in just a one hour period, the river at Silverton would have carried 6,700,000 gallons of water. The entire hydrologic signal was damped by incoming flow as the plume moved downriver, making this additional water harder and harder to see as a wave (but not as a plume of concentrated metals). We just were not able to separate the remainder of the plume volume at USGS gages in the Animas River within the accuracy of water level measurement devices at USGS gages. In addition, both the wave of water and the plume of metals mixed with ambient flow and were dispersed by drag forces, as clearly evident in concentrations measured after the peak of the plume traveled. We do not mean to imply that all 3 million gallons of water with its full mixture of metals did not migrate downstream. We will clarify that point in the final draft.

1) There are not 2 plume volumes. The plume volume was 3,000,000 gallons. The mass of metals was 490,000 kg. We discuss a "core" within the plume that coincided with observable changes in water level at USGS gages downstream for some distance that amounted to 1.2 million gallons of the total release and plume volume. This core does not represent the entire plume but it did coincide with peak concentrations. The remainder of the plume did not "disappear". It mixed with the river volume at levels that were within the detection limit of the measurement devices at USGS gages.

2) The plume is defined by the elevated metals concentrations—not the volume of water. Elevated metals concentrations persisted for a much longer time than measurable changes in water level at gaging stations.

All modeling uses the volume of 3,000,000 gallons and mass of 490,000 kg with defined concentrations as a basis. The 1.2 million gallon core is simply interesting.

Reviewer Question 2
My concern relates to apparent differences in volumes of water released by the GKM blowout (i.e, the wave, not the "plume"). Using USGS gage data I calculate approximately 3 million cumulative gallons above base flow over the course of the passing wave (12.45 through 20:00 on 8/5/16[1]) at the Cement Ck. station (09358550). Using the same approach, taking into account the falling base flow in the Animas, I calculate 1.24 million gallons above base flow over the same time period downstream at station 09359020. Conservation of mass tells me this can't be right.

Unfortunately, I cannot validate the discharge at 09359020 by summing all the contributions because we don't have any discharge data for Mineral Ck. (at least that I'm aware of).

Since your group has obviously worked extensively with this data, please explain what I seem to be missing.

[1] It is assumed that this reviewer means 8/5/15 instead of 8/5/16.

EPA Response to Reviewer Question 2

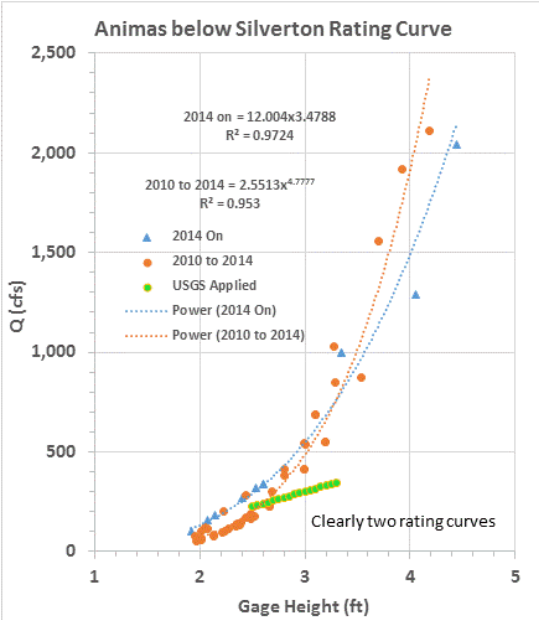
As the reviewer has noted, the plume clearly measures at about 3,000,000 gallons in Cement Creek. Once the plume enters the Animas River, the amount of water that can be clearly accounted for in the USGS gage records below Silverton is about 1,250,000 (depending on interpreting baseflow, etc). Approximately that same amount can be interpreted at the series of USGS gages down the Animas River. We describe that as the "core" of the plume where the largest concentrations of metals were also observed. The elevations in the hydrograph were small at downstream locations and lasted about 2 to 3 hours. But they are visible and when the volume above baseline is summed it adds up to the general figure of lets round to 1,500,000 gallons. Elevated plume metals concentrations lasted much longer than that. The reviewer wants to know where the other half of the plume volume went, especially in the Silverton area that was so close to Cement Creek.

We do not know where it went, although we also assume that conservation of mass would say it was somewhere. While we observed the phenomenon we did not worry about it a great deal in the report, as we were focused on the plume metals concentrations and not the water volume. We can offer some hydrologic observations and welcome any comments by the reviewer on this subject.

The release hydrograph in Cement Creek lasted for approximately 5 hours (although we could make a case it wasn't thoroughly complete for more like 20 hours.) 1.5 million gallons passed in about the first hour. The first 2 hours of flow at Cement Creek carried about 50% of the load and 77% of the water.

During this time, Cement Creek was about 48% of the flow in the Animas falling to 23% at the end of the hour, then declined to 15% in the next hour. The natural background rate is 11%. It is likely that these levels may not be obvious in the hydrograph in the Animas below Silverton (especially given the oscillations observed in the 15-minute records). The hydrographs at downstream gages could be interpreted as disturbed for about 2 hours, carrying about the same volume of flow. There are uncertainties in this as well in that reasonable estimates of travel time of the plume don't always match well with the apparent change in hydrographs.

Carrying on with the subject of uncertainty in the USGS gage records and coming back to the Animas River at Silverton. We preface this by saying it is always hard to understand everything about a reported gage discharge so we don't want to over-interpret and we are not in a position to correct any USGS gage records. However, let us offer this. Again, the following analysis is based on a great deal of experience with measurement of streamflow at gages like this one, including relatively small and steep streams and rivers.

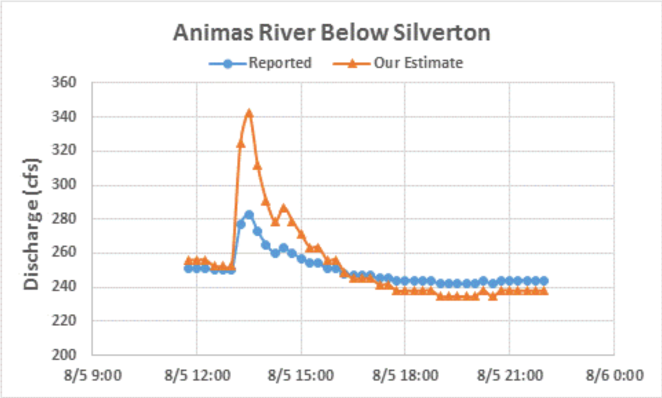


If we were to independently build the gage record ourselves from the gage records available, we would go to the published cross-section info for the gage and create a rating curve (Q=f(height or depth measured from the staff plate). We would go to the published flow record for the gage for the time period of interest and obtain the gage height record, along with the flow record. Our primary interest here is the height record. It is likely the depth from the rating curve and the height from the gage record may not agree, so they can be normalized with an offset between the two. (Height of the instrument can change with its position—therefore there is usually an instrument offset). One can then take the adjusted gage height and apply the rating curve and calculate q for each 15-mnute period.

The gaging record is shown above. The measurements are generally rated fair to poor by the observer. Within the record there is clearly an adjustment in the rating curve that occurs roughly in 2014. Normally, a new curve would be started when this is noted. A new gaging relationship should be applied until the next channel adjustment makes it no longer applicable. We will calculate the curve combined and separately as we don't know how the gage manager has elected to deal with this.

We obtained the gage height records early in our analysis. Actual recorded gage height ranged from 2.67 to 2.89 at peak during the first hour of the plume. That record is no longer available on the USGS website. Therefore, there could be some small error due to final versus provisional data.

We have also taken the reported provisional gage height and Q applied by the USGS in this narrow range of time at the gage and plotted it with the gage q to height relationships. One can see on the chart that the USGS discharge estimate in relation to the gage height is quite different than would be constructed from the more recent data. When examined closely, it doesn't match the older data that well or the combined set of data which should be used at a minimum.



Trusting our ability to create our own estimate of Q during the period of interest from our rating curves we obtain a much different estimate of Q during passage of the plume (see left). We prefer to use the most recent gage rating (2014 on in the figure above) given a clear shift in the curve, and we apply the offset described above to adjust to the gage height record to produce an alternative view of the Q record.

Interestingly, if one takes this estimated plume and subtracts out the baseflow, one accounts for 3,043,000 gallons of water from Aug 5 13:15 to Aug 5 18:00 (subject to assumptions about baseflow during the falling hydrograph).

Perhaps the volume was not actually lost in such a short distance. But the flow data is not ours and we are not privy to the processing that went on to produce the official record. The reviewers' thoughts are appreciated.

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 Soft-ware 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. SUMMARY OF PEER REVIEWER COMMENTS

III.1 General Impressions

The reviewers generally note that the report represents a considerable effort and contribution to understanding the Gold King Mine release. While acknowledging the report is in draft form, the reviewers also note that the report contains numerous editorial issues, including misspelling and incomplete sentences, which affect the reader's ability to understand the points the study authors are trying to make. Multiple reviewers suggested that the report should make more of an effort to identify, quantify, and qualify error. Numerous specific comments were made on the accuracy, clarity, and soundness of conclusions, which are presented in the responses to charge questions below.

III.2 Responses to Charge Questions

Part 1. Overall Project and Analysis

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

The reviewers all stated that the objectives of the study were clearly identified and addressed. However, one reviewer noted that the reasons for quantifying the release and fate and transport should be explained; in addition, the reviewer stated that the analyses are not always clearly or extensively presented. Another reviewer commented that the study itself is limited due to the lack of objective-critical data; specifically, the characterization of the discharge itself, 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. The reviewer also indicated that many data required filling through estimation methods and assumptions, which affects the error associated with conclusions reached in the study.

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

In general, the reviewers thought that the assumptions about data inclusion and use were appropriate. Some specific issues were, however, noted. In particular, one reviewer commented on three assumptions made to reconstruct the plume within Cement Creek. The reviewer stated that the report needs to include a discussion about the reconciliation and adjustments made to the flow data from one USGS gage (09359020) which was suspect. Also, justifications should be provided for assumptions on time-invariant mine discharge quality related to reconstructing the dissolved metal GKM plume. Finally, the reviewer noted that reconstruction of the suspended metal plume involves a different assumption (and model) relative to the dissolved metal plume. Two reviewers commented that data quality issues, such as dissolved metals greater than total metals, should be examined (if not already) and discussed in the report.

