United States
Environmental Protection
Agency

EPA/600/R-16/113 | June 2016 | www.epa.gov/ord

EPA Gold King Mine Analysis of Fate and Transport in the Animas and San Juan Rivers

Response to Peer Review Comments

Office of Research and Development
National Exposure Research Laboratory

Response to Peer Review Comments of EPA's

Gold King Mine Analysis of Fate and Transport in the Animas and San Juan Rivers

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

July 7, 2016

Disclaimer

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

Table of Contents

TABLE OF CONTENTS

I. Introduction 4

II. CHARGE TO REVIEWERS 6

III. GENERAL IMPRESSIONS 8

IV. RESPONSE TO CHARGE QUESTIONS 16

Part 1: Overall Project and Analysis 16
- Question 2 21
- Question 3 26
- Question 4 28

Part 2: Fate and Transport 31
- Question 5 31
- Question 6 39
- Question 7 44
- Question 8 49

Part 3: Geochemistry 50
- Question 9 50
- Question 10 51
- Question 11 52

Part 4: Water Quality Analysis Simulation (WASP) Modeling 53
- Question 12 53
- Question 13 63
- Question 14 68

Part 5: Groundwater Modeling 70
- Question 15 70
- Question 16 81
- Question 17 84
- Question 18 86

Part 6: EnviroAtlas Modeling 88
- Question 19 88
- Question 20 90

Part 7: Bioaccumulation 91
- Question 21 91

I. INTRODUCTION

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

A mid-project peer review was managed by Versar, Inc., an independent contractor, under contract No. EP-C-12-045 Task Order 80. Versar was tasked by EPA to coordinate an external peer review of EPA's project entitled Gold King Mine (GKM) Analysis of Fate and Transport in the Animas and San Juan River. The purpose of the three-day peer meeting, held at the EPA's Office of Research and Development (ORD) Laboratory in Athens, Georgia on February 23-26, 2016, was for five expert reviewers to evaluate the scientific integrity of EPA's analysis and characterization of the fate, transport, and potential impacts of acid mine drainage (AMD) release in the Animas and San Juan Rivers. The reviewers met with EPA scientists who presented their analysis and findings to the reviewers via Power Point™ presentations that had been provided prior to the meeting. The peer review process provided a documented, independent, and critical review of the draft analysis, and its purpose was to identify any problems, errors, or necessary improvements to the analysis prior to being published or otherwise released as a final assessment. Project findings included in a final report will also be independently peer reviewed.

Versar was charged with assembling the peer reviewers and coordinating the peer review. Versar evaluated the qualifications of peer review candidates, conducted a thorough conflict of interest (COI) screening process, independently selected the five peer reviewers, distributed review materials, managed the written peer review period, organized and hosted the peer review meeting. Versar identified candidate reviewers with expertise in the following areas: (1) geochemistry, (2) fate and transport (water/sediment), (3) water quality analysis simulation (WASP) modeling, (4) groundwater modeling, (5) geospatial analysis (EnviroAtlas modeling), and (6) bioaccumulation. Versar's in-depth and multi-staged evaluation of qualifications was based on each candidate's biosketch, curriculum vitae (CV), and publications.

Peer reviewers included:

Brian S. Caruso, Ph.D., P.E.. U.S. Geological Survey, Denver, CO
Charles R. Fitts, Ph.D., Fitts Geosolutions, LLC. Scarborough, ME
Henk M. Haitjema, Ph.D., Haitjema Consulting, Inc., Bloomington, IN
D. Kirk Nordstrom, Ph.D., U.S. Geological Survey, Boulder, CO
William A. Stubblefield, Ph.D., Chair, Oregon State University, Corvallis, OR

Versar developed a final peer review report that summarized the peer review comments provided during the meeting and presents the reviewers' individual written comments in response to a series of charge questions pertaining to hydrology, geochemistry, fate and transport, and potential impacts from the Gold King Mine release (Versar, Mar 9 2016).

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

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

Kate Sullivan, Ph.D
Christopher Knightes, Ph.D.
John Washington, Ph.D.
Mike Cyterski, Ph.D.
Steven Kraemer, Ph.D.
M. Craig Barber, Ph.D.
Anne Neale
Lourdes Prieto

II. CHARGE TO REVIEWERS

Part 1. Overall Project and Analysis

Question 1. Given the data that were available to the researchers at the time, were assumptions about data inclusion, formatting, and use appropriate? How so?

Question 2. Was the overall integration process of the various analyses conducted in a way that provided meaningful results and conclusions? Please explain.

Question 3. When looking at the full project, are there errors or gaps in the integration process that could have affected the overall analyses and/or the conclusions? Please explain.

Question 4. Were the overall conclusions that were drawn from these analyses appropriate and scientifically defensible based on the analysis? Why or why not?

Part 2. Fate and Transport

Question 5. Does the research appropriately characterize the metals concentrations and load produced at the Gold King Mine spill?

Question 6. The concentration of metals near the release site in the receiving waters had to be estimated from samples collected after the much of the plume had passed. Were the estimates of metals concentration at this location appropriately calculated through scientifically sound methods using available data?

Question 7. Were the data analyzed and visualized properly in regards to sediment metal concentrations in the post-plume period in Cement Creek and the Animas River?

Question 8. Were the data analyzed and visualized properly in regards to sediment metal concentrations in the post-plume period in Cement Creek and the San Juan River after receiving mine contaminated water from the Animas River?

Part 3. Geochemistry

Question 9. Were the geochemical principles to characterize transport and fate of acid mine drainage appropriately applied and interpreted? Please explain.

Question 10. Were precipitation and mineral saturation analyses of the acid mine drainage appropriately applied for interpreting metals fate in the river system? Please explain.

Question 11. Was the neutralization of acid mine drainage and subsequent fate of dissolved and colloidal/particulate metals appropriately interpreted? Why or why not?

Part 4. Water Quality Analysis Simulation (WASP) Modeling

Question 12. Did the WASP modeling appropriately apply modeling parameters to estimate the movement of plume water? Please explain.

Question 13. Did the application of assumptions and values in WASP modeling appropriately address particle transport and deposition of the acid mine drainage constituents? Please explain.

Question 14. Did the WASP modeling appropriately investigate the remobilization of metals during increased flow? Why or why not?

Part 5. Groundwater Modeling

Question 15. Is the analysis as presented sufficient to evaluate the potential for impact of the acid mine release from the GKM on pumping wells located in the floodplain aquifers downstream of the spill?

Question 16. Were the assumptions informing the choice and construction of the groundwater flow model appropriate for the intended use? Please explain.

Question 17. Were the assumptions informing the capture zone and particle tracking analysis appropriate for the intended use? How so?

Question 18. Did the method for calibration of the local scale groundwater flow model performance to the observed drawdown reported in the driller's log serve as an effective method? Please explain.

Part 6. Atlas Modeling

Question 19. Are the sources of the data included in the maps valid, complete, and adequately documented? Are there any points of confusion, gaps, or suggestions for improvement?

Question 20. Do all of the maps and charts communicate the analysis methods and results in such a way as to be readily understood by stakeholders with interest in the impacts of the Gold King Mine spill (e.g., First Nations; NGO's; news media; and State water, recreation, public health, and wildlife managers)? Are there points of confusion, gaps or suggestions for improvement?

Part 7. Bioaccumulation

Question 21. Given the limitations of the BASS model, how appropriate is the simulation of bioaccumulation of As, Cd, Cu, Pb, and Zn in the Animas River trout fishery? What are the strengths and weaknesses of using this approach?

III. GENERAL IMPRESSIONS

General Impressions		
Reviewer Name	**Reviewer Comment**	**EPA Response**
Brian Caruso	1) I commend EPA for gathering and analyzing all of this data in an attempt to understand the contaminant plume movement in the Animas and San Juan rivers from the Gold King Mine release. It is always challenging to collect and analyze in a consistent way existing data from a wide range of sources and with different levels of QA/QC. EPA has done a relatively good job in a short time frame at a first cut for this fate and transport analysis. However, the accuracy of information presented is questionable due to a number of reasons and assumptions, the clarity of presentation needs improvement, and the soundness of conclusions is also drawn into question based on these issues. One of the main issues is that the goal of the research appears to be too broad and not specific enough to determine if the information and conclusions are adequate. In some cases the goals and objectives are not entirely clear and appear to be somewhat different in various places in the presentation where they are presented. 2) In general, I believe that EPA should perform this work and prepare the research analysis so that it uses the best science available and presents results as clearly as possible in preparation for a number of issues, including potential lawsuits and Superfund investigations, monitoring plan development, and to inform all stakeholders of what occurred as best as possible. Although many of the conclusions seem generally appropriate based on the analyses performed, the quantification and accuracy of the conclusions are weak due to a great deal of missing information and lack of detailed uncertainty and sensitivity analysis.	1) The primary objective of EPA/ORD's Gold King Mine study was to quantify the amount and characteristics of AMD produced in the GKM release and its transport and fate in the Animas and San Juan Rivers during the plume and in the immediate period following the initial event. The scope of the project is specific and supported by data collected by EPA, states, and tribes during and following the event. Each analysis within the project supports the overall objectives with specific objectives. We will further clarify objectives and integration of analyses in the final report. 2) A basic principle of EPA is to conduct the highest quality science and to ensure this through peer review. This mid-project peer review contributes to meeting those goals. We will revise the organization of materials and visualization of results to more clearly communicate study objectives, methods, and findings in the final report. We will also include uncertainty and sensitivity analysis.

General Impressions		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	3) Several examples of where the analysis and presentation should be improved include: • better definition of goals and objectives to reflect critical information needs • use of EPA national criteria or standards for metals for drinking water and aquatic life as part of an initial screening of risk to select potential contaminants of concern for more detailed analysis and as indicators (instead of primarily evaluating total metals) • better use of other existing data and information from previous investigations to evaluate and help confirm background (pre-release) levels for comparison • inclusion of additional data and information for better reactive transport modeling, metals concentration and load calibration, and validation for WASP • better evaluation and presentation of uncertainty and sensitivity analysis of results should identify data gaps in the analysis and for future modeling	3). Our revised materials and report will include a clear expression of goals and objectives. In addition, the final report will include more detailed treatment of the topics identified by the reviewer, including: evaluation of the plume from an exposure perspective; analysis of pre-and post-GKM event water and sediment concentrations, additional data analysis for calibrating WASP (see questions 12 – 14). We will also increase discussion of uncertainty, model sensitivity, and data gaps.
Charles Fitts	4) It is hard to summarize since there are so many facets of these studies. The soundness of conclusions is discussed under question 4 below. There are many details that need attention, and many of these are just a matter of editing, polishing and fleshing out with more text and detail, which is to be expected in a more final draft. I felt that the overview and empirical analysis sections were generally logical and needed minor work. I have few comments on the geochemical and bioaccumulation portions since I have less background in those areas. The WASP presentation could use a good deal of clarification about the analyses and more caveats about the uncertainties involved and how the results may be used.	4) As part of the WASP calibration activities we have significantly improved the ability of the model to perform deposition/suspension analysis that reasonably reflects the observed data and that squares with published data. See responses to questions 12-14 for more discussion of these points. We have included more sophisticated 3D groundwater modeling to augment the 2D modeling that was reviewed by the panel. We have also gathered additional data for a greater number of

General Impressions		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	For example, the deposition/suspension analysis of WASP slides 25-27 seems to be quite uncertain and should be viewed skeptically since the deposition/suspension input parameters do not square with published ranges. In most cases, the WhAEM modeling was sufficient to characterize whether a well likely pulls in some river water, but the modeling approach was not sophisticated enough to predict accurately what fraction of a well's flow came from the river and what the plume breakthrough curve might look like in well concentrations. More sophisticated 3D and localized models could be constructed to improve predictions, but the Key Analysis Question (Groundwater slide 2) may not require such detail for most of the wells.	wells in the vicinity of the modeled wells to enhance calibration to river and well water levels. These responses are described in greater detail in a following section (Part 5 questions 15-18) responding to specific feedback on groundwater modeling.
Henk Haitjema	5) The overall goal of the research has been the topic of some mild confusion by me (and the group at large). The agency stressed that the current research does not constitute a formal risk assessment nor was it designed as such. However, the precise purpose of the research has not been articulated very clearly. I must assume that in the end the research presented is to be used as the basis for some form of risk assessment and, if needed, remedial action. As such I have been evaluating the research presented with this ultimate goal in mind. 6) Overall the work was well presented although the complexity of some issues and the necessary brevity of the presentations resulted in many interruptions of the presentations with questions or requests for clarifications by the reviewers. While I understand that the EPA researchers could only work with publically available data, it was observed by several reviewers that some important historic (background) data were missing, but might have been acquired from public sources (e.g. the USGS).	5) The project objectives were to quantify the volume and type of metals introduced into the Animas River system from the Gold King Mine release, the water quality characteristics during transport, and the fate of the metals in the receiving rivers. The final report will present results including the magnitude and duration of potential exposure of water users and assessment of possible future effects. 6) The EPA team has acquired publicly available historic data from the USGS and EPA STORET databases as well as data available from states and tribes to enable an expanded comparison of metals concentrations pre- and post-event in surface water and sediment.

General Impressions		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	7) Most conclusions seemed reasonable, taken into account the limited data and the basic nature of these initial studies. However, in several cases the data uncertainty could have been better alleviated with some sensitivity analyses and by presenting bracketing solutions showing both most favorable and most unfavorable (worst case) scenarios. 8) Finally, I have the impression that communication between different branches of EPA is less than optimal. On several occasions the quality of the studies suffered as a result. For example, the lack of coordination between the various sampling efforts and the lack of information about the sampling and quality assurance protocols cast some doubt on data integrity, hampers data comparisons, and may have resulted in unnecessary data gaps.	7). The final report will include increased treatment and presentation of uncertainty and sensitivity for empirical and modeling analyses. 8). The EPA/ORD team utilizes data developed by EPA Regions 8, 6, and 9 that were collected and processed following Quality Assurance Project Plans and SOPs that document sampling protocols, testing, and quality assurance. We also included data from the USGS, the states of Colorado, New Mexico, and Utah as well as the Southern Ute Indian Tribe and the Navajo Nation who have similar QA/QC related documentation.
Kirk Nordstrom	9) The Animas River Team (ART) of the EPA's Office of Research and Development (ORD) involved with research on the fate and transport of potential contaminants from the Gold King Mine (GKM) spill presented, summarized, and interpreted a very large set of diverse data collected by EPA and other technical groups under adverse conditions. Although the data set was large, many necessary parameters were missing and the quality was less than optimal for the objectives of the ART because the accidental release was unexpected and field and lab parameters were collected while the EPA was in an emergency response mode with little time for planning. Hence, the ART was working under a serious handicap and with very tight time constraints. Considering this overall situation, the presentations were impressive. They have made every effort to be thorough in collecting information, careful in most of their decisions on how to proceed with insufficient data, and they have been clear on what information is	9). See response (8). The EPA/ORD team has utilized available data to the fullest extent possible. Although there are some data gaps, such as several geochemistry-related parameters not routinely measured in field sampling in that would have enhanced geochemical analysis, the EPA/Ord team believes that the cumulative data set compiled from multiple agencies allows a robust analysis of pre-, during, and post event metals contamination by the Gold King Mine event in the 600 km of river sufficient to meet the objectives of this project which was to understand the source, transport and fate of the GKM plume metals. The source and use of data in analyses will be fully described in the final report.

General Impressions		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	based on fact, what assumptions were used, what aspects were largely speculative and require follow-up monitoring, and they have reviewed and revised many of their conclusions to keep them as sound as possible. They have been transparent about what they have tried to do and completely open to good suggestions. We had, in my opinion, excellent discussions about what can and cannot be done with the available data. 10) That's not to say that there isn't room for improvement. To be sure, some of their assumptions could use revision, some of the methods that were used need modification, and in one instance (bioaccumulation) the effort was highly questionable. Having an independent review to evaluate the work at this point was a wise choice. As long as the recommendations of the reviewers are carefully considered, this mid-point evaluation should prove extremely valuable in helping the ART to achieve its goal. 11) Some of the figures in the presentations were impossible to read either in the hard copy or in the various PowerPoint presentations. These should have been checked and improved. 12) A more logical and consistent sequence to the presentations would have helped also. A more helpful logical and consistent sequence means a clear statement of goals followed by an outline of available data with a tabulation of the logic on how to obtain said goals. Some of this was presented but it was a bit different for each group and the methodology was not always clearly stated. 13) It is difficult to appropriately characterize the metals concentrations and loads when a lot of the important field and lab data were not collected. Immediate field reconnaissance was	10) The EPA/ORD team agrees that that this mid-project review by independent experts was very beneficial. Reviewers' comments helped us improve our methods and presentation of information. Individual subject experts agree that the peer review team was deeply insightful in the scientific areas covered in this project and we appreciate the time and effort they gave us and the in-depth and lively discussions of our work. The review comments have strongly influenced the trajectory and details of all of the analyses and subsequent presentation of data that will be included in the final report. These will be further highlighted in response to specific comments that follow. 11). Figures and data presentation will be carefully prepared in the final report for clarity and lack of errors. 12). The final report will be organized to integrate findings along topic areas. The peer review was organized to review methods. We will include a table of approaches and other means do display the individual study elements and their integration. 13) When collecting water samples that were analyzed in the laboratory for dissolved and total metals, EPA field crews routinely measured water temperature, pH, specific conductance and dissolved oxygen.

General Impressions		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	challenging because of the unexpected accidental and sudden release of mine pool water, the time delay in notifying authorities of the accident, and the time delay in getting personnel and equipment to the field. Of course, under rapid emergency conditions it is difficult to collect enough of the right kind of data. However, it is hard to understand why more field parameters were not measured such as conductivity, pH, and temperature for all samples, why sulfate and Fe(II/III) were not determined when water samples were collected for analyses, and why no samples of GKM effluent were collected during the release. These parameters (pH, conductivity, and Fe(II/III) should always be measured for acid mine water contamination. This is not a criticism of the ART modeling efforts, obviously, but of the lack of guidelines for the field personnel who collected the samples. 14) The EPA should have a handbook that recommends what samples and field parameters need to be collected in an emergency mine water spill. Furthermore, the handbook should emphasize the importance of getting water samples of the source water (the Gold King Mine effluent) as soon as possible and throughout the main pulse of mine water release because its chemical composition could, and probably did, change during the release. It is imperative that the chemical composition of the pollutant source be properly characterized because substantial changes in its composition can occur and will affect downstream transport. If the source is not well characterized then it becomes extremely difficult for the team to characterize the changing conditions of the plume as it moves hundreds of kilometers downstream. If the proper parameters had been collected, the ART could have done far better at characterizing the metals and the load, the rate of movement of the plume, the partitioning of metals between dissolved and	14) The EPA/ORD team agrees that additional data would be useful for sophisticated geochemical analysis of an AMD plume such as occurred with Gold King Mine. These would include (in generally decreasing order of importance): i;) for field pHs above 4, simple field alkalinity titration; ii) dissolved Fe speciation, ferrous vs. ferric; iii) anions, especially $SO_4^{=}$; iv) for pHs less than 5, laboratory acidity, with the sample pre-oxidized with H_2O_2 and titrated over heat to accelerate reactions and SEM/AVS extraction as a metric of potential toxicity to aquatic organisms. We note that a limited amount of this type of geochemically-relevant data such as sulfate was collected by the states of Colorado, New Mexico and Utah as well as the Southern Ute Indian Tribe that were used in the analysis.

General Impressions		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	particulate forms, and the fate of the metals in the plume. What the Team did manage to do with this partial data set is highly commendable, appropriate, and the results were very reasonable. More on this below. 15) A general rule of thumb is that anybody trying to model the hydrogeochemical dynamics of a field site needs to see the field site. By visiting the sites, the team can get a much better idea of how appropriate their modeling and assumptions are for the goals of the project. I was surprised that no one had been allowed to see the area or had ever visited the area. A good field observer has a natural feel for how to model a complex and transient event with limited data. This disconnect between field and modeling effort can lead to inappropriate analyses and conclusions.	15). The EPA/ORD team agrees that direct field observation helps researchers gain insight when conducting analysis. However, the research team was able to use resources to help fill the gap. Members of the EPA/ORD team have had extensive field experience with AMD and hydrology and water quality measurement as well as with use of the models we deployed. We consulted with EPA regions onsite and were informed by other researchers and agency reports who conducted onsite sampling We were able to use remote sensing imagery to view the entire system. There were abundant photographs, news videos and images that provided visual impressions that were as good for many purposes as direct observation.
William Stubblefield	16) The presentations provided in the *Gold King Mine Analysis of Fate and Transport in the Animus and San Juan Rivers* reflected the high degree of effort and quality expended in their preparation. However, the overall objectives of the effort, technical approaches employed, and desired outcomes were not obvious. EPA NERL scientists were clearly at a disadvantage not having been involved in the design of the sampling plan, its implementation, and the assessment of the overall quality of the data. Two analysis objectives were stated in the overview presentation: • Characterize the release, transport and fate of the approximately 3 million gallons of released AMD, with a focus on a suite of metals • Identify the potential for water quality impacts, including municipal wells, and implications for future monitoring priorities.	16) The primary objective of ORD's Gold King Mine study was to quantify the amount and characteristics of AMD produced in the GKM release, and its transport and fate in the Animas and San Juan Rivers during the plume and in the immediate period following the initial event. We considered the potential for groundwater routes of exposure. We also attempted to provide some insight on what might happen to metals sequestered in the system in future runoff events. This analyses has been expanded since the peer review and will be included in the final report.

General Impressions		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	17) Clearly, a great deal of effort went into addressing the first objective and EPA scientists did a reasonable job of achieving this objective, given the limitations in data and the rapid nature of the response. It is not as clear how the second objective was to be addressed. Prior to the review it was explained that this was not an "ecological risk assessment;" however, to be able to address the second objective it is imperative that environmental exposures for individual metals be adequately described in terms of their magnitude and duration, as a minimum. Given the current state-of-the-science it would also be helpful to have information regarding those physicochemical parameters that can affect the toxicity of individual metals to aquatic organisms (e.g., dissolved organic carbon, pH, and hardness). It was also noted that there was a reasonable set of sediment data analyses (300 samples) but no detailed analysis of this data was presented. It was acknowledged that there is a large amount of data available and that the integration and interpretation of the data represents an onerous task, especially given the rapid analysis time available. 18) In conclusion, it was somewhat difficult to discern what the objectives of the integrated program were and whether they had been achieved or not. There seemed to be a lack of cohesiveness in the overall program objectives and the approaches taken to achieve these objectives.	17) The final report will contain information on the magnitude and duration of environmental exposures during the movement of the Gold King Mine plume through the affected rivers. The final report will contain the available data on physico-chemical parameters including hardness and pH that were collected during plume travel. There are no known organic carbon data collected during the event. The final report will also present data and analysis on metals concentrations in streambed sediments and the effect of the plume on them. 18) See responses (5) and (16). The final report will be organized to integrate findings along topic areas. We will include a table of approaches and other means to display the individual study elements and their integration.

IV. RESPONSE TO CHARGE QUESTIONS

Part 1: Overall Project and Analysis

Reviewer Name	Reviewer Comment	EPA Response
Brian Caruso	19) Some assumptions about data inclusion, formatting, and use were appropriate, some were not, and some were questionable. There appear to be many questions and issues with regard to the analysis methods and assumptions, many of which affect our evaluation of the assumptions about data inclusion, formatting, and use. Important questions and issues include: • a) The goals of the fate and transport analysis and modeling are not clear, and in some cases appear to be different in various parts of the presentation materials. • b) It appears that the WASP TOXI model for toxicants, including metals, was not used. This module incorporates Kd values for partitioning between dissolved and particulate forms, 1st order decay, and diffusion coefficients, for some reactive transport modeling. Also, why was the WASP add-on, Metals Transformation and Assessment (META4), not used for the fate and transport modeling? This module was developed by EPA and can handle reactive transport in complex acid mine drainage-metals systems with precipitation-dissolution reactions incorporating pH and other important parameters.	19) a) The project objectives were to quantify the volume and type of metals introduced into the Animas River system from the Gold King Mine release, the water quality characteristics during transport, and the fate of the metals in the receiving rivers. Each of the analytical methodologies reviewed individually had a more specific way of framing that charge depending on the subelement and nature of the method. The final report will ensure clarity on project objectives and scope. b) We used the WASP TOXI module for the fate and transport simulations. This was not made specifically clear in our presentations. We state this explicitly now. We have updated our model formulation following discussions with the reviewers. We have incorporated diffusion of dissolved species between water column and pore water and a lumped parameter Kd for individual metals for partitioning to and formation of particulates. We also have used the empirical estimates of total loads at selected locations to estimate constant settling velocities. We reached out to Dr. James Martin, one of the WASP architects, to ask about META4. META4 was used on mine sites in CO by Medine in the 1990s, which is based on

Reviewer Name	Reviewer Comment	EPA Response
	• c) Important or indicator individual metals should be analyzed and presented in more detail. These should probably include at least Cd, Cu, Pb, and Zn. Summary statistics of data should be calculated at time periods along the length of rivers and compared to EPA drinking water and aquatic life hardness-based criteria to evaluate potential contaminants of concern for fate and transport analysis and initial screening for potential risk. • d) Why did EPA not use Sondes for continuous monitoring of parameters such as conductivity and pH? • e) Why was pH and conductivity not measured in many samples? • f) Why were different sampling and analysis methods and detections limits used by different EPA organizations and for different samples?	WASP4. To use META4, the analyst would need to know META4 as well as WASP4. We are currently using WASP7, which has had many upgrades and updates since WASP4. META4 was never officially released. The routines available in WASP7, such as the new hydrology routines, which allows for the kinematic wave modeling for streams/rivers or the interface, are not be available in WASP4. An update to META4 was done in 2012, though it has never been fully implemented into WASP, and is not currently available or usable. Dr. Martin recommended that META4 not be used until it is finally incorporated into the updated WASP, and then only if there is good water chemistry data. c) The final report will include analysis of concentrations and mass of individual medals including at a minimum Cd, Cu, Pb, and Zn at various points along the rivers. These concentrations will be evaluated against various metals-related water quality criteria. d) Continuous sonde data on pH and specific conductance was collected by the state of New Mexico and the Southern Ute Indian Tribe at several locations in the lower Animas River and near Farmington in the San Juan River. The EPA/ORD team used these data in establishing plume movement.