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

All reviewers thought the results were meaningful. With respect to scientifically defensible conclusions, one reviewer stated the analysis is scientifically defensible, while another suggested that a fuller description and discussion of errors and their impact on finding might prove helpful. Two reviewers discussed the WASP modeling, stating that it was either not scientifically defensible (as discussed in Charge Question 5), or the reason for invoking a particularly complex model such as WASP needs to be made more clear. One reviewer provided suggestions for clarifying two of the conclusions in Chapter 10, including providing more information on the release characterization (conclusion 1) and adding field observations to augment the conclusion on metals released from Gold King Mine (conclusion 2). This reviewer also suggested presenting empirically modeled peak compositions (with respect to comment 89 and 91 of the mid-project review), and putting various conclusions related to increases relative to background or ambient conditions into context with some statistics (e.g., x% greater than the background mean).

Part 2. Fate and Transport

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

All reviewers commented on constraints from the lack of data. Given these constraints, one reviewer noted that 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. The reviewer further commented that 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. Another reviewer noted that the uncertainties on the estimates made should be better represented in report discussion, and suggests propagating a maximum and minimum source (Cement Creek discharge) through the subsequent downstream assessment to bind the conclusions. One reviewer also suggested categorizing some characteristics as "inferred" (Level 7 portal effluent and the derived slurry containing eroded dump material), and also noted that line 924 of the report may be misleading about the inability to collect pre-blowout samples because the GKM tunnel was sealed, as photos show water being released during and prior to construction activities (i.e., active flow in a corrugate ditch). Finally, one reviewer suggested providing either a qualitative or quantitative discussion on likely differences between pre-release and post-release concentrations (by possibly comparing pre- and post-release equilibrium pH and DO), and also stated that the report should always clearly state whether concentrations are dissolved, colloid/particulate, or total.

Charge 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.***

Most reviewers felt that the empirical methods and modeling used to assess plume water quality characteristics were appropriately applied and interpreted; one reviewer noted that 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 reviewers did, however, provide suggestions on the empirical methods and modeling. Two reviewers noted that the explanation of plume shape should be improved. One reviewer stated that there should be more detailed discussion of the broader WASP-simulated plume compared to the empirical data-based plume, as well as on numerical dispersion. The reviewer stated that if the differences in dispersion and total mass in the plume between WASP modeling and empirical modeling can't be overcome, 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. The reviewer also suggests using the aluminum signature discussed on pages 53-54 as empirical evidence of plume timing as it moved through the San Juan River.

Charge 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.***

The reviewers provided several comments regarding the empirical methods and modeling used to assess deposition and bed sediments. One reviewer stated that the empirical model for sediments appears appropriate as long it consisted of a mass balance based on 'colloidal/particulate' mass multiplied by wave volume, initialized as the calculated Cement Ck. 'colloidal/particulate' mass. For bed sediments, however, inconsistencies in the volume of the wave (discussed in Charge Question 2) cast doubt on the model. A table of decided upon 'wave' volumes at each gaging station should be provided, along with explanations of adjustments to the data. Another reviewer stated that since the WASP mass balance between water column and river bed is not correct, its results regarding deposition/resuspension are difficult to defend and should fixed or not presented. The reviewer points out that photos show significant deposition in the rapid canyon area below Silverton, yet the WASP model made the greatest underestimation in this reach. The reviewer also notes that sampled concentrations during the spring 2016 snowmelt are a more compelling approach to the resuspension questions, and this point should be made in the text. Another reviewer emphasized the need to discuss the magnitude of uncertainty on the GKM discharge and lost plume volume. The reviewer also suggested, if possible, calibrating WASP with the empirical data, especially for the San Juan River reach. For this reach, the reviewer also questions why a simple mass balance mixing model was not investigated to assess the transport of GKM contributions.

Charge 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.***

One reviewer stated that the data were appropriately analyzed using statistical methods. This reviewer also indicated that 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. Another reviewer suggested that Figures 8-2 and 8-3 might be more useful if they illustrated samples that were taken pre- and post-plume. Finally, another reviewer provided multiple suggestions, such as describing other statistical approaches that may have been considered and providing more details in the tables (date ranges, key to colors, identifying acronyms, and clarifying captions).

Charge 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.***

One reviewer stated that the data were properly analyzed and visualized in regards to sediment metal concentrations in the post-plume. Another reviewer reiterated that the GKM discharge is a relatively small component of the overall loading in the sediments. Additionally, one reviewer noted that while a lot of effort was expended in presenting the data in graphical form, many graphs and associated captions need clarifications; this reviewer provided specific comments for Figures 6-15 through 6-19.

Charge 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.***

Most reviewers stated that the geochemical principles used in the study were appropriately applied and interpreted. One reviewer noted that a geochemical modeling simulation to mix upper Animas River water with the characterized discharge from Cement Creek would be useful. Another reviewer noted some issues with the data and modeling used to implement geochemical principals. These issues included using analytical data compromised by the coarse (0.45µm) filtration, photos which show presence of bright orange water exiting the GKM portal prior to and during the initial minutes after the blowout (how does this affect the reconstruction of the GKM blowout chemistry?), and not providing the database (or at a minimum a list of all relevant/critical species considered with their corresponding log K values) for Geochemists Workbench. Also, updated values for log Ks for calcite and dolomite should be used.

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

One reviewer commented that the exposure analysis was done satisfactorily and the comparisons of the criteria concentrations were applied appropriately, with the exception of possible impacts on reproductive success. The reviewer noted that reproductive success was not determined and the only criteria that were used were acute toxicity; however, most of the exceedances were less than the 96-hour toxicity assessments. Another reviewer noted that there needs to be a discussion of the considerable uncertainty in chemical constituent concentrations required for the analysis, due to modeling plume peaks and Cement Creek discharge. The reviewer further noted that the exposure analyses may only be generally applicable given the uncertainties. In addition, given the transient nature of the GKM plume, the reviewer questioned the applicability of the results from models like BASS that are often dependent on reference data derived from long-term exposure.

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

One reviewer stated that the potential for groundwater uptake was appropriately addressed; noting that 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. Another reviewer stated that the assessment seems reasonable for uptake from the GKM, although geochemical constraints and challenges hamper the reliability of these model calculations. Also, one reviewer provided multiple suggestions on the topic. The suggestions included trimming Chapter 9 and Appendix D to avoid duplication and including a table in Appendix D which lists calibrated model properties for GFlow and Modflow (examples provided). The reviewer also recommended conducting 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 reviewer also noted that 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; the reviewer recommended testing a series of small-scale models at a couple of the wells of concern, using a range of K values.

Part 3. Application of Soft-ware Based Analytical Models

Charge 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.***

In general, the reviewers felt that the final report addressed the mid-project external peer review comments regarding the development and application of the WASP model. A few mid-project external peer review comments, however, still should be addressed. One reviewer commented that although the discrepancy between WASP and the empirical model does receive comment in the report, it is unclear why the model cannot be calibrated and reconciled with the empirical model. The reviewer also mentioned that sulfate concentrations were only mentioned a few times (including the total load in the release); at least a discussion on sulfate should be included since it can be used for indicating dilution of fresher water. Another reviewer provided specific comments on seven of the mid-project external peer review charge questions. Comments dealt with report presentation, absence of sensitivity analysis, inadequate explanation of methods used to estimate the GKM effluent quality, inadequate attempts to address deficits in the analytical data (i.e., empirical and modeled estimations and conclusions do not appear to take into account coarse (0.45 µm) filtration), and only mentioning "clay" 3 times in the report. Additionally, the reviewer stated that a mixing/titration simulation in which pH and SICAL are calculated could be compared to observations to address mid-project review comment 89.

Charge 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.***

In general, the reviewers felt that the final report addressed the mid-project external peer review comments regarding the development and application of groundwater modeling. The reviewers gave several examples of appropriate responses to comments. One reviewer did comment, however, that with respect to running several model realizations to test reasonable ranges of input values, the K ranges tested were not as wide as they should have been, given the heterogeneous nature of alluvium deposits in braided stream environments.

Charge 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.***

Two reviewers felt that the final report addressed the mid-project external peer review comments regarding the development and application of bioaccumulation modeling. One of the reviewers further noted that the use of the lack of an observable fish kill was criticized in the mid-project review as not being sufficiently conservative. However, the reviewer agreed 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. The reviewer indicated that the draft report does examine the potential for bioaccumulation of several metals, and the treatment of this issue is thorough.

III.3 Specific Observations

Numerous specific observations were made by page and paragraph number. The observations included editorial suggestions, additional references, technical inaccuracies, and additional details on technical issues mentioned in the charge question responses.

IV. GENERAL IMPRESSIONS

General Impressions		
Reviewer Name	**Reviewer Comment**	**EPA Response**
Charles Fitts	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. 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. 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.	
Glenn Miller	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	

General Impressions		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	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, 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.	
Ronald Schmiermund	<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	

General Impressions		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	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. 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. 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. <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 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. 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	

General Impressions		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	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 data should be presented and tabulated first, with as much reliance on graphics as possible, followed by the empirical model with its justification, and finally by the WASP model with clear objectives stated. 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. Soundness of Conclusions – 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 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.	
Mark Williamson	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. 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	

General Impressions		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	represents a considerable effort and contribution to understanding the Gold King Mine release. 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.	

V. 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	Yes, I think objectives were clearly stated, and I think generally these objectives were addressed.	
Glenn Miller	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 decade's 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.	
Ronald Schmiermund	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 … It seems that each objective was addressed via extensive data analysis, although the analysis is not always clearly or extensively presented.	

Question 1: Were project objectives clearly identified and did analyses address the objectives? Please explain.		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	Specific Comments: Should an additional objective be included? Specifically, identification of strategies for better preparing for monitoring future incidents? For example: • Collecting site specific background samples prior to any work that might lead to a change in conditions (e.g., sample of GKM discharge on August 4 would have been useful). This would have mitigated the greatest problem with the GKM analysis. • Guidelines for collecting samples after any change in discharge during an operation.	
Mark Williamson	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. 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.	

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	Yes. Any discrepancies or details are minor and included below.	
Glenn Miller	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.	
Ronald Schmiermund	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 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. 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	

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**
	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. • 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 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. 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.	

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**
	• 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. Doubling the estimated GKM discharge concentration (sentences 1048 and 1049) to account for a "first flush" seems numerically arbitrary – please justify. 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 like an autonomous debris flow with limited mixing with stream water. • 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.	
Mark Williamson	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.	

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**
	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).	

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	I would answer yes for the empirical model and no for the WASP model, as discussed below under Charge Question 5.	
Glenn Miller	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. 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	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 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.	