Reviewer Name	Reviewer Comment	EPA Response
		e) EPA crews measured specific conductivity and pH when collecting samples. These data are included in analyses. f). EPA samples were collected and laboratory tested according following Quality Assurance Project Plans in place in each Region. Laboratory testing was consolidated under one national laboratory several days into the GKM release, which accomplished standardization of reporting results.
Charles Fitts	20) There was some discussion about other possible sources of data from academics and other organizations. If there exist other data particularly at an earlier time near the GKM or Cement Creek, it would be helpful to get that data and include it in the analysis, since it would reduce the uncertainty about the source concentrations and mass.	20) The EPA/ORD team has obtained pre-event data from the USGS and EPA STORET data bases and all available data pre-, during, and post the GKM event from the states of Colorado, New Mexico, and Utah as well as the Southern Ute Indian Tribe, the Mountain Ute Tribe, and the Navajo Nation. Should additional data become available from sources with quality assurance documentation, we will strive to incorporate them into our analysis.
Henk Haitjema	21) In some cases data sources and limitations were not fully explained and required reviewer inquiries. While in most cases an appropriate attempt was made to overcome data scarcity and uncertainty by offering conservative (worst case) scenarios, these were not always well explained.	21) The EPA/ORD team presented a discussion of similarities and differences among data sets obtained from various agencies and cited some of the potential uncertainties associated with these data as well as gaps in available data. In the final report, we will pay close attention to highlighting and explaining important gaps and uncertainties in data and analyses.
Kirk Nordstrom	22) For the most part, the data that were available were properly included and appropriate. There is the distinct possibility that additional data was collected by university researchers, local stakeholders (such as the Animas Stakeholders Group), mine owners, the US Geological	22) We have acquired data collected by the U.S. Geological Survey (USGS) during the GKM plume event, as well as relevant historic data collected in the Animas and San Juan Rivers and used it in analysis. Our analysis has utilized data collected by

Reviewer Name	Reviewer Comment	EPA Response
	Survey (USGS), The Bureau of Land Management (BLM), and the US Forest Service that has not yet been discovered. For example, I am aware that some data was collected by the USGS which has not been included in the compilation and the presentations. These additional data sources, which included USGS data given to Steve Way of the EPA should be found and included if useful for the modeling. 23) It would have been helpful for me to have the team include chemical analyses of just a few waters samples such as GKM effluent in addition to the samples that defined the tail end of the plume. Then I could do some quick calculations to both confirm what the team had calculated and to see if there are any additional calculations that might need to be considered. The reviewers only saw a graph of a limited number of constituents. 24) Several of the plots were log plots that gave a strange symmetry to the data. I know that in many cases there is such a large range of values that a log plot is necessary but not in all cases. Log plots often make the data look better than it really is. I would suggest that some plots could be divided into 2 or 3 linear plots for better visualization.	EPA, states, tribes, and other organizations who have followed documented quality control procedures. Should additional data become available we will to attempt to incorporate it into our analyses. 23) The GKM effluent concentration data will be included in the final report or its appendices along with an explanation of which values were selected to represent the plume and why. 24) The EPA/ORD team uses log-scaled plots to display data during plume passage that typically spanned 3 orders of magnitude and sometimes ranged over 5 to 6 orders of magnitude over the length of the affected rivers. In the final report we will design graphical presentation of results to use the minimal scales possible with the least distortion.

Reviewer Name	Reviewer Comment	EPA Response
William Stubblefield	25) The scope and types of available data were adequately described and the limitations of the available data were also discussed. Obviously, there were limitations in the available data and in some cases key parameters that would have been useful for interpreting data were not available (e.g., dissolved organic carbon). The staff doing the analyses had to "make do" with the extant data and they seemed to do an adequate job with what was provided. 26) In some cases, questions were raised regarding the potential availability of data from other non-EPA sources that might exist. EPA is encouraged to seek out and obtain all potential data that would be useful in interpreting the extant data. Potential data sources that should be examined include the USGS and State Department of Environmental Quality and/or Departments of Fish and Wildlife. In addition, it is anticipated that there may be data held by researchers at local Universities and at various Native American organizations.	25) Although there is considerable data on metals in water and bed sediments, there are very little data on dissolved organic carbon during the plume, as well as some other parameters that would have facilitated geochemical analysis. However, we feel that a robust analysis addressing the project objectives can be accomplished with the available data. 26) The EPA/ORD team has acquired data from EPA, the states of Colorado, New Mexico, and Utah, and the Southern Ute Indian Tribe, the Ute Mountain Tribe, and the Navajo Nation. We have also acquired data collected by the US Geological Survey (USGS) during the GKM plume event, as well as their historic data. We continue to acquire data and reports by various entities as they are released. Should additional data become available we will strive to incorporate them into our analysis.

Question 2 Was the overall integration process of the various analyses conducted in a way that provided meaningful results and conclusions? Please explain.		
Reviewer Name	**Reviewer Comment**	**EPA Response**
Brian Caruso	27) In general the overall integration process of the various analyses at NERL was conducted in a way that provided some meaningful results and conclusions. However, the integration process outside of NERL appears to be a significant barrier to deriving more meaningful results. The lack of consistency in the data between different organizations, data gaps for some important analyses (such as pH and conductivity), and different detection limits and analytes even for the EPA labs, all make the overall integration appear weaker. In addition the apparent lack of integration between ORD NERL, other ORD labs, the regions, and other agencies in terms of response and future monitoring and modeling needs, limits the provision of meaningful results and conclusions. With regard to the presentations, it probably would have been more helpful to present the empirical results before the WASP modeling.	27) The EPA/ORD team used publically available data to address our primary research objectives of characterizing the source, transport and fate of metals in the GKM plume to a reasonable level. EPA samples were collected and laboratory tested following Quality Assurance Project Plans and Standard Operating Procedures in place in each Region. Field sampling included pH and conductivity. We also used state and tribal data collected following documented QA/QC procedures. Data among organizations was generally very consistent. Additional data obtained from various sources, including sondes, added insights into plume movement. Given that all of these entities were responding to an emergency event, all of the groups together managed to compile a robust collection of data that we feel provided a strong and coherent analysis of the GKM event at its origin and as it traveled 600 kilometers downstream. Often the mix of data sources enabled the gaps in one place or time to be filled by nearby measurements. Uncertainties are present, but the overall story was internally consistent and supported independently by the modeling that we deployed. Differences such as detection limits affected the ability to draw

Question 2 Was the overall integration process of the various analyses conducted in a way that provided meaningful results and conclusions? Please explain.		
Reviewer Name	**Reviewer Comment**	**EPA Response**
		conclusions about low concentrations of metals, raising uncertainty as to when and if the metals returned to pre-event levels, but did not deter detection of the GKM plume as it traveled. Differences in detection limits and lack of consistency in data sets were more significant in historic data that has been collected for many different reasons. Again, this increases uncertainty in determining if the GKM plume has completely left the system or when the river concentrations will return to pre-event levels.
Charles Fitts	28) I understand that there has been a pressing timeline for pulling these studies together and that we are looking at first drafts, which I think is the proper stage for having a review that allows time for revision. I expect more effort will go into integration, peer editing, and polishing, which the entire study could benefit from. 29) It would help to expand the overview section so that it explains clearly how each of these parts contribute to achieving the project's goals and describes to what extent each part depends on results from other parts. For example, the same analysis of source mass shows up in both the empirical and WASP sections. 30) Some portions of the work could benefit from additional review and input by additional experts within	28) The EPA/ORD team continues to improve the presentation of results and has worked to improve integration between software modeling systems and empirical analysis of observed data utilizing the suggestions and feedback of the peer reviewers. These improvements will be implemented in the final report. 29). The EPA/ORD team will include a "roadmap" of the analysis as to how data was used and how modeling approaches (empirical and analytical) interacted to address the research questions.

Question 2 *Was the overall integration process of the various analyses conducted in a way that provided meaningful results and conclusions? Please explain.*		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	EPA. Although I am not expert in this area, it seemed that the bioaccumulation study could use such review, as the reviewers indicated that it may need to consider alternate methods that are not based on factors that assume a ratio of river concentration to tissue concentration.	30) The EPA/ORD team agrees and will include effects-based expertise.
Henk Haitjema	31) There were some limited connections between the presentations, particularly between the presentation "Empirical Analysis of Metal Loads & Water Quality Trends Based on Observed Data" by Dr. Kate Sullivan and Dr. Mike Cyterski and the WASP modeling. However, there was no clear overarching structure in which the various presentations had a clear place. Consequently, the results and conclusions from the individual studies could not easily be related to each other. That said, I recognize that this review was conducted before all studies were fully completed and documented (written up in a report) and as a result the integration could not yet have happened. I believe that the timing of this review, prior to producing a final document, is very beneficial for an optimal impact of the review process. Thus the lack of integration observed is not to be interpreted as a critique on this research effort!	31) The connection between the empirical analysis of water quality concentrations and plume movement and the WASP modeling has been strengthened in response to the peer reviewer's comments. This includes stronger calibration of WASP against the empirically derived load estimates to calibrate particle settling. Integration of analysis will be stressed in the final report. In addition, the modeling will be used more prominently to animate findings to facilitate presentation of findings to the public.

Question 2 Was the overall integration process of the various analyses conducted in a way that provided meaningful results and conclusions? Please explain.		
Reviewer Name	**Reviewer Comment**	**EPA Response**
Kirk Nordstrom	32) The presentation of the various analyses could have been conducted in a logical sequence. The Empirical Analysis should always go before any modeling efforts based on the observations. Most people would want to see the data first and foremost. It is also better to get a feeling for the data to see what types of modeling approaches are reasonable and which ones aren't. Modeling is usually used to fill in data gaps, to gain more insight into the processes that might explain the data, and to explore possible scenarios to evaluate their consequences. So the data should come first and then the modeling results. Otherwise the sequence with Geochemistry, followed by WASP modeling, Bioaccumulation, and ending with Groundwater seemed appropriate. 33) One aspect that was problematic is that some of the results and the presentations changed several times. That is, we received one copy of PowerPoints by cyberspace before the meeting. At the meeting we received a paper copy of the PowerPoints in a binder where some things had been changed and then when people gave presentations they sometimes had made another update and handed that out to us separately. That tells me that the Team was not quite ready and were still finessing their results. It would have been more appropriate to wait another week or two to make sure there were no important changes before presenting to the reviewers. Last minute modifications are not helpful for a review meeting.	32) The EPA/ORD team will change the sequence of presentation of the multiple elements of the analysis in presentations and reports to reflect this comment and will more carefully orchestrate the description of data concordant with modeling discussions in the final report. 33) The analysis was ongoing at the time of the review which was scheduled to evaluate progress to date. Input from members of the panel helped further refine and improve analyses. The final written report will fully describe updated analyses and will undergo additional external peer review.

Question 2 Was the overall integration process of the various analyses conducted in a way that provided meaningful results and conclusions? Please explain.		
Reviewer Name	**Reviewer Comment**	**EPA Response**
William Stubblefield	34) Is not entirely clear what is meant by the "overall integration process" of the various analyses. For example, some of the reported metals data are presented on the basis of "total metals." This is a fairly nonstandard approach for presenting metals data especially if one of the objectives of the evaluation is to assess potential impacts to exposed aquatic organisms. The array of metals present in the Gold King Mine AMD will have vastly different toxic potencies and will be present in the AMD at greatly different concentrations (ppm to ppb). To conduct an appropriate evaluation of potential effects to exposed organisms, one needs to consider the exposure to the individual metals. It might be better if evaluations were conducted on a few different metals representing a range of toxicities, proportional presence in the AMD, and environmental fate processes. Evaluating metals such as iron, aluminum, copper, and zinc would cover a range of toxicity profiles and presence in the AMD.	34) In the interest of maintaining a reasonable number of individual figures, the EPA/ORD team sometimes presents summed metals to demonstrate the general patterns of metals transport by concentration or mass that were observed with the plume. The peer review panel was generally dissatisfied with that simplification as individual metals are of interest biologically and geochemically and probably behaved differently within the system and with differing levels of importance. The final report will describe general plume movement with summed metals as an overall description of movement but will analyze individual metals as well.

Question 3 ***When looking at the full project, are there errors or gaps in the integration process that could have affected the overall analyses and/or the conclusions? Please explain.***		
Reviewer Name	**Reviewer Comment**	**EPA Response**
Brian Caruso	35) Please see comment on question 2 above.	35) See response (27)
Charles Fitts	36) Since the study focused mostly on total metal concentrations, it is possible that it overlooked behaviors of specific species of metals that would be important in subsequent risk assessments and monitoring plans. I also mention this in the fate and transport section, and suggest analyzing the fate and transport of a few metals that are likely to pose risk and may be representative of groups of similar metals. 37) Most of the concentration data we saw in the presentations was from water samples. However, slides 6 and 20 of the overview alluded to over 320 bed sediment samples. Presentation of the sediment data was limited, so if there is more to that story, perhaps more should be presented.	36) The EPA/ORD team will emphasize analysis of individual metals in the final report. 37) The presentation of data included a representative number of sediment related analyses. Future presentations and reports will include a full analysis of bed sediment data.
Henk Haitjema	38) While the "Overview" presentation offered a "Summary of Findings" (slide 25) that I found relevant and important, there was no overarching presentation that put the various studies together to substantiate these final conclusions. What is needed in addition to the work presented to the reviewers is a document with a clear statement of purpose and explanation of the motivations for the various studies. That same document then must also have a concluding section in which these studies are referenced, and the conclusions integrated into an overall set of conclusions and, where appropriate, recommendations. I did not observe fundamental flaws in the studies that negatively affected the conclusions presented.	38) The EPA/ORD team appreciates the reviewer's comments regarding method of presentation. The final report will be structured to address the recommendations of the reviewer.
Kirk Nordstrom	39) The integration process could have been improved by better communication between the Geochemical Analysis and the	

Question 3 When looking at the full project, are there errors or gaps in the integration process that could have affected the overall analyses and/or the conclusions? Please explain.		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	Empirical Analysis groups. These 2 sections are very closely aligned and have clear overlaps on the source term composition. More discussion was probably needed between these groups to have a better consensus on how to characterize the source term. It seems to me that when writing up the final report these 2 sections might be merged into one. Alternatively, writers should make clear what deserves to be called geochemical analyses and what is empirical. Whenever geochemical modeling is involved it would seem necessary to call it a geochemical analysis, however, mass balances is also considered geochemical modeling. Very often some geochemical reactions need to be assumed or modeled for the mass balances to make sense. Hence, these two sections should probably be merged.	39) The EPA/ORD team will improve description of the coordination between geochemical analysis with the various modeling and empirical approaches used and integrate discussions of these topics in the final report.
William Stubblefield	40) One of the stated objectives of the effort was "Identify the potential for water quality impacts, including municipal wells, and implications for future monitoring priorities near-term and long-term." It is not clear how this objective was going to be met. Few "exposure" concentrations were provided as a result of the Agencies analysis and little to no indications of how "impacts" were going to be assessed were discussed.	40) We have added more presentation of the potential exposure to metals for various water uses during and following the GKM event relative to water quality criteria. These will be included in the final report.

Question 4 Were the overall conclusions that were drawn from these analyses appropriate and scientifically defensible based on the analysis? Why or why not?		
Reviewer Name	**Reviewer Comment**	**EPA Response**
Brian Caruso	41) The overall conclusions drawn from these analyses generally seem appropriate, but this is somewhat difficult to determine due to the lack of clarity in the goals and objectives of the research. In addition, the conclusions are not entirely scientifically defensible based on the analysis. The primary reasons for this are generally discussed in the overall impressions above and include: • lack of clarity of goals and objectives to reflect critical information needs • lack of use EPA national criteria or standards for metals for drinking water and aquatic life as an initial screening of risk to select potential contaminants of concern for more detailed analysis and as indicators (instead of primarily evaluating total metals) • inadequate use of other existing data and information from previous investigations to evaluate and help confirm background levels for comparison • lack of inclusion of additional data and information for better reactive transport modeling, concentration and load calibration, and validation for WASP • very limited evaluation and presentation of uncertainty and sensitivity analysis of results • lack of identification of data gaps in the analysis and for future modeling	41) The EPA/ORD team presented the peer review team details of data analysis and modeling of volumes and metals concentrations at the GKM source, the timing and concentrations of metals at the plume traveled through the Animas and San Juan River systems, the fate of metals mass in surface water and sediments, including possible transport through alluvial sediments into groundwater wells. Quantifying these aspects of the GKM release was the main objective of the project as described to the peer review team in each individual unit of analysis. As recommended, we will include additional analysis of historical data, increase calibration of WASP with observed data, and will address sensitivity and uncertainties associated with data and modeling. These will be included in the final report.
Charles Fitts	42) I think the conclusions presented in overview slide 25 are generally sound and on-target. It think that the 4th bullet point about most of the metals being deposited in the Animas streambed could be more specific. The presentation could point	42) This reviewer interpreted the information correctly as to where most of the deposition occurred. The EPA/ORD team will ensure that these locations are highlighted as the areas with greatest

Question 4 ***Were the overall conclusions that were drawn from these analyses appropriate and scientifically defensible based on the analysis? Why or why not?***		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	out the specific stretches of the Animas River that received the greatest mass of deposition (RK 13-16 and RK 64-96, as discussed in question 7). There is no bullet point about the impact on wells located near the river. I think there should be an additional point made about the potential for impact in wells close enough to the river, but that sampling data showed only well 35m66km with a noticeable plume signal, which was at levels that did not pose any significant risk.	deposition in the final report. We will also include discussion of the potential impacts on wells in these areas and include findings on groundwater potential exposure in summary of findings in the final report.
Henk Haitjema	43) I believe they were, but as outlined in my response to various questions below, additional work and better documentation are needed.	43) The EPA/ORD team agrees and will address in the final report.
Kirk Nordstrom	44) Not entirely. (1) The geochemical analysis used some flawed assumptions to estimate the GKM effluent composition (see below), (2) alternative approaches to the GKM effluent composition were not considered (see below), (3) sensitivity analyses need to be employed for many of the analyses and modeling with a propagated range of uncertainty; this approach would result in upper and lower bounds for the plume at several locations downstream, and (4) I have a difficulty in seeing any scientifically defensible conclusions coming out of the bioaccumulation study – the lack of fish kills and the caged fish study are much more appropriate to address fish toxicity for such a short transient event than the attempt at modeling that was presented.	44) Items (1), (2) will be discussed in later responses following Dr. Nordstrom's suggestion to see his responses below where he provides more detail. (3) The EPA/ORD team will include discussion of uncertainty and sensitivity to assumptions in the final report. (4) We view some treatment of bioaccumulation modeling as helpful in understanding the potential uptake of metals during relative rapid movement of metals through their habitat and help to corroborate what was observed in fish within the system and to assess potential for dietary uptake. The revised work will be presented in the final report that will also receive an external peer review before publication.

Question 4 ***Were the overall conclusions that were drawn from these analyses appropriate and scientifically defensible based on the analysis? Why or why not?***		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	45) A variety of conclusions were provided in a number of the presentations; however, for the purposes of this response, we are assuming that the "summary of key findings" from the overview presentation captures the “overall” conclusions. For the most part these findings were supported by the data provided in the presentations. However, in some cases it is difficult to point specifically to the data that support a given conclusion. This is in part due to the sheer volume of data and the way that the presentations were organized based on the available time for presentation. It is anticipated that a detailed report outlining the analysis that was conducted would provide an opportunity to present an analysis in greater detail. For example, providing metal specific data rather than “total metals data” would provide greater support for the conclusions.	45). The final report will present data organized in a manner that facilitates integration of findings. This will include mapping between data and analyses used for or generated by them. Key individual metals will be showcased.

Part 2: Fate and Transport

Question 5 ***Were the overall conclusions that were drawn from these analyses appropriate and scientifically defensible based on the analysis? Why or why not?***		
Reviewer Name	**Reviewer Comment**	**EPA Response**
Brian Caruso	46) The research makes an attempt to characterize the metals concentrations and loads produced at the Gold King Mine spill. However, it is extremely surprising and unfortunate that EPA collected no samples from the release itself until what appears to be a substantial time period after the release. In addition, no samples were collected at the mouth of Cement Creek (CC) until about 4 hours after the release and after the release/plume had passed. The volume of the release was estimated by the USGS based on the change in the hydrograph at the CC mouth. Four samples were collected at the adit release up to about September 23, 2015. It was not made clear when the first sample at the adit was collected, but appears to be at least many hours to a day after the release. One of these samples collected by EPA was selected to characterize the release and use in subsequent calculations and modeling. It was stated that this was selected because it was the most comprehensive analysis. However, it is not known or made clear why the other samples were not analyzed the same way. The samples are presented on a log graph for most metals, so the variability of the results is not entirely clear. The variability and uncertainty of these adit release results should be analyzed and presented in more detail, and perhaps a mean or median over this time period should be used instead of just one sample. Also, it is not clear if any other samples from inside the adit itself, or from the ongoing drainage, had been collected and analyzed previously, prior to the release. If so, these should be compared to what was observed in the release.	46) The first samples of mine effluent were collected by Colorado DPHE within 2 days of the event and 3 more measurements were made by CO and EPA over the next 6-weeks. These data were shown to the peer review team and will be provided in the final report. In the final report, we will provide detailed description of these data and analysis used to select a value to represent the effluent concentration, including a discussion of uncertainty.
Charles Fitts	47) The data were mostly presented as total metals and did an adequate job of portraying the distribution of total metals. The presentation seldom presented data on subgroups of metals or individual metals. It might be instructive to look at empirical data for a few individual metals of interest, selected because of their importance in terms of risk and their characteristic behavior representative of a	47) The EPA/ORD team will continue to strive to find parsimonious ways to present data as appropriately as we can. The final report will describe general plume

Question 5 *Were the overall conclusions that were drawn from these analyses appropriate and scientifically defensible based on the analysis? Why or why not?*		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	group of similar metals (e.g. one metal that precipitates at a low pH range and another that precipitates at a higher pH range). Since subsequent studies will be examining risk and monitoring plans that aim to minimize risk, the metals chosen for individual analysis should include ones that are most likely to pose risk.	movement with summed metals as an overall description of movement. Individual metals concentrations, mass loads, and exposure analysis will be included in the final report.
Henk Haitjema	48) As explained below the total metals load leaving Cement Creek were probably underestimated. However, this was recognized in the analyses presented to the reviewers and could not have been avoided in lieu of the lack of more pertinent sampling (sampling of the peak of the plume in Cement Creek).	48) The EPA/ORD team and the peer reviewers discussed mass estimates from Cement Creek at great length. There are important uncertainties that would not likely be bridged with data collected after the fact given the unique conditions during the event.
Kirk Nordstrom	49) Characterizing the composition and load of the Gold King Mine spill is problematic. No samples of the mine effluent were taken during the spill event. Samples were taken some days later. When the plume hit the first gage at Cement Creek (at the mouth), samples for chemical analysis were taken well after the peak of the plume had passed. Furthermore, the first 2 samples at the gage were incomplete (no pH, conductivity, or sulfate determinations). In addition, when the plume hit the Cement Creek gage it had picked up additional sediment and dissolved substances that were not part of the original mine pool discharge. Consequently, it makes sense to consider the source water as the plume that was recorded in the Cement Creek gage right before it entered the Animas River. It is still a problem characterizing the water composition at the peak of the Cement Creek discharge because the first sample collected for analysis at the gage was about 5 hours after the spill began and contained only about 20% of the Gold King effluent as well as missing some critical parameters. I think the ART did important calculations to estimate the water composition at the Cement Creek gage peak flow from the GKM release and I shall suggest additional considerations.	49-50) The EPA/ORD team agrees. The metals concentrations and therefore mass at the peak of the GKM plume in Cement was likely much greater than first sampled at 16:00 hr after most of the plume had passed. Our methods to reconstruct the plume that were shared with the panel are generally described by this reviewer in (50) and (51).

Question 5 ***Were the overall conclusions that were drawn from these analyses appropriate and scientifically defensible based on the analysis? Why or why not?***		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	50) The Team did a straightforward conservative calculation assuming straight mixing of GKM effluent with upper Cement Creek water with no reaction. This result would normally give a bounding limit to the chemical composition of the plume. But which limit? High or low? If there is a reaction in progress, is that increasing or decreasing metal concentrations? Both are possible. Oxidation and precipitation of iron would tend to remove metals. Dissolution of soluble salts from the eroded waste piles and Cement Creek would increase metals. Erosion of fine clays might provide more surfaces for metal sorption and partitioning from dissolved to the solid phase. From my experience with weathering of mine tailings and waste rock during storm events, there is a brief and sudden increase in dissolved metals during the early rise of the discharge and then a decrease from dilution. In this instance, dilution is with GKM release water and upper Cement Creek flow because it is not a rainstorm event. But there is still likely to be a sudden increase early in the plume movement and then a drop to the concentrations of the GKM effluent for the remaining majority of the plume release followed by decrease to Cement Creek baseline once the GKM plume has passed. This early spike in concentrations would be from the addition of soluble salts and films of concentrated acid mine water contained within the tailings pile downstream and separate from the effluent composition released from the mine. I would anticipate sorption processes to be largely ineffective at this pH (~3) and with higher than normal metal concentrations. The plume is moving too fast for much oxidation and precipitation of iron. Hence, I would argue that the total plume load would be greater than that expected from just the analyses of the GKM effluent in both dissolved and fine particulate matter combined with the estimated discharge. Further, I would argue that the first measured concentrations at the gage on Cement Creek should be close to conservative mixing (20% of GKM and 80% upper Cement Creek water) but that the dissolved concentrations were higher	50) We will revisit how to characterize the dissolved fraction of the metals concentrations at the peak of the plume in Cement Creek. The final approach that we use to characterize the metals in Cement Creek will be fully described in the final report.

Question 5 Were the overall conclusions that were drawn from these analyses appropriate and scientifically defensible based on the analysis? Why or why not?		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	during the first ½ hour of the GKM release. How much higher is very difficult to say so this calculation would be a lower bound that can be compared to another estimate. It can also be compared to a loading calculation that takes a constant composition GKM release as a lower limit after mixing with Cement Creek baseline water. This constant composition chosen by the ART was the August 15 sample because it was the first complete analysis of the mine effluent after the plume had passed. There were 3 other samples that I would say could be used as well from other time periods. Although pH, sulfate concentration, and conductivity data were sometimes missing, it is possible to reconstruct these by optimizing pH and sulfate concentrations using charge balance for pH (using the PHREEQC program) and conductivity balance (using either PHREEQC or WATEQ4F although WATEQ4F would be preferable because it is more reliable for acid mine waters). 51) With regard to estimating the composition of the Gold King effluent water during the spill, the explanation could have been clearer, especially since this composition is critical to the entire interpretation of downstream fate and transport. Unfortunately, the data available is sparse and incomplete which adds to the confusion. As I understand it, there are two key sets of data: (1) direct analyses of the Gold King effluent but collected after most of the spill had occurred with dates of 8-07-15 and 8-11-15 collected by CDPHE and dates of 8-15-15 and 9-21-15 collected by the EPA and (2) Cement Creek samples collected during the tail end of the plume movement (first sample was collected about 5 hours after the spill began). The CDPHE samples are missing critical data such as pH, temperature, conductivity, iron and sulfate concentrations. The Cement Creek samples are Gold King effluent mixed with 80% or more of upper Cement Creek water, possibly mixed with some dissolved soluble salts, eroded sediments, and their pore waters. The GKM effluent composition had to be estimated from these	51) Regarding alunite: based on review of the literature, important details of alunite chemistry are still in need of resolution. Also, based on discussions during peer review, the EPA/ORD team has performed kinetic calculations on alunite dissolution rate with data recently published in the literature; these calculations also argue against alunite dissolution accounting for the high aluminum reported in the Cement Creek sample. Because of these uncertainties, the EPA/ORD team agrees that alunite chemistry cannot be invoked to estimate load limits with any confidence and load and source estimates based on equilibrium chemistry driven by alunite will not be used in the final report.