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**
	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 for WASP. I assume the latter is to allow for a contaminant mass balance. 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. • 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	

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**
	compositions) were not included in the report. I recommend that some empirically modeled peak compositions be presented. 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).	
Mark Williamson	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.	

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	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.	
Glenn Miller	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	

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**
	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.	
Ronald Schmiermund	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. 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.	
Mark Williamson	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.	

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**
	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 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.	

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	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. 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. 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, 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,	

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**
	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 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.	
Glenn Miller	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.	
Ronald Schmiermund	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 figures, I got the essence of the approach, but the reader should not need to do that. Specific Comments:	

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**
	The explanation of plume shape seems especially weak.	
Mark Williamson	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. 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 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	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 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. 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. 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.	
Glenn Miller	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 the particulate sorption that promotes attachment to the particles.	

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**
Ronald Schmiermund	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. 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.	
Mark Williamson	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. The GKM discharge and lost plume volume notwithstanding, the discussion of uncertainty and the 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? 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	

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**
	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.	

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	Not my area of expertise.	
Glenn Miller	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.	
Ronald Schmiermund	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 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.). Table 8.6 addresses some pre- and post-event dissolved and total metal concentration statistics. 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. 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	

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**
	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?	
Mark Williamson	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.	

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	Not my area of expertise.	
Glenn Miller	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.	
Ronald Schmiermund	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. 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 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? 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. 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'?	

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**
	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-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. 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).	
Mark Williamson	As with charge question #8 above, I find no particular concerns with the presentation of metal concentrations post-plume.	

<table>
<tr><th colspan="3">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.</th></tr>
<tr><th>Reviewer Name</th><th>Reviewer Comment</th><th>EPA Response</th></tr>
<tr><td>Charles Fitts</td><td>Not my area of expertise.</td><td></td></tr>
<tr><td>Glenn Miller</td><td>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.</td><td></td></tr>
<tr><td>Ronald Schmiermund</td><td>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 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?
• 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?</td><td></td></tr>
</table>

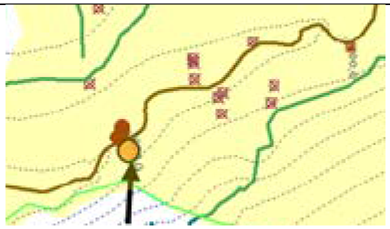
Question 9: ***Were the geochemical principles to characterize transport and fate of acid mine drainage regarding neutralization, precipitation and mineral saturation appropriately applied and interpreted? Please explain.***		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	• The use of geochemical 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. 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 dolomite. I was pleased to see the updating of the conventional assumption for atmospheric log CO_2 fugacity to -3.4 from -3.5.	
Mark Williamson	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	

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**
	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.	

Question 10: ***Were exposure analyses based on GKM concentration results appropriately applied and interpreted? Please explain.***		
Reviewer Name	**Reviewer Comment**	**EPA Response**
Charles Fitts	Not my area of expertise.	
Glenn Miller	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.	
Ronald Schmiermund	No comment	
Mark Williamson	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 the GKM plume, I wonder how applicable	

Question 10: ***Were exposure analyses based on GKM concentration results appropriately applied and interpreted? Please explain.***		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	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	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. 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. 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. 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.	

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**
	 Since the irrigation ditches and river are head-specified boundaries, what happens beyond them has very little impact on the simulated well capture zone. 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.	
Glenn Miller	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 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.	

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**
Ronald Schmiermund	No comment	
Mark Williamson	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.	

Part 3: Application of Soft-ware 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	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.	
Glenn Miller	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. 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. 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 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.	

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**
	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. 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, 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. 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.	
Ronald Schmiermund	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: • 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	

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**
	the reader that they were following the study correctly. I spent a lot of time backtracking to understand the context. • 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. • 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. • 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. • 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. • Question 7 – "Clay" only appears 3 times in the final draft, so I don't think this recommendation has been adequately addressed. • 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.	
Mark Williamson	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	

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**
	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.	

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	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.	
Glenn Miller	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. 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	No comment	
Mark Williamson	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	

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**
	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.	

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	Not my area of expertise.	
Glenn Miller	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.	
Ronald Schmiermund	No comment	
Mark Williamson	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.	

VI. SPECIFIC OBSERVATIONS

Specific Observations on Main Document, Figures and Tables			
Reviewer Name	**Page**	**Line**	**Comment or Question**
Charles Fitts	vii		Correct *Moutain Studies Institute*
Charles Fitts	9	8	*low acidity* should be *low pH*
Charles Fitts	9	22-23	Omit *not* from this sentence?
Charles Fitts	10	56	Change *specific* to *specify*
Charles Fitts	10	64	Fix "..”
Charles Fitts	10	66	*as the plume passed*?
Charles Fitts	10	72	Change *by* to *of*
Charles Fitts	10	81	Do you mean *detected* or *exceeded*? Seems like the latter.
Charles Fitts	13	189	*features*, not *feature*
Charles Fitts	14	242	*3* should be subscript, not superscript
Charles Fitts	14	250	*high*, not *low*
Charles Fitts	15	284	Weird font for *n* in **Remediatio**n.
Charles Fitts	19	409	*flow*, not *flows*
Charles Fitts	Fig.. 2-1		Schematic at Baker's Bridge is missing
Charles Fitts	20	443	*to the rest*, not *to rest*
Charles Fitts	20	453	Close quote after *providers*
Charles Fitts	Fig.. 2-3		Photo missing
Charles Fitts	21	481-482	*provider*, not *provide* Figure numbers off by 1, and actual Fig. 2-4 out of order.
Charles Fitts	21	489	*these three exceptions* needs explaining or rewording

Specific Observations on Main Document, Figures and Tables			
Reviewer Name	**Page**	**Line**	**Comment or Question**
Charles Fitts	21	492	Omit *of*
Charles Fitts	22	517	*then* not *than*
Charles Fitts	23	564	*sampled* not *sampling*
Charles Fitts	23	568	. not ..
Charles Fitts	Table 2-8		Add a *river km* column and sort in ascending order of this.
Charles Fitts	26	680	*Table ?*
Charles Fitts	26	689	*EPA* not *the EPA*
Charles Fitts	26	697	*or* not *of*
Charles Fitts	26	710	No *()* around *(40 microns)*
Charles Fitts	27	727	*release* not *blowout* for consistency
Charles Fitts	27	730	Not clear what *Two perspectives* are after reading paragraph
Charles Fitts	27	744, 749	Here and elsewhere *soft-ware* is used, but so is *software*. I like the latter
Charles Fitts	27	758	*at* not *as*
Charles Fitts	28	780	Use *background* instead of *normal?*
Charles Fitts	28	785-786	*applying bioaccumulation analysis* not *applying a bioaccumulation*
Charles Fitts	28	788	Omit 2nd *BASS*
Charles Fitts	28	801	*contaminants* not *contaminant*
Charles Fitts	29	808	Omit 1st *through*
Charles Fitts	29	825	Maybe *at a lower elevation within* not *at the lowest elevation on*
Charles Fitts	29	836	*in consultation* not *with consultation*

Specific Observations on Main Document, Figures and Tables			
Reviewer Name	**Page**	**Line**	**Comment or Question**
Charles Fitts	30	849	*where there was* not *where*
Charles Fitts	30	856	Omit comma
Charles Fitts	30	861	Omit *that is*
Charles Fitts	Fig. 3-5		*for each gage* not *for gage* in 1st sentence of caption. Last sentence of caption needs clarification. *11.33 million liters* not *1.33 million liters*.
Charles Fitts	Fig. 3-6		Omit 1st sentence of caption. Period and space after 2nd sentence. Omit last sentence.
Charles Fitts	31	892-894	The gage measures height directly – omit 1st part of sentence?
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.
Charles Fitts	31	908	Omit *is determined*
Charles Fitts	31	914	m^3/s not *m/s*
Charles Fitts	31	918, 920	*11,333,000 liters* is too precise (5 digits vs 1 digit in *3,000,000 gallons*).
Charles Fitts	31	927	*chemistry would return*
Charles Fitts	Table 3-1		*Sample concentrations* not *Samples concentrations*. Table lacks average and selected columns alluded to in caption.
Charles Fitts	32	942	*On* not *An*.
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?
Charles Fitts	Fig. 3-10		in last sentence of caption: *used to estimate* not *used estimate*
Charles Fitts	32	958	Omit parentheses.
Charles Fitts	32	967	*was about 0.5 g mercury released* not *was no mercury released*. See Fig. 3-11

Specific Observations on Main Document, Figures and Tables			
Reviewer Name	**Page**	**Line**	**Comment or Question**
Charles Fitts	Fig. 3-11		Part C is blank in my copy.
Charles Fitts	Fig. 3-13		Caption: *40 years of debris* not *40 years debris.* Omit *forty years of mine waste in this pile* at the end.
Charles Fitts	33	995	*was* not *were*
Charles Fitts	33	996	Omit *be.* This sentence is an unclear run-on and should be re-written to clarify.
Charles Fitts	33	999-1000	Clarify *all metals estimates reference from the 16:00 hour sample.*
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 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	Fig. 3-14		The low pH values before 12:45 seem erroneous (low), since the plume flow peaked at 12:45, and background pH ~ 4.7.
Charles Fitts	34	1048	*2001* not *201*
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.
Charles Fitts	35	1060	*were estimated to be roughly* not *were roughly*
Charles Fitts	Fig. 3-17		Caption: *at 12:45*, not *as 12:45*
Charles Fitts	35	1073	*the end of the flow defined period* - clarify this.
Charles Fitts	35	1083-1089	Choppy writing, especially in 1st sentence. It's not clear how the estimates of eroded volume and mass fit into the discussion, since it starts with saying this is not a way to validate total mass loading.
Charles Fitts	37	1100	Omit *dissolved* (this includes colloid/particulate)