Question 5 Were the overall conclusions that were drawn from these analyses appropriate and scientifically defensible based on the analysis? Why or why not?		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	limited pieces of data. The approach taken was to use the Cement Creek USGS gage data to determine the proportion of that water containing GKM effluent. Then unmix the water assuming conservative mixing. Then most of the concentrations were increased by an amount that was estimated by assuming that alunite saturation equilibrium was achieved in the GKM effluent and increasing the aluminum concentration accordingly. Alunite saturation equilibrium was indicated in a paper by Eary (1999) and this is the first time I have heard of making this assumption to estimate a mine water composition. The question is whether this assumption is reasonable and whether there are other, more reasonable approaches. Alunite is a relatively insoluble mineral which is slow to dissolve and precipitate unless the temperature is increased substantially above ambient. I have read the Eary (1999) paper and the case made for alunite solubility equilibrium at low pH is extremely speculative. I say that because the plots that Eary showed (1) had considerable scatter, (2) were not done the normal way with the log of the activity of the free aluminum ion vs pH – he used dissolved aluminum concentrations vs. pH which doesn't really tell you much and cannot be directly compared to solubility of alunite, and (3) he doesn't show saturation indices for alunite as he does for gypsum, fluorite, and other carbonate and sulfate minerals. Further, he was looking at a pit lake which can be different than underground mine effluent. Not to mention that there are a range of thermodynamic properties for alunite so we really don't know how the solubility might change with solid solution substitution, particle size and crystallinity, and uncertainty in the thermodynamic properties. I am sure that alunite does reach equilibrium solubility in some environments but I would be very hesitant to apply it for this situation. Hence, I would discourage using this type of modeling approach to correct the mine effluent chemistry to the original composition.	52) We will refer to the QAPPs used by EPA regions for sample collection

Question 5 Were the overall conclusions that were drawn from these analyses appropriate and scientifically defensible based on the analysis? Why or why not?		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	Instead, I would take the range of composition of the mine effluent water (max and min as bounding conditions) that was sampled later, correcting pH and sulfate concentrations as mentioned above, and compare that to the conservative estimate made from the mixing calculation that the team did from the 1600 hour sample. Then I would consider a 50% to a 100% increase in concentrations during the first ½ hour only of the GKM release to account for washout of the tailings pile for an upper bound of the loading and concentrations. 52) The characterization of the metals concentrations and the loads begins with the field collection of water samples and field parameters, followed by laboratory analyses. The ART did not participate in these activities. There may have been some QA/QC (quality assurance/quality control) tasks done by individuals in the team, but, apparently not as a group effort. Consequently, some unexplained discrepancies occurred in the results presented, such as several elements in which the total (unfiltered, acidified) concentration is substantially less than the dissolved (filtered, acidified) concentration. This discrepancy is most apparent with As, Sb, Pb, Mo, and V in the Cement Creek samples that were used to estimate the source effluent composition from the mine, which are sometimes discrepant by an order of magnitude or more and that is far greater than the analytical error. One way of avoiding these problems is for the team to engage in conversation with the field collection personnel and with the laboratory and any QA/QC examiners to determine if there were any sampling problems or analytical problems that could explain these anomalies. I have seen similar discrepancies before with metal concentration data from mine-influenced water at Superfund mine sites and the main problem seemed to be the lack of communication between those collecting the samples, those analyzing the samples, and those providing QA/QC. Without knowing field difficulties in collecting samples and whether there were any modifications of normal procedures (waters should be filtered and	and testing and include this information in the final report. This project also operates under an approved Quality Assurance Project Plan with provisions for data quality control.

Question 5 *Were the overall conclusions that were drawn from these analyses appropriate and scientifically defensible based on the analysis? Why or why not?*		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	acidified immediately on collection; unfiltered samples acidified immediately except for anion sample) and without knowing if any serious interferences or possible contamination occurred with the analytical procedures, it becomes impossible to know how best to interpret the data. The higher dissolved concentration could be a contamination problem and the lower total value closer to the truly dissolved value OR the dissolved concentration could be more accurate, and the total concentration could be a result of the sample being collected in a different part of the river or an analytical interference. These are important issues that can affect any attempts at interpreting the results for fate and transport. 53) For this report, everything that can be known about sampling, preservation, and analytical procedures should be spelled out more. There were probably different procedures employed by State, Federal, tribal groups and other parties (for example, were samples sometimes stored for some time before acidification? Was the same acid used among agencies for acidification? Was acidification done with the same strength acid and with the same volume per volume of sample or to the same pH? If samples were filtered, what was the filter pore size? Instead of providing the EPA method numbers for the analytical method, it would be better for the reviewers to simply have the actual instrumental technique employed (ICP-AES or ICP-MS, etc.) which might be more useful when comparing results from different agencies. Reviewers and stakeholders might want to know the QA/QC for the data. I recommend a table that lists what samples were collected when, by whom, whether filtered on site or not, if filtered what pore size was used, whether acidified on site or later, if later how much later, what and how much acid was used. A separate table can cover QA/QC data (blanks, spiked recoveries, standard reference water samples, alternate methods). These tables can be appendices in the report, but it is essential to include this information because it supports the credibility and usefulness of the data for modeling and interpretation.	53) The EPA/ORD team will further research and compile information on data collections methods for data used in this study. The final report will provide summary tables and links to quality assurance documentation.

Question 5 Were the overall conclusions that were drawn from these analyses appropriate and scientifically defensible based on the analysis? Why or why not?		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	54) Also, several metal concentrations that were reported are of questionable value such as cobalt, barium, and beryllium. I know these are easy to determine by ICP-AES and ICP-MS but if there are no obvious toxicological concerns and the concentrations are quite low, then that could be stated explicitly. It could also be stated that certain metals were selected (and others not) for continued description in the plume movement because of their concentrations and their potential toxicity. 55) My understanding is that grab samples were collected rather than width-integrated composite samples. Under the given conditions, it might be that grab samples were the only ones possible at many of the sites, however, some width-integrated samples should have been possible or at least near-central-velocity samples collected. If the team doesn't know what the velocity of river was where the sample was collected, it could easily affect the results. Some information on this aspect should be provided in the final report.	54) A number of metals were present in only very low concentrations and below detection limits. The final report will discuss how samples at detection limit concentrations are used and what metals are more fully analyzed and reported and why. 55) There is inherently variability in both surface water and sediment samples due to collection techniques used by different organizations. Many surface water samples were grab sampled rather than using depth integrated techniques. This could add to variation among samples, although much of the river system was sufficiently turbulent to minimize these errors. The final report will describe river conditions that would contribute to mixing assumptions.
William Stubblefield	56) It is difficult to address this question given the "total metals" approach used in the analysis of the data. It would seem logical that there are sufficient individual metals data to permit a "by metal" analysis of exposures. This would be helpful in addressing the questions associated with potential impacts to organisms and would allow for better characterization of the fate and transport of individual metals.	56) Information and analyses of individual metals will be emphasized in the final report.

Question 6 ***The concentration of metals near the release site in the receiving waters had to be estimated from samples collected after the much of the plume had passed. Were the estimates of metals concentration at this location appropriately calculated through scientifically sound methods using available data?***		
Reviewer Name	**Reviewer Comment**	**EPA Response**
Brian Caruso	57) There was a reasonable attempt made to estimate metals concentrations at this location (adit and or CC?) using scientifically sound methods based on available data. As stated above, however, there appears to be many questions and issues with regard to the analysis methods and assumptions. 58) At the adit release, estimated concentrations and loads were only based on one sample, whereas the summary statistics, variability, and uncertainty of the four samples collected over the month and half after the release should have been better presented and perhaps used in the analysis. Downstream at the CC mouth, an attempt was made to back calculate the concentrations and loads during the peak flow, and to account for dissolved and particulate metals scoured from CC by the passing flood wave. It is not clear how WASP was used to calculate the Maximum Total Concentration to aid with this. This appears to be done outside of WASP as input to the model as a simple mass balance using the estimated release concentrations, estimated background upstream CC concentrations and flow, and downstream measured flow. This mass balance approach seems to be appropriate. However, the analysis is not clear and background concentrations in CC appear to have been based on post-plume concentrations at the mouth, even though there are many pre-release sample and analysis data available for CC. These previous data could have, and probably should have been used, or at least collected and compared to the background estimates used.	57) The data and analysis method of metals concentrations in the mine will be fully discussed in the final report. 58) The data and analysis method of metals concentrations in the mine will be fully discussed in the final report. We clarify that we did not use WASP to calculate the metals concentration at the peak flow in Cement Creek but rather we used the mass balance approach described by this reviewer. This method will be fully described in the final report. If background concentrations in Cement Creek are needed for the analysis the supporting data will be fully described.

<table>
<tr><th colspan="3">Question 6
The concentration of metals near the release site in the receiving waters had to be estimated from samples collected after the much of the plume had passed. Were the estimates of metals concentration at this location appropriately calculated through scientifically sound methods using available data?</th></tr>
<tr><th>Reviewer Name</th><th>Reviewer Comment</th><th>EPA Response</th></tr>
<tr><td></td><td>59) The explanation of how the plume concentrations were re-constructed at the CC mouth is not entirely clear. It is not clear whether the PHREEQ modeling was needed, or what value the WASP modeled concentrations are considering; these are estimates based on conservative constituents with no reactive transport.</td><td>59) WASP was not used for this purpose of reconstructing the dissolved and colloidal/particulate concentrations during the GKM plume at the mouth of Cement Creek. The methods used will be fully described in the final report.</td></tr>
<tr><td>Charles Fitts</td><td>60) The calculations that lead to the “estimated peak” concentrations shown in the bar chart of empirical slide 20, the WASP and Empirical concentrations in slide 27, and “Simulated Load” in WASP slide 13 needs to be explained in more detail. This is critical since the extrapolation needed at early times strongly affects the estimated total load in the plume. In the following table, I analyzed total concentrations (Ct) vs. discharge (Q) for the early time observations and the early time simulated concentrations in Silverton.

<table>
<tr><th>Time</th><th>Q (cms)</th><th>Ct simulated (mg/L)</th><th>Ct observed (mg/L)</th><th>Ratio Ct simul./Q (mg/L/cms)</th><th>Ratio Ct observ./Q (mg/L/cms)</th><th>Source</th></tr>
<tr><td>12:45</td><td>3.5</td><td>37000</td><td></td><td>10571</td><td></td><td>WASP slide 13 (W13)</td></tr>
<tr><td>12:45</td><td>3.5</td><td>29557</td><td></td><td>8444</td><td></td><td>Empirical slide 20</td></tr>
<tr><td>16:00</td><td>1.1</td><td>10500</td><td>11485</td><td>9545</td><td>10441</td><td>W13</td></tr>
<tr><td>19:25</td><td>0.1</td><td>3000</td><td>998</td><td>30000</td><td>9980</td><td>W13</td></tr>
</table></td><td>60) We have ensured that the initial value of Ct is consistent across all analyses and we will clarify the description of the method in the final report.</td></tr>
</table>

Question 6 ***The concentration of metals near the release site in the receiving waters had to be estimated from samples collected after the much of the plume had passed. Were the estimates of metals concentration at this location appropriately calculated through scientifically sound methods using available data?***		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	It makes sense that higher stream discharge and velocity would correlate to higher suspended sediment load and higher total concentration. Based on observations, it appears that the ratio Ct/Q was about 10,000 at the earliest observations (blue). Lacking other evidence, this ratio may be reasonably applied to earlier times, but there is uncertainty in this extrapolation that should be acknowledged in the report. The simulated ratio Ct/Q was about 10,000 for both the 12:45 and 16:00 Ct estimates (purple), but about 30,000 for the 19:25 Ct estimate (red). It seems reasonable to keep the 10,000 ratio to estimate the early time concentrations, which is close to what was done. I think the bullet item on WASP slide 13 should read "Assume total concentration (Ct) is proportional to flow." Also the 12:45 and early time Ct numbers should be made consistent across the study and consistent with the explanation for how the early concentrations were extrapolated. 61) Looking at the graph in WASP slide 13 and the 19:25 row in the above table, the simulated concentrations from about 18:00 onward are systematically higher than observed, and they are noisy, bouncing up and down as though the simulated concentrations could only move in large quantum leaps. This portion of the simulated Ct should be modified to remove the noise and to better match observed Ct, even if the impact on simulation results downstream is minor. 62) Since the source mass is critical to all analyses, this deserves attention. If other concentration data becomes available from other	61) We will examine model output. Results will be reflected in the final report. 62) We will incorporate any new data in the final analysis and report.