Specific Observations on Main Document, Figures and Tables			
Reviewer Name	**Page**	**Line**	**Comment or Question**
Charles Fitts	37	1105-1107	Awkward, unclear sentence. Suggest re-writing it.
Charles Fitts	37	1113	Omit *would have*
Charles Fitts	Fig. 4-2		Caption: fix *with elevated with elevated. as a coherent mass* not *as coherent mass*
Charles Fitts	37	1125	*consideration* not *considerations*
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.
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.
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.
Charles Fitts	39	1211	*enable us to approximate the movement of metals* not *enable us move the metals*
Charles Fitts	39	1215-1217	I presume this sentence is defining the empirical model. Add *(empirical model)* at the end of the sentence to clarify if this is true.
Charles Fitts	39	1221	*empirical model fully understands timing* – models don't understand anything, but people sometimes do. Clarify what is meant here.
Charles Fitts	Fig. 4-6		*samples* not *sample*. *used for flow* not *used to for flow*
Charles Fitts	39	1225	*sampled* not *measured*
Charles Fitts	40	1246	*4-7* not *4-8*
Charles Fitts	40	1251	*that* not *there*
Charles Fitts	40	1262	*45 minutes later than the nearest sample in time*? Please clarify
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?
Charles Fitts	Fig. 4-8		Fix *fitting fitting*, omit *(A).*

Specific Observations on Main Document, Figures and Tables			
Reviewer Name	**Page**	**Line**	**Comment or Question**
Charles Fitts	41	1289-1290	*of the plume model* not *of the plume as it traveled*. Also *simulated* not *transported* (mixing terms for reality with those for a model)
Charles Fitts	Fig. 4-9		*model* not *mode*. The caption only mentions particulate mass, but WASP was simulating both particulate and dissolved mass. Please clarify.
Charles Fitts	Fig. 4-10		*but not necessarily* not *but not necessarily consists*
Charles Fitts	41	1304	*are* not *is*
Charles Fitts	42	1348	Omit *occur*
Charles Fitts	44	1399	Omit *in*
Charles Fitts	44	1403	*simulated* not *a whole enabled* (keep clear terminology that distinguishes model vs. reality)
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.
Charles Fitts	44	1412-1415	These sentences have the same problem as described in the preceding point about Fig. 4-12, and use wording that confuses model-simulated behavior with reality. Also, say *WASP* not *GK WASP* or *GKM WASP* to be consistent throughout the text and Figures.
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 (conductance or whatever it is, rather than the undefined *Sonde shape factor*).
Charles Fitts	44	1429	Omit *on the leading side of the plume*, since dispersion affects leading and trailing edges (note symmetry of WASP plumes in Fig. 4-13).
Charles Fitts	45	1463	*low background* not *no background*

Specific Observations on Main Document, Figures and Tables			
Reviewer Name	**Page**	**Line**	**Comment or Question**
Charles Fitts	45	1467	*estimate* not *determine*
Charles Fitts	45	1470	Omit *what*
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.
Charles Fitts	45	1479	*as a coherent* not *as coherent*
Charles Fitts	46	1483	Omit *past this point due to lack of data*
Charles Fitts	46	1507	*a shorter duration plume* not *for a shorter period…that the bulk*
Charles Fitts	46	1511-1512	*good estimate of the travel time and timing* not *reliable record of the travel time and can provide the timing*
Charles Fitts	46	1516	*at peak times predicted by* not *as predicted by*
Charles Fitts	46	1517-1521	I find this paragraph confusing. Clarify what is meant by *for data providers* and the last two sentences.
Charles Fitts	Table 4-4		The caption uses wording that confuses simulation results with reality. I think this is all about WASP simulation results and should be clearly labeled as such (e.g. *Simulated Plume Duration* not *Plume Duration*). Explain source of estimated time at peak (is it WASP simulated peaks or observed peaks? I get the latter notion from the text).
Charles Fitts	47	1529	Omit 2nd *movement*
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.
Charles Fitts	47	1545	*concentrations were generally much closer* to what?
Charles Fitts	47	1547	Omit *from*
Charles Fitts	47	1549	Omit *shown in Chapter 4*

Specific Observations on Main Document, Figures and Tables			
Reviewer Name	**Page**	**Line**	**Comment or Question**
Charles Fitts	Fig. 5-2		*empirical model*. Not *empirical effort (cite).*
Charles Fitts	47	1558	*very large influx* not *very influx*
Charles Fitts	47	1560	*in the San Juan*
Charles Fitts	48	1596	*By the time*
Charles Fitts	Fig. 5-4		Caption and Figure mismatched (no orange dotted line, has criteria lines which are not discussed in text)
Charles Fitts	Fig. 5-5		Caption: *reaching 0.1% of the initial concentration after mixing with the San Juan River* (See graph)
Charles Fitts	49	1636	*favor generation of Fe*
Charles Fitts	Fig. 5-9		*dipped slightly as the front of the plume passed.* Omit *as the plume.* The notion of *linear* change in pH from 0-100 km is based on data with a huge central gap. I would not tout *linear* based on this.
Charles Fitts	50	1664	*(Fig. 5-7)* not *(Fig. 9)*
Charles Fitts	50	1677	Omit *is*
Charles Fitts	50	1681	*water sampling suggesting by application* Reword to clarify
Charles Fitts	50	1688	*virtually linear* - see comment for Fig. 5-9
Charles Fitts	Fig. 5-11		Caption: *11,000 kg* doesn't match with plot (~15,500 kg)
Charles Fitts	Fig. 5-12		Make parts A and B, not A and C. Color of As symbol wrong in legend. Caption: *were mostly sorbed* not *one mostly sorbed*
Charles Fitts	Fig. 5-13		2^{nd} and 3^{rd} panels are same. Label on 4^{th} panel says *Aztec*, not *Farmington*.
Charles Fitts	52	1741	*at the peak*
Charles Fitts	52	1763	Omit *has*
Charles Fitts	Fig. 5-15		*was very turbid and masked*

Specific Observations on Main Document, Figures and Tables			
Reviewer Name	**Page**	**Line**	**Comment or Question**
Charles Fitts	Fig. 5-16		*sediment increased by 10-fold. …increase as the San Juan flowed westward.*
Charles Fitts	52	1772	*were elevated in water that entered the San Juan from the Animas*
Charles Fitts	52	1775	*in the San Juan*
Charles Fitts	54	1822	*the* not *tge*
Charles Fitts	Fig. 5-24		Copper plots are repeated in panels 2 and 3
Charles Fitts	54	1838	*in this who Animas concoction*? Some mixed drink!
Charles Fitts	55	1853	*the alkalinity of the Animas River neutralized*
Charles Fitts	56	1897	*Fig.. 6-3*
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 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.
Charles Fitts	Figs. 6-4, 6-5		State in the caption that these are based on the empirical model.
Charles Fitts	Fig. 6-6		Clarify/edit this sentence: *Major exception was lead that had source as much lead as the San Juan.*

Specific Observations on Main Document, Figures and Tables			
Reviewer Name	**Page**	**Line**	**Comment or Question**
Charles Fitts	Fig. 6-7		Text says the data is from the empirical model, caption says it is from both empirical and WASP models. It should be only the empirical model, since the WASP model overstates mass.
Charles Fitts	Fig. 6-8		Panel D is referred to in caption, but not A,B,C, E. Figures are not labeled with letters.
Charles Fitts	Fig. 6-13		No mention of time frame for image A. Maybe it is not needed.
Charles Fitts	6-14		Panel B is referred to, but panels have no letters.
Charles Fitts	59	2051	*Metal mass in the bed of the San Juan River is very low…*
Charles Fitts	59	2054	*Shown on the right side of Figure 6-18...*
Charles Fitts	60	2058	*exceed the highest natural metals concentrations*
Charles Fitts	Fig. 6-17		Wording should be clearer throughout that the orange line is simulated. As noted elsewhere, these WASP results are not consistent with observed concentrations in water or sediment, especially in the canyon reach.
Charles Fitts	Fig. 6-18		*declining to historic present* – reword, clarify
Charles Fitts	60	2079	*Generally, sediment metals concentrations*
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 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	2112	*Downstream of Silverton, mass from the Gold King release increased pre-existing mass by 2-3%.* This analysis is very crude since the 5 cm vertical thickness of metals-rich deposits is quite uncertain. Uncertainty in this parameter and the results should be discussed.
Charles Fitts	Fig. 6-21		Similar comments to Fig. 6-19.

Specific Observations on Main Document, Figures and Tables			
Reviewer Name	**Page**	**Line**	**Comment or Question**
Charles Fitts	61	2124	*relatively uniform*
Charles Fitts	61	2135	*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.
Charles Fitts	62	2150-2151 and 2153-2156	Awkward, unclear sentences could use editing.
Charles Fitts	Fig. 6-22		Not referenced in text.
Charles Fitts	62	2177	*are strongly associated with the streambed?* Needs clarification.
Charles Fitts	62-63		Discussion of Figs 6-24 and 6-25 doesn't draw any conclusions. It seems important to the resuspension discussion to note that spring 2016 dissolved concentrations are within historic high-flow ranges, and so are suspended concentrations except for Cadmium.
Charles Fitts	64	2229-2230	Unclear sentence.
Charles Fitts	64	2237	*simulations for total metals are shown*
Charles Fitts	64	2246	*of individual metals were*
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*.
Charles Fitts	64	2271	*differences account for*
Charles Fitts	65	2290	*Even holding*
Charles Fitts	Fig. 6-31		Is missing.
Charles Fitts	97	3459	Omit 2nd *plume*
Charles Fitts	97	3464	*are private*
Charles Fitts	97	3466	Missing period at end of sentence

Specific Observations on Main Document, Figures and Tables			
Reviewer Name	**Page**	**Line**	**Comment or Question**
Charles Fitts	97	3466-3468	Sentence needs editing.
Charles Fitts	97	3469	Omit *Due to the higher metals concentrations during plume movement,*
Charles Fitts	97	3479	*…to identify wells, which due to their geographic setting or pumping history, may have had the potential…*
Charles Fitts	97	3486	Omit *characteristics of*
Charles Fitts	Fig. 9-1		I assume last two sentences in caption are notes to self?
Charles Fitts	97	3493	*This chapter emphasizes*
Charles Fitts	97	3494	*are provided in Appendix D*
Charles Fitts	98	3497	*deposits that snake*
Charles Fitts	98	3512-3513	*high permeability sands and gravels and low permeability silts*
Charles Fitts	98	3514	*Fig. 9-3*. This and other insert Figure numbers are off.
Charles Fitts	Fig. 9-5		Caption give too much detail, given this is just an example, and this site from this project, which may be confusing.
Charles Fitts	98	3528-3529	*some from water flowing towards the stream from upland areas*. This was a steady analysis, so no change in aquifer storage.
Charles Fitts	98	3529	*that come from*
Charles Fitts	99	6553	the elevated metals signal soon after the river plume passed.
Charles Fitts	Fig. 9-10		*Only three wells (including 5 community wells)?* Make sure the wording makes it clear these are simulation results, not reality.
Charles Fitts	Fig. 9-11		*Flushing of the aquifer occurs in about*
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?
Charles Fitts	101	1620	*household wells*