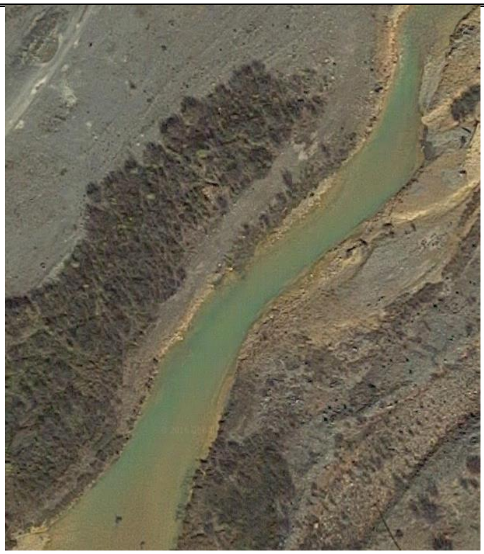
Question 6 *The concentration of metals near the release site in the receiving waters had to be estimated from samples collected after the much of the plume had passed. Were the estimates of metals concentration at this location appropriately calculated through scientifically sound methods using available data?*		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	sources for the early hours in Cement Creek or GKM, it should be incorporated in revised source estimates.	
Henk Haitjema	63) On slide 20 of the "Empirical Analysis" presentation, two approaches are mentioned to arrive at the maximum total concentration (CMAX) in the peak of the plume at 12:45. These are using WASP for CMAX and PHREEQ for maximum dissolved concentration. In fact, as I understand it, WASP was not involved in determining CMAX but a mass balance calculation outside of WASP was used (see discussion under question 14). I cannot comment on the PHREEQ method due to unfamiliarity with this code and the processes it simulates. 64) My overall assessment is that the dissolved concentrations in the peak are probably fairly well estimated, but that the suspended total metals concentration in the peak is almost certainly significantly underestimated. In fact, this is recognized in the current study on slide 20 with the comment on the graph: "Concentrations at 12:45 peak probably much higher." In summary, the current study does offer reasonable estimates of peak concentrations and recognizes the underestimation due to the unknown amounts of suspended materials in the peak of the plume in Cement Creek.	63) We will clarify and add additional description of the methods in the final report. 64) We understand that the peak concentration as the GKM plume passed may be higher than the later 16:00 hr sample. The panel offered insight on how the total and dissolved fractions could be estimated. The modified approach that we use will be fully described in the final report.
Kirk Nordstrom	65) Not entirely. More use should have been made of historical data. This was mostly addressed above. There is a serious problem with some of the analyses (e.g. CC06 and GKM13 collected on 8/15/2015) in that many of the total concentrations of metals were lower than the dissolved concentrations. This can occur from problems with field sampling and samples that were not filtered and	65) Each EPA Region followed a Sampling and Analysis Plan when collecting samples (The sampling and quality assurance plans can be accessed at https://www.epa.gov/goldkingmine/s

Question 6 *The concentration of metals near the release site in the receiving waters had to be estimated from samples collected after the much of the plume had passed. Were the estimates of metals concentration at this location appropriately calculated through scientifically sound methods using available data?*		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	acidified on site (which probably did not happen for the earlier collected samples). Hence, a table summarizing the information on how water samples were collected and when filtered and acidified, is crucial to interpreting the results. Also, alkalinities of 5 mg/L are reported for these samples when the pH is too low for there to be any detectable alkalinity. This contradiction needs to be resolved. Further, the acidities are reported but I am not sure they are used or needed anywhere. There are several different methods for acidity so the result is very method dependent. If there is a need to report these, then the method used needs to be documented.	eptember-14-2015-interim-sampling-plans. Samples were filtered and acidified on site. The EPA/ORD team will compile information on data collection methods for data used in this study. The final report will provide summary tables and links to quality assurance documentation We will check on the referenced values of alkalinity. We are not aware of measured values of acidity. We reported calculated acidity estimates. We will explain the methods of calculation and limitations of these values in the final report.
William Stubblefield	66) A number of questions were raised regarding the accuracy of the estimated metals concentrations in the original AMD release. EPA staff acknowledged that there was a degree of uncertainty associated with the estimates and this was reflected in the presentations. It was recommended by the reviewers that EPA adopt an approach that characterizes the degree of uncertainty associated with the discharge estimates and incorporate that into the overall presentation. This would result in something of a "sensitivity analysis" that would bound the "best-case" and "worst-case" scenarios.	66) The EPA/ORD team will include discussion of variability and sensitivity in future presentations and the report.

Question 7 ***Were the data analyzed and visualized properly in regards to sediment metal concentrations in the post-plume period in Cement Creek and the Animas River?***		
Reviewer Name	**Reviewer Comment**	**EPA Response**
Brian Caruso	67) This comment relates to both questions 7 and 8 since they are related. The intent of this question is not clear. Is the intent of this for the post-plume period for all rivers, or just for the post-plume period in CC and during the plume in the Animas and San Juan rivers? Post-plume is important for later or ongoing resuspension or dissolution of metals to the water column during higher flows. Of course during the plume is also important for a number of reasons. Does 'sediment' refer to bed sediment or particulates (or colloids) in the water column? We assume this refers to bed sediment. 68) The methods used for estimating sediment metal concentrations are not clear. Although the methods used for estimating the dissolved and sediment/particulate metals load from CC during the release generally seem appropriate based on the mass balance approach, there appear to be a number of concerning issues. Any previous data from other studies on bed sediments in CC are not presented or used. These could include sediment physical characteristics (particle sizes) and sediment chemistry and metals concentrations. Similarly, any previous background metals concentrations in water are not presented. A significant amount of previous studies by USGS and others in CC and the Animas have been performed with these types of data. 69) It appears that with the exception of sediment data collected by an academic researcher (Dr. Williams?), EPA collected no sediment samples or data during or after the plume to help evaluate sediment metal concentration estimates or calibrate the WASP modeling.	67) The project analyzed and presented informations on metals in surface water and sediment during the GKM plume passage and in the fall months post-event. Sediment refers to bed sediment. We refer to the non-dissolved fraction in the water column as colloidal/particulate. It may become sediment if it transfers to the bed. 68) The EPA/ORD team is generally not focused on the bed sediments of Cement Creek but we do evaluate metals concentrations in streambed sediments throughout the Animas River. The final report will include analysis of historic data available from USGS and other studies in this area. 69) The EPA/Regional teams sampled bed sediments throughout the Animas and San Juan River systems during and after the GKM plume, collecting over 300 samples at once or repeatedly at multiple locations and greatly expanding on sites that had previously been measured. The sampling and quality assurance plans can be accessed at https://www.epa.gov/goldkingmine/september-14-2015-interim-sampling-plans. There is also some pre-event data at a few selected locations

Question 7 ***Were the data analyzed and visualized properly in regards to sediment metal concentrations in the post-plume period in Cement Creek and the Animas River?***		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	70) It appears that sediment (bed) metals concentrations were not estimated except for those estimated with the WASP model. There are issues with these estimates, as discussed in comments below.	in the Animas River headwaters streams (several of which coincide with USGS sampling sites). These data are used to analyze deposition from the Gold King Mine plume and to calibrate WASP. The use of this data will be fully described in the final report. 70) Observed bed concentrations were analyzed and discussed in the "Empirical" section of the presentations and in the WASP section of the presentation. Metals mass (rather than concentrations) were emphasized in the peer review. The final report will fully discuss bed sediment concentrations and mass.
Charles Fitts	71) The presentation about measured sediment concentrations was brief: slides 44-48 of the empirical section. There are inherent difficulties in distinguishing plume-event sediment from other sediment, and in concentration variations with sample location. Therefore, it may be difficult to conclude much from sediment concentration data. There did appear to be declining trends in sediment concentrations after the plume passed the lower Animas (lower two graphs of empirical slide 47). There were WASP simulations that indicated where sediments from this event were likely to have deposited (WASP slides 25-27), but as I say elsewhere, these WASP results should be viewed as qualitative, not quantitative.	71) The EPA/ORD team agrees that it will be difficult to separate the effects of the GKM plume relative to pre-existing AMD contamination of bed sediments, except to look for higher concentrations post event relative to pre-event. WASP predictions of deposition zones are corroborated with observed samples in empirical analysis. These comparisons will be fully discussed in the final report. 72) We agree that the mass analysis derived empirically at individual locations based on

Question 7 Were the data analyzed and visualized properly in regards to sediment metal concentrations in the post-plume period in Cement Creek and the Animas River?		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	72) A better approach to estimating where plume sediment was deposited would be to examine time-series of total metal concentrations at gages along the Animas River. Estimate the total mass of metals passing a point in the river by numerically integrating Ct*Q data through time, like appears to be shown in slides 30, 34, 35, and 44 of the empirical presentation. Mass changes from one station to the next one downstream could be due to deposition or mass added or subtracted at tributaries or diversions. From the graph in the lower right of slide 44, it appears that most metals deposited just below Silverton (~RK 13-16) and above Durango (~RK 64-96). This empirical approach to estimating deposition trends has a much stronger basis than the deposition results shown in WASP slides 25-27. This analysis should be expanded and highlighted in the empirical presentation and overview, and the WASP analysis section should compare its results to the empirical analysis estimates of deposition. 73) The relatively flat stretch of the Animas River below the confluence with Cement Creek (~RK 13-16) is an area where a significant fraction of plume suspended sediment probably was deposited. Average plume flow velocity would have dropped dramatically beyond the confluence due to the gentler gradient and wider channel, and the abrupt increase in pH would have promoted precipitation and sorption. This stretch of the Animas has alternating riffles and pools and the larger pools would have been particularly ripe for deposition. The image below shows one >100m pool in the Animas channel in this area.	flow and measured concentration is the strongest evidence for how much and where metals mass deposited and this is the primary method used to quantify deposition. We have also used those calculations as part of the strong calibration of WASP similar to the recommendations. Methods and results will be fully discussed in the final report. 73) We agree with the comment regarding the likelihood of deposition in the reach of the Animas River between the confluence with Cement Creek and before the river goes into the canyon below Silverton. There are other likely deposition zones as well. These will be discussed in the final report. 74) We agree that WASP which moves downstream at an average velocity calibrated to USGS at-a-station hydraulic geometry would not capture velocity variability at the riffle/pool scale. Although our subsequent calibrations of WASP have improved its deposition and erosion performance, we note that the results shared with the peer reviewers did show a large decline in concentration in the reach described at left and again at the exit point of the canyon and that future erosion

Question 7 ***Were the data analyzed and visualized properly in regards to sediment metal concentrations in the post-plume period in Cement Creek and the Animas River?***		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	 74) The WASP modeling was too large-scale and homogenized to capture the local differences in velocity between riffles and pools. WASP slides 25-27 indicated minor deposition in this area, but empirical slide 44 indicates deposition of about 40% of the metals mass from Cement Creek in this stretch. I suspect significant plume mass was deposited in these pools and some will move downstream during subsequent high-discharge events. 75) Note: a typo in empirical slide 38: 2nd line should say "High acidity" or "Low pH". 76) In empirical slide 39, explain what blue dots are in lower right plot.	could occur at both points. We agree that deposited material from the GKM plume will likely be entrained with bed sediments in future flow events. WASP cannot capture localized effects such as pools, since each WASP segment behaves as a well-mixed reactor. Our settling rates, however, are based on the empirical reconstruction of the plume from observed concentrations, so we capture any net loss due to these locations as a reach average. Model simulations will not be able to capture mobilization of these localized effects, however. We will discuss these uncertainties and possibilities in future presentations and the report. 75) We have corrected the typo. 76) The triangles are measured pH values and the blue dots are calculated at specific locations based on the trend line shown in the regression. We will explain this in future presentations and the report.

Question 7 ***Were the data analyzed and visualized properly in regards to sediment metal concentrations in the post-plume period in Cement Creek and the Animas River?***		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	77) The graphs in empirical slide 48 need axis labels and a better explanation. 78) In empirical slide 50, it would provide helpful perspective to show estimated total metal transport during a typical spring runoff season, in addition to the estimates for the late August storm.	77) The axis labels with descriptions will be included in the final report. 78) The final report will include analyses of the types described.
Henk Haitjema	79) Generally they were, although other reviewers were often critical of the lumping of metals into a total metals load or concentration.	79) The final report will provide data on key metals individually.
Kirk Nordstrom	80) It must be stated much more clearly that the sediment load is a mix of (1) clays eroded from the tailings pile during GKM release, (2) clays (mostly Fe and Al oxyhydroxides) formed during oxidation and mixing with downstream transport, and (3) clays eroded from Cement Creek during turbulent mixing of the GKM plume. Further, these sediments should be compared to suspended or clay sediments that have been determined earlier in USGS studies to see what the chemical differences are and how much they can be related to the actual GKM release.	80) The EPA/ORD team will describe the likely mix of sediments entrained between the mine entrance and the bottom of Cement Creek as suggested. We will characterize sediment in the Animas River during or after the GKM release where data are available.
William Stubblefield	81) As previously discussed, very little information regarding bed sediment metal concentrations were provided. Currently, the state-of-the-science for evaluating metal concentrations in sediments and the potential impacts on sediment dwelling organisms requires information about the acid volatile sulfide content of the sediment and the simultaneously extracted metal concentrations of other metals present in the sediment. It did not appear that this information was available for the sites downstream of the Gold King Mine.	81) The EPA/ORD team is not aware of any publically available data as suggested that was collected during or after the GKM release. We will report metals concentrations determined using the standard methodologies employed by EPA and others.