Specific Observations on Main Document, Figures and Tables			
Reviewer Name	**Page**	**Line**	**Comment or Question**
Charles Fitts	101	3626	*water levels after the GKM release*
Charles Fitts	101	3627	Omit *the potential to have*
Charles Fitts	101	3651	Omit *using particle tracking the*
Charles Fitts	Fig. 9-16		*Flushing of the aquifer occurs in about*
Charles Fitts	102	3673	*Dissolved background zinc concentrations*
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?
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.
Charles Fitts	Fig. 9-17 and D-39		Use the same color in the legend for both dissolved and colloidal (e.g. both blue). At present it is confusing with river vs. well colors.
Charles Fitts	103	3719	Omit *(*
Charles Fitts	104	3752	*break-through*
Charles Fitts	104	3753	*at these wells closest to the river would be days to weeks*
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.
Charles Fitts	105	3769	Omit *Animas River to the San Juan River*
Charles Fitts	105	3776	*the weeks during the release*

Specific Observations on Main Document, Figures and Tables			
Reviewer Name	**Page**	**Line**	**Comment or Question**
Charles Fitts	106	3801-3804	Grammar problems, needs editing
Charles Fitts	106	3812	*Gold King Mine event released*
Charles Fitts	106	3818	*entrained or dissolved metals mass in Cement Creek from previously deposited sediment.*
Charles Fitts	106	3824-3825	*varying by organism and exposure.*
Charles Fitts	106	3832	*due to dispersion.*
Charles Fitts	107	3862-3865	This discussion neglects the empirical derived deposition in the canyon reach (see Fig. 6-9a), which is more data-based than the WASP deposition pattern.
Charles Fitts	107	3874	*once it entered the San Juan River*
Charles Fitts	108	3916	*as the plume passed, to prevent*
Charles Fitts	108	3923	*of a variety of organisms*
Charles Fitts	109	3939	*exceeded* not *detected*
Charles Fitts	109	3943	*a large*
Charles Fitts	109	3952	*would likely dissolve*
Charles Fitts	109	3962	Omit *documented by monitoring during the plume period*
Charles Fitts	110	4015	Add that except for the one mid-Animas well, no metal concentration anomalies in well water were detected near the time of the release.
Charles Fitts	111	4023	*when the sediments*
Charles Fitts	111	4033	Omit *from*
Charles Fitts	111	4049	*It is not known whether this pattern persisted*
Glenn Miller	9	8	"very low acidity" should be "very high acidity"
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

Specific Observations on Main Document, Figures and Tables			
Reviewer Name	**Page**	**Line**	**Comment or Question**
			metals tends to minimize the issue (although it is correct, a new reader will think that the Document is trying to minimize the impact)
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?
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.
Glenn Miller	9	41	What metals have increased toxicity in the particulate form? Are there any?
Glenn Miller	11	113	Add " or lower concentrations of calcium and sulfate"
Glenn Miller	13	208	Change "reducing" to "oxidizing" AMD is created when the highly reduced iron pyrite is oxidized to sulfuric acid and ferrous/ferric iron
Glenn Miller	14	258	Remove "oxidizing" Raising the pH changes the solubility, not the valence
Glenn Miller	21	492	Remove "of"
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.
Glenn Miller	45	1457	Remove the "e"
Glenn Miller	47	1558	"There was a very influx"???
Glenn Miller	49	1618	This should be "neutralizing", not oxidizing
Glenn Miller	90	3241; 3243	Subscripts, not superscripts.
Glenn Miller	104	3752	Break-through
Glenn Miller	108	3943	Remove "al"
Glenn Miller	Fig. 3-10		The legend is confusing. What is the "selected sample" Where was "cc" taken?
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?

Specific Observations on Main Document, Figures and Tables			
Reviewer Name	**Page**	**Line**	**Comment or Question**
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.
Glenn Miller	Fig. 6-1		Where is the figure?
Glenn Miller	Table 7-10		What are the units of time?
Glenn Miller	Figures 9-13, 9-14 and 9-16 to 9-18		All of these figures are very difficult to read. In fact, on my computers, many of the figures are very fuzzy.
Ronald Schmiermund	General		Given the importance of "dissolved" versus "colloidal/particulate", the word "metal(s)" should always be preceded by a modifier for clarity
Ronald Schmiermund	8	Fig 1-6, line 5	… where basic (i.e., alkaline) …
Ronald Schmiermund	9	8	Replace "low acidity" with either "high acidity" or "low pH"
Ronald Schmiermund	14	246-247	Mine-waste rock has not been pulverized to remove sulfides – only applies to tailings
Ronald Schmiermund	14	250-251	Replace "low to moderate acidity" with either "high to moderate acidity" or "low to moderate pH"
Ronald Schmiermund	14	265-259	Dissolved metal concentrations are *generally* suppressed… Increasing pH *allows* oxidizing. Secondary iron sulfate minerals exist and form in *acidic* conditions due to oxidation typically accompanied by evaporation.
Ronald Schmiermund	14	259	Spelling "oxyhydroxides"
Ronald Schmiermund	15	263-266	Aluminum precipitates are white and may exist in the absence of iron
Ronald Schmiermund	15	268	By definition "colloids" will never be lost from the water column, especially in moving water. Only after sufficient aggregation/flocculation, typically in response to changes in ambient chemistry and/or time, can they be lost from the water column.

Specific Observations on Main Document, Figures and Tables			
Reviewer Name	**Page**	**Line**	**Comment or Question**
Ronald Schmiermund	15	273-283	Please clarify the differences (composition, stability and distribution) between "waste rock/dumps" and "tailings/tailings ponds or piles"
Ronald Schmiermund	15	299-300	Clarify that GKM is one of the 80 mines mentioned above
Ronald Schmiermund	16	327	Remove "of" Clarify " … the same geochemical reactions (of) routinely observed near Silverton …"
Ronald Schmiermund	20	453	Close quote after *data providers*
Ronald Schmiermund	Table 2-1	footnote	Define "SADIE"
Ronald Schmiermund	21	502	Reword sentence …"potential risk to contaminants …"
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 table. There should be a reference to the SOPs for each entity.
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).
Ronald Schmiermund	Fig. 2.5		Include "RCWWN" in list of abbreviations

Specific Observations on Main Document, Figures and Tables			
Reviewer Name	**Page**	**Line**	**Comment or Question**
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 acknowledgement, that a reason for using that convention should be offered and an explanation of the consequences provided.
Ronald Schmiermund	27	718	Sp. "ware"
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?
Ronald Schmiermund	27	724-725	"latter" should be "later". The sentence implies that the only mechanism for subsequent metal mobilization is re-entrainment of settled solids and does not acknowledge desorption.
Ronald Schmiermund	27	730	For clarity modify the sentence "….. throughout the analysis: one based on contaminant concentration and one based on contaminant mass."
Ronald Schmiermund	27	737	Should read "… concentration (expressed as mass of contaminant per unit volume of water or unit mass of sediment) …"
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 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	29		Include citation for "BOR 2015" and "EPA 2016"
Ronald Schmiermund	Fig. 3-3	Caption last line	"A much *smaller?* flood wave ……."
Ronald Schmiermund	Fig. 3-5	caption	Last sentence is a fragment.
Ronald Schmiermund	Fig. 3-8		Seems that iron should be included in the plots (secondary Y axis)?

Specific Observations on Main Document, Figures and Tables			
Reviewer Name	**Page**	**Line**	**Comment or Question**
Ronald Schmiermund	31	926-927	"…. pool re-established, but that chemistry *would?* return …. This expectation is likely to depend on the degree of aeration/oxidation that becomes established after the blowout.
Ronald Schmiermund	Table 3-1	Caption, last line	Sentence fragment
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.
Ronald Schmiermund	Fig. 3-14	caption	State the location for this data (USGS gage station 12.5 km. downstream)
Ronald Schmiermund	33	1004	"concentration" used twice. Should also clarify that "colloidal/particulate" concentrations are being used.
Ronald Schmiermund	33	1009	Where is the equation?
Ronald Schmiermund	37	1100	"…. acidic water, dissolved *and suspended* metals …."
Ronald Schmiermund	Fig. 4-7	Plot	Does the red triangle signify something else?
Ronald Schmiermund	Fig. 4-2	caption	"with elevated" is repeated '… traveled as *a* coherent mass …"
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.
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?
Ronald Schmiermund	Fig. 4-5A		No units on vertical axis
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

Specific Observations on Main Document, Figures and Tables			
Reviewer Name	**Page**	**Line**	**Comment or Question**
Ronald Schmiermund	39	1198-1200	The conductivity as measured by the sondes does not necessarily confirm that metal concentrations were behaving 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.
Ronald Schmiermund	40	1246	Should be Fig. 4-7
Ronald Schmiermund	40	1247-1248	Reword sentence after "(SUIT)"
Ronald Schmiermund	Fig. 4-8	citation	2nd line, "fitting" is repeated 3rd line, extraneous "(A)" Site should be identified. Please plot water volume across entire plot.
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.
Ronald Schmiermund	41	1309	Site 09358550 is 0.72+ miles upstream of the confluence – hardly "just upstream"
Ronald Schmiermund	41	1304	"is" should be "are"
Ronald Schmiermund	Fig. C-9		I'm quite familiar with Geochemist's Workbench, but I can't follow the figure caption
Ronald Schmiermund	Fig. C-10		Explain differences between plots
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 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.

Specific Observations on Main Document, Figures and Tables			
Reviewer Name	**Page**	**Line**	**Comment or Question**
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
Ronald Schmiermund	107	3853	'acidity' is an intensive, not an extensive quantity, quantity (no volume association)
Ronald Schmiermund	40	1244-1245	Explain "basin-scale relationship"
Ronald Schmiermund	Fig. 6-1		missing
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.

Specific Observations on Appendices				
Reviewer Name	**Appendix**	**Page**	**Line**	**Comment or Question**
Charles Fitts	D	2	23	*Animas River is gaining water*
Charles Fitts	D	4	81	*cumulative tributary stream flows upstream and the measured streamflow downstream…*
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.
Charles Fitts	D	5	96-97	*verifying locations of the sampled wells using hand-held GPS, and surveying well water levels* (keep the items in a list with parallel grammar)
Charles Fitts	D	5	101	*Animas River is losing water*
Charles Fitts	D	8	136-140	Omit discussion of FDM – most readers will know this, and if they don't they can easily look it up.