Question 8 ***Were the data analyzed and visualized properly in regards to sediment metal concentrations in the post-plume period in Cement Creek and the San Juan River after receiving mine contaminated water from the Animas River?***		
Reviewer Name	**Reviewer Comment**	**EPA Response**
Brian Caruso	82) Please see comment on question 7 above.	82) See response to comments (67)-(69)
Charles Fitts	83) I assume the question refers to the San Juan River, not Cement Creek. By the time the plume reached the San Juan River, the muted plume signal was hard to detect and sediment concentrations in the San Juan did not show a discernable plume signal. The data presented on this subject was brief (empirical slide 46). Hopefully the accompanying text, when written, will clearly explain the origin of the data and conclusions drawn from this slide.	83) The final report will describe the sediment data and conclusions clearly and will present analyses of the plume in the San Juan River.
Henk Haitjema	84) Generally they were, although other reviewers were often critical of the lumping of metals into a total metals load or concentration.	84) Future presentations and reports will provide data on key metals individually.
William Stubblefield	85) Very little information was provided regarding sediment concentrations in the post-plume period in waters downstream from the GKM and Cement Creek. Information regarding individual metals would be helpful, however it is recognized that it will be difficult to attribute specific metals concentrations to the GKM incident given the ongoing contamination that exists in the area as a result of other operations and abandoned mines.	85) The final report will provide more information on bed sediment concentrations of metals collected throughout the Animas and San Juan Rivers.

Part 3: Geochemistry

Question 9 ***Were the geochemical principles to characterize transport and fate of acid mine drainage appropriately applied and interpreted? Please explain.***		
Reviewer Name	**Reviewer Comment**	**EPA Response**
Charles Fitts	86) I am less of an expert in geochemistry, so my comments in this section are limited. I found the discussion of the American tunnel plugging and rising groundwater levels (slide 17) quite interesting from a hydraulics and geotechnical standpoint. I suspect that even if EPA had not done earthwork near the GKM entrance, the plug of loose fill at the GKM entrance may have eventually failed by internal erosion (piping) in a manner similar to what occurred on 5 August 2015. As heads inside the mine rose, the hydraulic gradient across the plug increased. Excavating activities also increased the gradient across the plug, but it is quite possible that even without that activity, the increasing gradient could have eventually triggered a piping failure and a sudden release of water stored behind the plug.	86) The EPA/ORD team found the background on the subsurface hydrology of these mines compiled by the US Bureau of Reclamation to be very informative as well.
Kirk Nordstrom	87) Much of the geochemistry followed well-accepted principles but there were some exceptions. I have addressed these in my comments above.	87) No response required. .

Question 10 ***Were precipitation and mineral saturation analyses of the acid mine drainage appropriately applied for interpreting metals fate in the river system? Please explain.***		
Reviewer Name	**Reviewer Comment**	**EPA Response**
Kirk Nordstrom	88) For the most part, yes. First, the application of dissolved iron oxidation rates was helpful to point out the enormous increase with pH. However, the fact that microbes can speed up the rate enormously at low pH was not mentioned. This should be mentioned along with the caveat that microbes would not have enough time to develop sufficient colonies in the short time of the release to affect much oxidation. There is often a 1-2 week lag time necessary before microbial colonies are of sufficient concentration to show detectable changes in the ferrous iron concentration. 89) Second, the saturation indices for calcite and dolomite were very pertinent and appropriate. This is especially important in pointing out the neutralizing capacity of the Upper Animas River. It would be really useful for the final report to do a simple mixing with reaction calculations with PHREEQC to simulate the effect of mixing the estimated plume (at or near the peak GKM release) at the mouth of Cement Creek and the Animas River to show the strength of the Animas in neutralizing the plume. The geochemical analysis has made a start down this path but a little more work should be done to complete this effort. I see it as a very important part of the overall characterization. 90) The saturation indices for amorphous gibbsite was an appropriate figure, but it is a little disturbing that the saturation state shows considerable oversaturation for many of the data points. We have not seen quite as much of this amount of oversaturation. Because it is not reasonable to have such supersaturation relative to freshly precipitating Al hydroxides, it should be assumed that some particles were not fully filtered out. As pointed out for the large supersaturation for iron hydroxides, considerable Fe particles get through the filter and indeed, may have been formed during storage of samples if they were not filtered and acidified right away.	88) We mention microbial oxidation rate when discussing Cement Creek and will make that clear in the final report. 89) We agree with the suggestion to use geochemical computer coding to model reactions at selected locations. Results of this modeling will be included in the final report. 90) The EPA/ORD team also noted the spread on the gibbsite saturation indices, including an indication of supersaturation. There are several possible contributing factors discussed by the group during the review, ranging from filtering inefficiency to estimation techniques. This possibilities will be further explored and discussed in the final report.

Question 11 Was the neutralization of acid mine drainage and subsequent fate of dissolved and colloidal/particulate metals appropriately interpreted? Why or why not?		
Reviewer Name	**Reviewer Comment**	**EPA Response**
Kirk Nordstrom	91) Yes, a very good start on the neutralization and fate of colloids was done. As mentioned above, a PHREEQC simulation of mixing with reaction to compare with the more qualitative description would wrap this part up nicely. 92) The sorption calculations are considerably speculative, at least in the way they were described. The ART should use Dzombak and Morel's (1990) book on sorption to apply modeling because it is the only place where a self-consistent set of data is available. I am still not sure that scientifically justifiable results can come out of this but at least this would be a starting point. Also, it should be noted that Webster et al. (1998) EST 32, 1361-1368 found that the sorption of acid mine drainage precipitates and schwertmannite were different than ferrihydrite, shifting the sorption edge.	91) We will use geochemical computer coding to model reactions at selected locations. Results of this modeling will be included in the report. 92) Modeling of trace metal sorption on ferric hydroxides was supplied only to offer qualitative support of the conceptual model that dissolved trace metals would be scavenged from solution by the combined phenomena of electrostatic and chemi-sorption on the incipient ferric and aluminum hydroxides. We will review Dzombak and Morel's book for insight that may offer improvements to the analyses and we will consider adding Webster et al. to our discussion of sorption onto mineral surfaces in our analyses.

Part 4: Water Quality Analysis Simulation (WASP) Modeling

Question 12 ***Did the WASP modeling appropriately apply modeling parameters to estimate the movement of plume water? Please explain.***		
Reviewer Name	**Reviewer Comment**	**EPA Response**
Brian Caruso	93) The methods and results for the WASP modeling were unclear and it appears that a number of common and well-accepted modeling practices were not used, with no clear or acceptable explanation of why. Although I understand that there is a lack of data in some areas and that some time constraints contributed to the approach used, I feel that the problem is important enough that the best modeling approach possible should be used to derive the most accurate and useful results possible. Primary comments and issues are itemized below: • 94) Although the primary goal of the modeling was presented, this goal is very general and vague, which leads to a great deal of generality and uncertainty in the model results. • 95) The segmentation and structure of the model for surface water and sediment segments were briefly discussed, but these should be clearly presented in a map or schematic. • 96) Although the modeled discharge was calibrated to the flow measured at the USGS gages, there was no attempt to include or model major tributary inputs to, or irrigation or other	93) Through the peer review and associated comments, we have updated the WASP modeling effort. These include partitioning of individual metals in dissolved or particulate form using WASP7 TOXI as well tight calibration of model particle deposition to observations. We will fully describe the implementation and calibration of WASP along with calibration statistics in the final report. This work will be further peer-reviewed before being finalized. 94) We will ensure that modeling goals are specific in the final report. 95) A map for the model domain will be incorporated into the report. 96) WASP modeling utilized USGS river gage data to parameterize model hydrodynamics. Hydraulic geometry relationships determining the depth and width exponents of power equations were determined from the gage available cross-section data (Leopold and Maddock, 1953). We calibrated the Manning's roughness coefficients so that velocities matched observed flow results. We assume that irrigation influences were negligible during plume movement since use had been curtailed,

Question 12 *Did the WASP modeling appropriately apply modeling parameters to estimate the movement of plume water? Please explain.*		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	takes from, the main stem of the rivers. These should have been included. It is also not clear whether and why both Geometry equations and Manning's Equation using roughness and slope were used to estimate flow parameters such as velocities. • 97) There appears to have been no attempt to calibrate or validate the modeled concentrations or loads. This is standard practice and there appears to have been adequate data collected during and immediately after the release to at least calibrate the model, so this should have bene done. • 98) It is not clear why the TOXI module of WASP, for some reactive transport of toxicants including metals, was not used. This module includes Kd for partitioning between dissolved and particulate forms, first order decay, and a diffusion coefficient. Why were only total metals modeled, whereas both dissolved and total (or particulate) could have (and probably should have) been simulated. Although WASP cannot model equilibrium precipitation-dissolution reactions based on pH and other parameters like some other models (META4 and OTEQ), it can incorporate Kd values and diffusion coefficients to simulate adsorption and partitioning between the dissolved	which was clearly evident in the hydrographs for the period. Flow volumes were adjusted at gages. Given the abundance of gages, this was considered sufficient for the scale needed to represent plume movement. We will include discussion of factors affecting flow estimates in the final report. 97) We have revised the calibration of WASP settling and resuspension of total particulates based on empirical estimates of metals loads during the plume using observations at selected sites with multiple sampling. We use the constant settling velocity based on Stokes' law. To determine the settling particle size, we back-calculate the particle sizes that would result in these settling velocities that produce the loss of mass determined from the empirical modeling of mass at sites based on observed data. The methods used to estimate the loads from observations as well as how they were used to calibrate WASP will be fully described in the final report. This method of calibration suggests silt-size deposition, which appears reasonable. WASP was not calibrated to individual samples of metal concentrations sampled in what was thought to be the plume, given how limited this data actually was and lack of clarity of exactly when each sample was collected relative to the passage of the plume. 98) The WASP7 TOXI module was applied and described in the peer review, which we make clear in the final report.

Question 12 *Did the WASP modeling appropriately apply modeling parameters to estimate the movement of plume water? Please explain.*		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	and solid or particulate phases with suspended sediment in the water column and with bed sediments. It appears that this should have been done. • 99) Also, why was the WASP add-on, Metals Transformation and Assessment (META4), not used for the fate and transport modeling? This module was developed by EPA and can handle reactive transport in complex acid mine drainage-metals systems with precipitation-dissolution reactions incorporating pH and other important parameters.	We have incorporated Kd which is segment and metal specific. The Kd is a lumped parameter of dissolved metal and particulate metal to capture precipitation and sorption. We simulate total particulate metals and the individual metals As, Cu, Pb, and Zn, simulated as dissolved and particulate. Diffusion between the water column and the sediment pore water has been incorporated. WASP application with these additional elements will be fully described in the final report. 99) We reached out to Dr. James Martin, one of the WASP architects, to ask about META4. META4 is based on WASP4 and was used on mine sites in CO by Medine in the 1990s. To use META4, requires user knowledge of META4 and WASP4. We are currently using WASP7. The routines available in WASP7, such as the new hydrology routines that allow for kinematic wave modeling for streams/rivers, or the interface, are not available in earlier versions of WASP. In addition, META4 was never officially released and it does not have the level of testing that WASP7 has undergone. An update to META4 was done in 2012, but it was never been fully implemented into WASP, and is not currently available or usable. Dr. Martin recommended that META4 not be used until it is finally incorporated into the updated WASP, and then only if there is good water chemistry data. We do not use META4 for this project.

Question 12 *Did the WASP modeling appropriately apply modeling parameters to estimate the movement of plume water? Please explain.*		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	• 100) It is not clear why Scenario 1 was used. Although this scenario could provide a very conservative estimate and upper bound on transport estimates, this scenario may not be conservative if scour of metals in sediment is important, and it is too simplistic and not realistic. • 101) It is not clear what initial and boundary conditions were used for metals and sediment (Total Suspended Solids) concentrations and loads. For concentrations it appears that post-plume metals concentrations were used, whereas previous pre-release data are available and perhaps should have been used. • 102) Although sediment/particulate settling and resuspension was included in Scenario 2, it is not clear why there was no attempt to use any existing information on surface (bed) sediment metals concentrations for model input or calibration. Some of these data may have been available from previous studies to use as model input, and from data collection afterwards for calibration.	100) Scenario 1 has been removed. We now simulate a case where we only look at the total particulate load so we can determine the movement of the plume independent of other factors. 101) We set initial concentrations to 0 throughout the system so that concentrations rose and fell solely due to the passage of the release plume. This eliminates background concentrations of dissolved calcium, potassium, magnesium and sodium that are the primary constituents in the water at this time of year. The upstream boundary condition for total particulate metals in the San Juan River is upstream of the confluence with the Animas River and is based on observations in the lower Animas in the San Juan immediately downstream from the confluence. Any estimates of background concentrations used in the modeling are based on a few pre-plume samples or post-plume data collected well after the plume passed. We have evaluated post plume water quality trends and identified that adjustments to water chemistry have occurred. However, they are very small relative to the plume and should not introduce significant error to plume modeling. 102) See Answer 97. We have revised the calibration method by calibrating to computed metals mass based on observed metals data.