Specific Observations on Appendices				
Reviewer Name	**Appendix**	**Page**	**Line**	**Comment or Question**
Charles Fitts	D	8	150	Omit *(water flows down the hydraulic potential gradient)*
Charles Fitts	D	8	158	*more coarse representations*
Charles Fitts	D	8	159-160	Omit these 2 sentences and Fig. D7 – too much detail. You've given plenty of references for the curious.
Charles Fitts	D	9	177	*baseflow compared to field observed baseflow*
Charles Fitts	D	9	181	*an individual pumping well.*
Charles Fitts	D	11		No need for Figure 8 – MODFLOW references are enough.
Charles Fitts	D	12	292-294	Awkward sentence should be edited.
Charles Fitts	D	16	370 and 372	*slide 12?*
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.
Charles Fitts	D	23	476	*map of simulated hydraulic head contours*
Charles Fitts	D	23	479	There is no Fig. D-16c
Charles Fitts	D	25	487	*the ditches in Farmington is shown*
Charles Fitts	D	Fig. D-20		Figure doesn't include lateral flows from rock into alluvium or irrigation flows lost to ET.
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.
Charles Fitts	D	30	552	*support*
Charles Fitts	D	31	557	*nominated* for what?

Specific Observations on Appendices				
Reviewer Name	**Appendix**	**Page**	**Line**	**Comment or Question**
Charles Fitts	D	31	558-559	Explain what you mean by *sanitation wells.*
Charles Fitts	D	32	578	*Animas River, Durango CO*
Charles Fitts	D	38	657, 659	Tau symbol is missing.
Charles Fitts	D	Fig. D-37		The minimum travel time would be to the top of the well screen. Perhaps start particles at the top of the well screen if that is possible.
Charles Fitts	D	44	718	*associated with water entering the lower half of the well screen* I assume the well used the lower two layers and maintained consistent head between levels? Please spell that out in the text.
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.
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).
Charles Fitts	D	47	777	*community well was perhaps*
Charles Fitts	D	48	794	*than was observed*
Charles Fitts	D	48	796	*colloidal forms.*
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.
Glenn Miller				[Reviewer provided no comment]

Specific Observations on Appendices				
Reviewer Name	**Appendix**	**Page**	**Line**	**Comment or Question**
Ronald Schmiermund	F			missing
Mark Williamson				[Reviewer provided no comment]

VII. INDIVIDUAL PEER REVIEWER COMMENTS

Review by:

Charles Fitts, Ph.D.

Peer Review Comments on EPA's Draft Document "*Analysis of the Fate and Transport of Metals Released from the Gold King Mine in the Animas and San Juan Rivers*"

Charles Fitts, Ph.D.
Fitts Geosolutions, LLC
October 16, 2016

I. GENERAL IMPRESSIONS

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.

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.

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.

II. RESPONSE TO CHARGE QUESTIONS

Part 1. Overall Project and Analysis

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

Yes, I think objectives were clearly stated, and I think generally these objectives were addressed.

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

Yes. Any discrepancies or details are minor and included below.

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

I would answer yes for the empirical model and no for the WASP model, as discussed below under Charge Question 5.

Part 2. Fate and Transport

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

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.

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

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.

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.

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, 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 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.

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

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 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.

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.

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.

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.

Not my area of expertise.

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.

Not my area of expertise.

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.

Not my area of expertise.

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

Not my area of expertise.

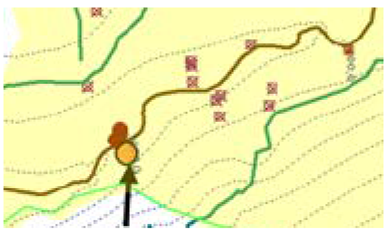
11. Was the potential for groundwater uptake from the Gold King Mine appropriately applied and interpreted? Please explain.

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.

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.

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.

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.



Since the irrigation ditches and river are head-specified boundaries, what happens beyond them has very little impact on the simulated well capture zone.

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.

Part 3. Application of Soft-ware Based Analytical Models

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.

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.

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.

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.

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.

Not my area of expertise.

III. SPECIFIC OBSERVATIONS

Specific Observations on Main Document, Figures, and Tables		
Page	**Line**	**Comment or Question**
vii		Correct *Moutain Studies Institute*
9	8	*low acidity* should be *low pH*
9	22-23	Omit *not* from this sentence?
10	56	Change *specific* to *specify*
10	64	Fix "..”
10	66	*as the plume passed*?
10	72	Change *by* to *of*
10	81	Do you mean *detected* or *exceeded*? Seems like the latter.
13	189	*features*, not *feature*
14	242	*3* should be subscript, not superscript
14	250	*high*, not *low*
15	284	Weird font for *n* in **Remediatio**n.
19	409	*flow*, not *flows*
Fig. 2-1		Schematic at Baker's Bridge is missing
20	443	*to the rest*, not *to rest*
20	453	Close quote after *providers*
Fig. 2-3		Photo missing

Specific Observations on Main Document, Figures, and Tables		
Page	**Line**	**Comment or Question**
21	481-482	*provider*, not *provide* Figure numbers off by 1, and actual Fig. 2-4 out of order.
21	489	*these three exceptions* needs explaining or rewording
21	492	Omit *of*
22	517	*then* not *than*
23	564	*sampled* not *sampling*
23	568	. not ..
Table 2-8		Add a *river km* column and sort in ascending order of this.
26	680	*Table ?*
26	689	*EPA* not *the EPA*
26	697	*or* not *of*
26	710	No *()* around *(40 microns)*
27	727	*release* not *blowout* for consistency
27	730	Not clear what *Two perspectives* are after reading paragraph
27	744, 749	Here and elsewhere *soft-ware* is used, but so is *software*. I like the latter
27	758	*at* not *as*
28	780	Use *background* instead of *normal?*
28	785-786	*applying bioaccumulation analysis* not *applying a bioaccumulation*
28	788	Omit 2nd *BASS*
28	801	*contaminants* not *contaminant*
29	808	Omit 1st *through*
29	825	Maybe *at a lower elevation within* not *at the lowest elevation on*
29	836	*in consultation* not *with consultation*
30	849	*where there was* not *where*
30	856	Omit comma
30	861	Omit *that is*
Fig. 3-5		*for each gage* not *for gage* in 1st sentence of caption. Last sentence of caption needs clarification. *11.33 million liters* not *1.33 million liters*.
Fig. 3-6		Omit 1st sentence of caption. Period and space after 2nd sentence. Omit last sentence.
31	892-894	The gage measures height directly – omit 1st part of sentence?
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.
31	908	Omit *is determined*
31	914	*m^3/s* not *m/s*
31	918, 920	*11,333,000 liters* is too precise (5 digits vs 1 digit in *3,000,000 gallons*).
31	927	*chemistry would return*

Specific Observations on Main Document, Figures, and Tables		
Page	**Line**	**Comment or Question**
Table 3-1		*Sample concentrations* not *Samples concentrations*. Table lacks average and selected columns alluded to in caption.
32	942	*On* not *An*.
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?
Fig. 3-10		in last sentence of caption: *used to estimate* not *used estimate*
32	958	Omit parentheses.
32	967	*was about 0.5 g mercury released* not *was no mercury released*. See Fig. 3-11
Fig. 3-11		Part C is blank in my copy.
Fig. 3-13		Caption: *40 years of debris* not *40 years debris.* Omit *forty years of mine waste in this pile* at the end.
33	995	*was* not *were*
33	996	Omit *be.* This sentence is an unclear run-on and should be re-written to clarify.
33	999-1000	Clarify *all metals estimates reference from the 16:00 hour sample.*
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 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.
Fig. 3-14		The low pH values before 12:45 seem erroneous (low), since the plume flow peaked at 12:45, and background pH ~ 4.7.
34	1048	*2001* not *201*
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.
35	1060	*were estimated to be roughly* not *were roughly*
Fig. 3-17		Caption: *at 12:45*, not *as 12:45*
35	1073	*the end of the flow defined period* - clarify this.
35	1083-1089	Choppy writing, especially in 1st sentence. It's not clear how the estimates of eroded volume and mass fit into the discussion, since it starts with saying this is not a way to validate total mass loading.
37	1100	Omit *dissolved* (this includes colloid/particulate)
37	1105-1107	Awkward, unclear sentence. Suggest re-writing it.
37	1113	Omit *would have*

Specific Observations on Main Document, Figures, and Tables		
Page	**Line**	**Comment or Question**
Fig. 4-2		Caption: fix *with elevated with elevated. as a coherent mass* not *as coherent mass*
37	1125	*consideration* not *considerations*
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.
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.
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.
39	1211	*enable us to approximate the movement of metals* not *enable us move the metals*
39	1215-1217	I presume this sentence is defining the empirical model. Add *(empirical model)* at the end of the sentence to clarify if this is true.
39	1221	*empirical model fully understands timing* – models don't understand anything, but people sometimes do. Clarify what is meant here.
Fig. 4-6		*samples* not *sample*. *used for flow* not *used to for flow*
39	1225	*sampled* not *measured*
40	1246	*4-7* not *4-8*
40	1251	*that* not *there*
40	1262	*45 minutes later than the nearest sample in time*? Please clarify
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?
Fig. 4-8		Fix *fitting fitting*, omit *(A).*
41	1289-1290	*of the plume model* not *of the plume as it traveled*. Also *simulated* not *transported* (mixing terms for reality with those for a model)
Fig. 4-9		*model* not *mode*. The caption only mentions particulate mass, but WASP was simulating both particulate and dissolved mass. Please clarify.
Fig. 4-10		*but not necessarily* not *but not necessarily consists*
41	1304	*are* not *is*
42	1348	Omit *occur*
44	1399	Omit *in*
44	1403	*simulated* not *a whole enabled* (keep clear terminology that distinguishes model vs. reality)
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

Specific Observations on Main Document, Figures, and Tables		
Page	**Line**	**Comment or Question**
		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.
44	1412-1415	These sentences have the same problem as described in the preceding point about Fig. 4-12, and use wording that confuses model-simulated behavior with reality. Also, say *WASP* not *GK WASP* or *GKM WASP* to be consistent throughout the text and Figures.
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 (conductance or whatever it is, rather than the undefined *Sonde shape factor*).
44	1429	Omit *on the leading side of the plume*, since dispersion affects leading and trailing edges (note symmetry of WASP plumes in Fig. 4-13).
45	1463	*low background* not *no background*
45	1467	*estimate* not *determine*
45	1470	Omit *what*
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.
45	1479	*as a coherent* not *as coherent*
46	1483	Omit *past this point due to lack of data*
46	1507	*a shorter duration plume* not *for a shorter period…that the bulk*
46	1511-1512	*good estimate of the travel time and timing* not *reliable record of the travel time and can provide the timing*
46	1516	*at peak times predicted by* not *as predicted by*
46	1517-1521	I find this paragraph confusing. Clarify what is meant by *for data providers* and the last two sentences.
Table 4-4		The caption uses wording that confuses simulation results with reality. I think this is all about WASP simulation results and should be clearly labeled as such (e.g. *Simulated Plume Duration* not *Plume Duration*). Explain source of estimated time at peak (is it WASP simulated peaks or observed peaks? I get the latter notion from the text).
47	1529	Omit 2nd *movement*
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

Specific Observations on Main Document, Figures, and Tables		
Page	**Line**	**Comment or Question**
		for part B. Why present the red line (San Juan = distilled water) since it is so unreasonable? Correct *XXXX* in caption.
47	1545	*concentrations were generally much closer* to what?
47	1547	Omit *from*
47	1549	Omit *shown in Chapter 4*
Fig. 5-2		*empirical model*. Not *empirical effort (cite).*
47	1558	*very large influx* not *very influx*
47	1560	*in the San Juan*
48	1596	*By the time*
Fig. 5-4		Caption and Figure mismatched (no orange dotted line, has criteria lines which are not discussed in text)
Fig. 5-5		Caption: *reaching 0.1% of the initial concentration after mixing with the San Juan River* (See graph)
49	1636	*favor generation of Fe*
Fig. 5-9		*dipped slightly as the front of the plume passed.* Omit *as the plume.* The notion of *linear* change in pH from 0-100 km is based on data with a huge central gap. I would not tout *linear* based on this.
50	1664	*(Fig. 5-7)* not *(Fig. 9)*
50	1677	Omit *is*
50	1681	*water sampling suggesting by application* Reword to clarify
50	1688	*virtually linear* - see comment for Fig. 5-9
Fig. 5-11		Caption: *11,000 kg* doesn't match with plot (~15,500 kg)
Fig. 5-12		Make parts A and B, not A and C. Color of As symbol wrong in legend. Caption: *were mostly sorbed* not *one mostly sorbed*
Fig. 5-13		2nd and 3rd panels are same. Label on 4th panel says *Aztec*, not *Farmington*.
52	1741	*at the peak*
52	1763	Omit *has*
Fig. 5-15		*was very turbid and masked*
Fig. 5-16		*sediment increased by 10-fold. …increase as the San Juan flowed westward.*
52	1772	*were elevated in water that entered the San Juan from the Animas*
52	1775	*in the San Juan*
54	1822	*the* not *tge*
Fig. 5-24		Copper plots are repeated in panels 2 and 3
54	1838	*in this who Animas concoction*? Some mixed drink!
55	1853	*the alkalinity of the Animas River neutralized*
56	1897	*Fig. 6-3*
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 mass, so the only way the WASP water column mass can be higher at downstream locations

Specific Observations on Main Document, Figures, and Tables		
Page	**Line**	**Comment or Question**
		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.
Figs. 6-4, 6-5		State in the caption that these are based on the empirical model.
Fig. 6-6		Clarify/edit this sentence: *Major exception was lead that had source as much lead as the San Juan.*
Fig. 6-7		Text says the data is from the empirical model, caption says it is from both empirical and WASP models. It should be only the empirical model, since the WASP model overstates mass.
Fig. 6-8		Panel D is referred to in caption, but not A,B,C, E. Figures are not labeled with letters.
Fig. 6-13		No mention of time frame for image A. Maybe it is not needed.
6-14		Panel B is referred to, but panels have no letters.
59	2051	*Metal mass in the bed of the San Juan River is very low…*
59	2054	*Shown on the right side of Figure 6-18...*
60	2058	*exceed the highest natural metals concentrations*
Fig. 6-17		Wording should be clearer throughout that the orange line is simulated. As noted elsewhere, these WASP results are not consistent with observed concentrations in water or sediment, especially in the canyon reach.
Fig. 6-18		*declining to historic present* – reword, clarify
60	2079	*Generally, sediment metals concentrations*
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 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.

Specific Observations on Main Document, Figures, and Tables		
Page	**Line**	**Comment or Question**
61	2112	*Downstream of Silverton, mass from the Gold King release increased pre-existing mass by 2-3%.* This analysis is very crude since the 5 cm vertical thickness of metals-rich deposits is quite uncertain. Uncertainty in this parameter and the results should be discussed.
Fig. 6-21		Similar comments to Fig. 6-19.
61	2124	*relatively uniform*
61	2135	*concentrations*
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.
62	2150-2151 and 2153-2156	Awkward, unclear sentences could use editing.
Fig. 6-22		Not referenced in text.
62	2177	*are strongly associated with the streambed?* Needs clarification.
62-63		Discussion of Figs 6-24 and 6-25 doesn't draw any conclusions. It seems important to the resuspension discussion to note that spring 2016 dissolved concentrations are within historic high-flow ranges, and so are suspended concentrations except for Cadmium.
64	2229-2230	Unclear sentence.
64	2237	*simulations for total metals are shown*
64	2246	*of individual metals were*
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*.
64	2271	*differences account for*
65	2290	*Even holding*
Fig. 6-31		Is missing.
97	3459	Omit 2nd *plume*
97	3464	*are private*
97	3466	Missing period at end of sentence
97	3466-3468	Sentence needs editing.
97	3469	Omit *Due to the higher metals concentrations during plume movement,*
97	3479	*…to identify wells, which due to their geographic setting or pumping history, may have had the potential…*
97	3486	Omit *characteristics of*
Fig. 9-1		I assume last two sentences in caption are notes to self?
97	3493	*This chapter emphasizes*
97	3494	*are provided in Appendix D*
98	3497	*deposits that snake*

Specific Observations on Main Document, Figures, and Tables		
Page	**Line**	**Comment or Question**
98	3512-3513	*high permeability sands and gravels and low permeability silts*
98	3514	*Fig. 9-3*. This and other insert Figure numbers are off.
Fig. 9-5		Caption give too much detail, given this is just an example, and this site from this project, which may be confusing.
98	3528-3529	*some from water flowing towards the stream from upland areas.* This was a steady analysis, so no change in aquifer storage.
98	3529	*that come from*
99	6553	the elevated metals signal soon after the river plume passed.
Fig. 9-10		*Only three wells (including 5 community wells)?* Make sure the wording makes it clear these are simulation results, not reality.
Fig. 9-11		*Flushing of the aquifer occurs in about*
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?
101	1620	*household wells*
101	3626	*water levels after the GKM release*
101	3627	Omit *the potential to have*
101	3651	Omit *using particle tracking the*
Fig. 9-16		*Flushing of the aquifer occurs in about*
102	3673	*Dissolved background zinc concentrations*
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?
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.
Fig. 9-17 and D-39		Use the same color in the legend for both dissolved and colloidal (e.g. both blue). At present it is confusing with river vs. well colors.
103	3719	Omit *(*
104	3752	*break-through*
104	3753	*at these wells closest to the river would be days to weeks*
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.

Specific Observations on Main Document, Figures, and Tables		
Page	**Line**	**Comment or Question**
105	3769	Omit *Animas River to the San Juan River*
105	3776	*the weeks during the release*
106	3801-3804	Grammar problems, needs editing
106	3812	*Gold King Mine event released*
106	3818	*entrained or dissolved metals mass in Cement Creek from previously deposited sediment.*
106	3824-3825	*varying by organism and exposure.*
106	3832	*due to dispersion.*
107	3862-3865	This discussion neglects the empirical derived deposition in the canyon reach (see Fig. 6-9a), which is more data-based than the WASP deposition pattern.
107	3874	*once it entered the San Juan River*
108	3916	*as the plume passed, to prevent*
108	3923	*of a variety of organisms*
109	3939	*exceeded* not *detected*
109	3943	*a large*
109	3952	*would likely dissolve*
109	3962	Omit *documented by monitoring during the plume period*
110	4015	Add that except for the one mid-Animas well, no metal concentration anomalies in well water were detected near the time of the release.
111	4023	*when the sediments*
111	4033	Omit *from*
111	4049	*It is not known whether this pattern persisted*

Specific Observations on Appendices			
Appendix	**Page**	**Line**	**Comment or Question**
D	2	23	*Animas River is gaining water*
D	4	81	*cumulative tributary stream flows upstream and the measured streamflow downstream…*
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.
D	5	96-97	*verifying locations of the sampled wells using hand-held GPS, and surveying well water levels* (keep the items in a list with parallel grammar)
D	5	101	*Animas River is losing water*
D	8	136-140	Omit discussion of FDM – most readers will know this, and if they don't they can easily look it up.
D	8	150	Omit *(water flows down the hydraulic potential gradient)*
D	8	158	*more coarse representations*
D	8	159-160	Omit these 2 sentences and Fig. D7 – too much detail. You've given plenty of references for the curious.
D	9	177	*baseflow compared to field observed baseflow*
D	9	181	*an individual pumping well.*
D	11		No need for Figure 8 – MODFLOW references are enough.
D	12	292-294	Awkward sentence should be edited.
D	16	370 and 372	*slide 12?*
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.
D	23	476	*map of simulated hydraulic head contours*
D	23	479	There is no Fig. D-16c
D	25	487	*the ditches in Farmington is shown*
D	Fig. D-20		Figure doesn't include lateral flows from rock into alluvium or irrigation flows lost to ET.
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.
D	30	552	*support*
D	31	557	*nominated* for what?
D	31	558-559	Explain what you mean by *sanitation wells*.

Specific Observations on Appendices			
Appendix	**Page**	**Line**	**Comment or Question**
D	32	578	*Animas River, Durango CO*
D	38	657, 659	Tau symbol is missing.
D	Fig. D-37		The minimum travel time would be to the top of the well screen. Perhaps start particles at the top of the well screen if that is possible.
D	44	718	*associated with water entering the lower half of the well screen* I assume the well used the lower two layers and maintained consistent head between levels? Please spell that out in the text.
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.
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).
D	47	777	*community well was perhaps*
D	48	794	*than was observed*
D	48	796	*colloidal forms.*
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.

Review by:

Glenn C. Miller, Ph.D.

Peer Review Comments on EPA's Draft Document "*Analysis of the Fate and Transport of Metals Released from the Gold King Mine in the Animas and San Juan Rivers*"

Glenn C. Miller, Ph.D.
University of Nevada
October 15, 2016

I. GENERAL IMPRESSIONS

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, both as a catastrophic failure, but also from continual smaller drainage.

II. RESPONSE TO CHARGE QUESTIONS

Part 1. Overall Project and Analysis

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

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 decade's 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.

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

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.

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

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. 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.

Part 2. Fate and Transport

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

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.

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

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.

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

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 the particulate sorption that promotes attachment to the particles.

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.

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.

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.

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.

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.

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.

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

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.

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

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 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.

Part 3. Application of Soft-ware Based Analytical Models

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.

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.

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.

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.

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 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.

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, 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.

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.

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.

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. 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.

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.

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.

III. SPECIFIC OBSERVATIONS

Specific Observations on Main Document, Figures, and Tables		
Page	**Line**	**Comment or Question**
9	8	"very low acidity" should be "very high acidity"
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

Specific Observations on Main Document, Figures, and Tables		
Page	**Line**	**Comment or Question**
		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)
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?
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.
9	41	What metals have increased toxicity in the particulate form? Are there any?
11	113	Add " or lower concentrations of calcium and sulfate"
13	208	Change "reducing" to "oxidizing" AMD is created when the highly reduced iron pyrite is oxidized to sulfuric acid and ferrous/ferric iron
14	258	Remove "oxidizing" Raising the pH changes the solubility, not the valence
21	492	Remove "of"
	1088	The term "mine waste" is correct. "ore" is an economic material, and since it was deposited outside the mine, it is waste.
45	1457	Remove the "e"
47	1558	"There was a very influx"???
49	1618	This should be "neutralizing", not oxidizing
90	3241; 3243	Subscripts, not superscripts.
104	3752	Break-through
108	3943	Remove "al"
Fig. 3-10		The legend is confusing. What is the "selected sample" Where was "cc" taken?
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?
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.
Fig. 6-1		Where is the figure?
Table 7-10		What are the units of time?
Figures 9-13, 9-14 and 9-16 to 9-18		All of these figures are very difficult to read. In fact, on my computers, many of the figures are very fuzzy.

Specific Observations on Appendices			
Appendix	**Page**	**Line**	**Comment or Question**
			[Reviewer provided no comment]

Review By:

Ronald L. Schmiermund, Ph.D.

Peer Review Comments on EPA's Draft Document "*Analysis of the Fate and Transport of Metals Released from the Gold King Mine in the Animas and San Juan Rivers*"

Ronald L. Schmiermund, Ph.D.
Economic & Environmental Geochemistry, Inc.
October 14, 2016

I. GENERAL IMPRESSIONS

<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.

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.

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.

<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 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.

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 model with its justification, and finally by the WASP model with clear objectives stated.

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.

Soundness of Conclusions – 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 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.

II. RESPONSE TO CHARGE QUESTIONS

Part 1. Overall Project and Analysis

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

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 …

It seems that each objective was addressed via extensive data analysis, although the analysis is not always clearly or extensively presented.

Specific Comments:

Should an additional objective be included? Specifically, identification of strategies for better preparing for monitoring future incidents? For example:

- Collecting site specific background samples prior to any work that might lead to a change in conditions (e.g., sample of GKM discharge on August 4 would have been useful). This would have mitigated the greatest problem with the GKM analysis.
- Guidelines for collecting samples after any change in discharge during an operation.

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

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 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.

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.

- 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 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.

  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.

- 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.

  Doubling the estimated GKM discharge concentration (sentences 1048 and 1049) to account for a "first flush" seems numerically arbitrary – please justify.

  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 like an autonomous debris flow with limited mixing with stream water.

- 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.

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

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 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.

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 for WASP. I assume the latter is to allow for a contaminant mass balance.

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.

- 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.

- 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).

Part 2. Fate and Transport

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

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.

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.

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

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 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.

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

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.

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.

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.

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 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.).

Table 8.6 addresses some pre- and post-event dissolved and total metal concentration statistics. 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.
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?

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.

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.

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 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?

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.

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'?

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-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.

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).

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.

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 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?

- 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?

- 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.

- 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 dolomite. I was pleased to see the updating of the conventional assumption for atmospheric log CO_2 fugacity to -3.4 from -3.5.

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

No comment

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

No comment

Part 3. Application of Soft-ware Based Analytical Models

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.

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:

- 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.

- 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.

- 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.

- 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.

- 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.

- Question 7 – "Clay" only appears 3 times in the final draft, so I don't think this recommendation has been adequately addressed.

- 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.

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.

No comment

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.

No comment

III. SPECIFIC OBSERVATIONS

Specific Observations on Main Document, Figures and Tables		
Page	**Line**	**Comment or Question**
General		Given the importance of "dissolved" versus "colloidal/particulate", the word "metal(s)" should always be preceded by a modifier for clarity
8	Fig 1-6, line 5	… where basic (i.e., alkaline) …
9	8	Replace "low acidity" with either "high acidity" or "low pH"
14	246-247	Mine-waste rock has not been pulverized to remove sulfides – only applies to tailings
14	250-251	Replace "low to moderate acidity" with either "high to moderate acidity" or "low to moderate pH"
14	265-259	Dissolved metal concentrations are *generally* suppressed… Increasing pH *allows* oxidizing. Secondary iron sulfate minerals exist and form in *acidic* conditions due to oxidation typically accompanied by evaporation.
14	259	Spelling "oxyhydroxides"
15	263-266	Aluminum precipitates are white and may exist in the absence of iron
15	268	By definition "colloids" will never be lost from the water column, especially in moving water. Only after sufficient aggregation/flocculation, typically in response to changes in ambient chemistry and/or time, can they be lost from the water column.
15	273-283	Please clarify the differences (composition, stability and distribution) between "waste rock/dumps" and "tailings/tailings ponds or piles"
15	299-300	Clarify that GKM is one of the 80 mines mentioned above

Specific Observations on Main Document, Figures and Tables		
Page	**Line**	**Comment or Question**
16	327	Remove "of" Clarify " … the same geochemical reactions (of) routinely observed near Silverton …"
20	453	Close quote after *data providers*
Table 2-1	footnote	Define "SADIE"
21	502	Reword sentence …"potential risk to contaminants …"
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 table. There should be a reference to the SOPs for each entity.
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).
Fig. 2.5		Include "RCWWN" in list of abbreviations
27	708-714	I approve of the acknowledgement that 0.45 µm is neither natural nor effective, but I think, having made the acknowledgement, that a reason for using that convention should be offered and an explanation of the consequences provided.
27	718	Sp. "ware"
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?
27	724-725	"latter" should be "later". The sentence implies that the only mechanism for subsequent metal mobilization is re-entrainment of settled solids and does not acknowledge desorption.
27	730	For clarity modify the sentence "….. throughout the analysis: one based on contaminant concentration and one based on contaminant mass."

Specific Observations on Main Document, Figures and Tables		
Page	**Line**	**Comment or Question**
27	737	Should read "… concentration (expressed as mass of contaminant per unit volume of water or unit mass of sediment) …"
27	739-752	These paragraphs are difficult to follow and interpret, and might be taken by the public to be obfuscation. I personally 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"?
29		Include citation for "BOR 2015" and "EPA 2016"
Fig. 3-3	Caption last line	"A much *smaller?* flood wave ……."
Fig. 3-5	caption	Last sentence is a fragment.
Fig. 3-8		Seems that iron should be included in the plots (secondary Y axis)?
31	926-927	"…. pool re-established, but that chemistry *would?* return …. This expectation is likely to depend on the degree of aeration/oxidation that becomes established after the blowout.
Table 3-1	Caption, last line	Sentence fragment
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.
Fig. 3-14	caption	State the location for this data (USGS gage station 12.5 km. downstream)
33	1004	"concentration" used twice. Should also clarify that "colloidal/particulate" concentrations are being used.
33	1009	Where is the equation?
37	1100	"…. acidic water, dissolved *and suspended* metals …."
Fig. 4-7	Plot	Does the red triangle signify something else?
Fig. 4-2	caption	"with elevated" is repeated '… traveled as *a* coherent mass …"
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.
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?
Fig. 4-5A		No units on vertical axis
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
39	1198-1200	The conductivity as measured by the sondes does not necessarily confirm that metal concentrations were behaving consistently between sampling points, only that the combined effect of sulfate and

Specific Observations on Main Document, Figures and Tables		
Page	**Line**	**Comment or Question**
		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.
40	1246	Should be Fig. 4-7
40	1247-1248	Reword sentence after "(SUIT)"
Fig. 4-8	citation	2nd line, "fitting" is repeated 3rd line, extraneous "(A)" Site should be identified. Please plot water volume across entire plot.
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.
41	1309	Site 09358550 is 0.72+ miles upstream of the confluence – hardly "just upstream"
41	1304	"is" should be "are"
Fig. C-9		I'm quite familiar with Geochemist's Workbench, but I can't follow the figure caption
Fig. C-10		Explain differences between plots
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 interpret if paired with a plot of important aqueous concentrations.
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.
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
107	3853	'acidity' is an intensive, not an extensive quantity, quantity (no volume association)
40	1244-1245	Explain "basin-scale relationship"
Fig. 6-1		missing
Appendix F		missing

Review by:

Mark A. Williamson, Ph.D.

Peer Review Comments on EPA's Draft Document
"*Analysis of the Fate and Transport of Metals Released from the Gold King Mine in the Animas and San Juan Rivers*"

Mark A. Williamson, Ph.D.
Geochemical Solutions, Loveland, Colorado
October 14, 2016

I. GENERAL IMPRESSIONS

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.

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.

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.

II. RESPONSE TO CHARGE QUESTIONS

Part 1. Overall Project and Analysis

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

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.

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.

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

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.

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).

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

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.

Part 2. Fate and Transport

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

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.

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 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.

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

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.

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 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.

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

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.

The GKM discharge and lost plume volume notwithstanding, the discussion of uncertainty and the 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?

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.

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.

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.

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.

As with charge question #8 above, I find no particular concerns with the presentation of metal concentrations post-plume.

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.

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.

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

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 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.

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

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.

Part 3. Application of Soft-ware Based Analytical Models

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.

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.

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.

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.

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.

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.

III. SPECIFIC OBSERVATIONS

Specific Observations on Main Document, Figures, and Tables		
Page	**Line**	**Comment or Question**
		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.

Specific Observations on Appendices			
Appendix	**Page**	**Line**	**Comment or Question**
			[Reviewer provided no comment]

Appendix A

Curricula Vita