Question 12 ***Did the WASP modeling appropriately apply modeling parameters to estimate the movement of plume water? Please explain.***		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	• 103) In addition, there was no evaluation of sediment particle sizes from any previous data for modeling of settling and resuspension based on velocity and shear stress. It is not clear why the larger shear stresses in Scenarios 2b-2 and 2b-3 were selected and what kind of particle sizes these may relate to. • 104) For Scenario 3 for Long-term Effects, it is not clear why Nov 2010 – Dec 2011 was selected, what the magnitudes of the high flows modeled were, and how representative they are of high flows in these rivers.	103) The particle characteristics of the plume made up of precipitates of iron and aluminum oxides and perhaps gibbsite were probably unlike sediment that may have been previously measured in most parts of this system. Furthermore, the geochemical reactions during plume movement would likely have caused changing particle sizes as the plume traveled. We are not aware of any particle data collected as the plume passed. Because of the lack of data to support the erosion/settling equations, of the GKM model, we have changed our model formulation and are now using the dynamic erosion equations to improve the settling functionality velocities. We use the constant settling velocity based on Stokes' law. To determine the settling particle size, we back-calculate the particle sizes that would result in these settling velocities that produce the loss of mass determined from the empirical modeling of mass at sites based on observed data. Estimated particle size (~silt) compares well with what would be expected from precipitates in the plume. The methods and calibration will be fully described in the final report. 104) The Nov 2010 - Dec 2011 was chosen because it was the most recent year with a complete hydrological record. Coincidentally, it appears that spring snowmelt that year may be very close to that occurring in 2016. The final

Question 12 ***Did the WASP modeling appropriately apply modeling parameters to estimate the movement of plume water? Please explain.***		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	105) Using the model to estimate plume movement is good, but using Scenario 1 (due to reasons stated above) and qualitatively matching modeled to observed concentrations visually are very questionable. Why were no standard objective functions (such as Nash Sutcliffe or correlation coefficients) used? This is standard practice.	report will provide the rationale for selection of the flow record period used to model and analysis of its magnitude. 105) Scenario 1 has been eliminated. The final report will show the relationship between observed data, empirical reconstruction at selected locations and the WASP simulations. Nash-Sutcliffe is a standard metric for hydrological modeling, but not standard for water quality. In hydrological models, you have a rich data set of observed and simulated, while in water quality, you often have limited data. We have included in our updated work, the use of the correlation coefficient, R-squared.
Charles Fitts	106) Overall, the WASP modeling seemed to show that the program could be made to simulate migration and dispersion of the plume in the river that is fairly consistent with observations of the plume's passage. It appeared to be useful for simulating the approximate dilution and dispersion of contaminants, but I felt less comfortable with the analysis of deposition/resuspension since required erosion/deposition parameters were far outside of published ranges. 107) It was not clear to me how the equations for velocity, depth, width (slide 9) and Manning's equation (slide 12) were applied. Our 25 Feb phone call helped clear this up for me, but that section could use clearer explanations about how Q and V as a function of (t, river distance) were calculated. If the velocity equation on slide 9 was	106) Following the peer review discussions, we have changed our approach for settling/resuspension. See (103) 107) The power relations and their predicted exponents from regressions were used to relate depth and width of the channel geometry to flow. Flow and velocity were calculated by solving the continuity and momentum equations, incorporating Manning's equation. The figures and equations will be improved in the final report.

Question 12 ***Did the WASP modeling appropriately apply modeling parameters to estimate the movement of plume water? Please explain.***		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	not used, as Chris mentioned in our phone call, that equation should be eliminated. Also, there are corrections as noted in our discussion: the constants "a" and "b" in slide 9 equations are duplicated but should not have been, and the graph on that page needs axis labels. 108) It appears that Q was assumed constant from one gage down to the next gage, where the Q abruptly jumps up or down. It seems to me that it would be better to assume a more gradual transition of Q from one gage to the next, because the abrupt jumps in Q ripple through the calculations to cause abrupt jumps in concentrations. If you know where larger tributaries join, you could improve the assumed distribution of Q between gages using that knowledge. I understand from our phone call that WASP is limited to 50 such discharge changes, which is many more than were used in this simulation. So more, smaller jumps in Q could be incorporated into the model to give smoother, less distracting results. 109) Unless I am mistaken, the comparison shown on slide 11 means little, since the model input was constrained to match Q at gages, and all this slide shows is that the constraint worked as expected for a spring hydrograph record. 110) In slide 12 the discrepancies are not large, but they are systematic – modeled velocities are consistently high.	108) To improve the incorporation of Q into the model, we increased the number of inflows (boundaries) to refine spatial representation of tributary inflow. We distributed the change in Q by adding 3 more locations between gages in the Animas River, and added 1 more location on the San Juan. 109) This slide was presented to demonstrate exactly what the reviewer points out--that our structuring of the flow matched observations as expected. 110) We have further adjusted Manning's roughness coefficients to produce both positive and negative residuals. We will discuss how assigned Manning's roughness are high due to the averaging nature of a WASP segment, which does not capture the roughness effects associated with reach scale channel roughness including curving and braiding of the natural system.

Question 12 *Did the WASP modeling appropriately apply modeling parameters to estimate the movement of plume water? Please explain.*		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	The deviation should be explained, and the need for unusually high Manning coefficients should also be discussed and rationalized. 111) Note these typos in slides 15 and 16: the mass flux between the segments n and n+1 should be “QnCn”. Numerical dispersion is noted in slide 15. If it is possible to quantify that and compare it to simulated dispersion, that would be helpful. 112) Slide 23 shows two graphs that appear to show the same data. If there is a need for two graphs, explain what they show and how they differ. If there are time-series observed concentrations, they should be shown on these graphs for comparison. 113) In slide 41, it would help to color code the sample dots to indicate how close the measurement was to the passage of the peak plume (e.g. red for within 1 hour of peak, orange for 1-2 hours before or after peak, and so on). 114) Slide 42 should show lines of relevant criteria other than recreation, such as for aquatic life and drinking.	111) Typos will be corrected in the final report. We will also include discussion of dispersion, both real and numerical. Numerical dispersion causes the plume to have an appearance like a normal distribution, when it may be skewed and peaked. 112) Figures and graphs will be clarified in the final report. 113) We will show the distribution of concentrations at multiple sites during plume travel using color coding or other means in the final report. 114) Plume concentration estimates will be compared to water quality criteria in the final report.

Question 12 ***Did the WASP modeling appropriately apply modeling parameters to estimate the movement of plume water? Please explain.***		
Reviewer Name	**Reviewer Comment**	**EPA Response**
Henk Haitjema	115) For the most part it seems that proper use has been made of available data and, when necessary, data from the literature. However, this was not always fully explained during the presentation. Specifically, the formulas presented on slide 9, used to calculate the average water velocity, stream depth, and stream width for a particular model segment of the stream were not fully documented (and contained some erroneous coefficients). A more complete description of exactly what was measured where and how the regression analysis was applied to arrive at the coefficients "a" through "f" must be provided in the final report. 116) In follow up discussions it appeared that WASP did not use the first formula on slide 9 - the formula for velocity. These velocities were obtained using Manning's equation and calibration using observed velocities at USGS gauges, see slide 12. This is of course confusing. The velocity calculation on slide 9 is best removed. It should also be explained why the calibration on page 12 left all observed velocities below the modeled velocities.	115) We have corrected equations and will fully explain the data and method of analysis in the final report. 116) Velocities are determined with the Manning equation. Reference to the velocity estimates based on hydraulic geometry will be eliminated.
Kirk Nordstrom	117) It should probably be mentioned that there are other transport codes for this situation (e.g. OTEQ and PHREEQC) and some justification should be given why the team used WASP instead of something else. Especially in light of the fact that OTEQ has been used on mountain streams containing acid mine drainage for about 20 years and PHREEQC has been used longer than that	117) We will discuss in our presentation of WASP and model selection the availability of other models, such as OTEQ. We used WASP because we have expertise in this model and felt that it could represent the travel characteristics associated with basin hydrology (dilution), local hydraulics and deposition of a plume of AMD in a river system 600 km in length. We recognize that

Question 12 Did the WASP modeling appropriately apply modeling parameters to estimate the movement of plume water? Please explain.		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	for geochemical modeling of acid mine drainage chemistry. 118) I was glad to see that the team did not try to combine transport with reaction because there is not sufficient data to constrain such modeling.	important geochemical transformations of plume dissolved metals would take place, especially in the first hours of plume travel once the plume passed into the Animas River. The suggested META4 module would potentially have assisted with this but it is not currently publicly available or incorporated into WASP7. Such a modeling effort at this geographic scale would also have been hampered by a lack of geochemically-relevant data everywhere in the system of interest. Instead, the EPA/ORD team relies on geochemical analyses conducted partially through independent geochemical equilibria models and supported by what data was available. We believe for purposes of characterizing transport, fate, and exposure that the resolution achieved with this approach at this scale of analysis meets project objectives. 118) The EPA/ORD team agrees that to combine reaction with transport is ill-advised given the limited data.

Question 13		
Did the application of assumptions and values in WASP modeling appropriately address particle transport and deposition of the acid mine drainage constituents? Please explain.		
Reviewer Name	**Reviewer Comment**	**EPA Response**
Brian Caruso	119) This comment addresses both question 13 and 14 below, since they are related. It is not entirely clear how the application of assumptions and values in WASP modeling addressed particle transport and deposition of the acid mine drainage constituents. As stated above, although sediment/particulate settling and resuspension was included in Scenario 2, it is not clear why there was no attempt to use any existing information on surface (bed) sediment metals concentrations for model input or calibration. Some of these data may have been available from previous studies to use as model input, and from data collection afterwards for calibration. In addition, there was no evaluation of sediment particle sizes from any previous data for modeling of settling and resuspension based on velocity and shear stress. It is not clear why the larger shear stresses in Scenarios 2b-2 and 2b-3 were selected and what kind of particle sizes these may relate to. 120) It is not clear exactly what the settling results (slides 25-28) of the WASP presentation are showing or how they were computed. There appears to be total metals concentrations in the water column, and a certain fraction of this mass settles out into the sediment, creating total metals concentrations in the surface sediments, a certain faction of which can be re-suspended under high flows. But these are	119) Given an improved understanding of the system, particularly the calculated loads and decrease in load traveling downstream, As also described in (103), we have based the model formulation on a constant settling velocity rather than the from dynamic erosion algorithm presented to the review panel. Settling velocity is determined from using the empirically quantified plume mass based on observed data at selected locations. The calculated settling rates are calculated using the particle sizes that would have these settling rates. The settling velocity is applied to the reaches between calibrated locations. The EPA/ORD team feels that the movement of this plume through the system was a unique event whose conditions would not be well represented in existing data. Methods for model parameterization has generally described above will be fully explained and compared to existing data to the extent possible in the final report. 120) WASP outputs results representing sediment concentrations in non-standard units. In the final report we will report WASP output in standard units of mg/kg. Assuming low flow and suspended sediment on the day of the GKM plume on August 5, the initial concentrations of sediment were assumed to be 0 for the analysis.

Question 13 ***Did the application of assumptions and values in WASP modeling appropriately address particle transport and deposition of the acid mine drainage constituents? Please explain.***		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	presented in units of mg/L instead of mg/kg, which is typical for metals in sediment. Typically you would have particulate metals in the water column which can settle out. In addition, as stated above there appears to be no initial metals concentrations or mass in the surface sediments as an initial condition, although in some locations there should be some of this type of data from previous studies. It is also not clear if there was any initial suspended sediment loads or concentrations in the water column as part of initial conditions.	
Charles Fitts	121) The estimation of the GKM release load (WASP slide 13) is discussed in question 6 above. The WASP model simulated only the metals loads in the discharge of Cement Creek where it joins the Animas. It assumed zero metals loads in the Animas above Cement Creek and in the San Juan above the Animas. It would make for more meaningful comparisons with observations (e.g. slide 22) if estimated loads from the San Juan and upper Animas were added to the model. Adding these inputs may also impact the simulations of settling and resuspension. I imagine that reasonable estimates of the upper Animas and San Juan loads could be made from longer-term monitoring data.	121) During the summer low flow when the GKM release occurred, metals constituents in the Animas River are dominated by dissolved calcium, potassium, magnesium and sodium with low or non-detectable levels of the key metals of interest that were significantly elevated in the GKM plume. We felt that inclusion of "background" metals would not improve modeling results. However, major cations introduced with the mine effluent were treated as plume load while background concentrations of major cations were ignored. As noted by the reviewer, WASP models the GKM metals plume as if it were moving through distilled water. WASP loads were adjusted as the plume entered into the San Juan to represent the large sediment load from upstream that joined the GKM plume at the junction of the Animas and San Juan Rivers. In the updated simulations, Kd is applied as a lump parameter for precipitation reactions and sorption to the

Question 13 Did the application of assumptions and values in WASP modeling appropriately address particle transport and deposition of the acid mine drainage constituents? Please explain.		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	122) In slide 2 (and 28), the 3rd item says "…upon entering Cement Creek", but I think that should be "…where Cement Creek enters the Animas River." The last item on these slides should start with "Simulations indicated that high flow periods…" to clarify that this is a simulation result, not a measured result. 123) The simulations of sediment deposition and subsequent re-suspension indicated that a likely stretch of river to receive such deposition was from about 65km to about 95 km (WASP slides 25-27). However, the 2b-2 simulation used erosion/deposition critical shear stress thresholds that were far outside expected ranges, thus the results should be viewed only as qualitative. There should be a discussion explaining possible factors that required such large thresholds, and the degree to which these factors render the results useful or not. 124) One problem with the WASP modeling of erosion/deposition is that it must treat long stretches of river as homogeneous with respect to velocity, which is far different than the actual riffle and pool nature of the Animas River where it occupies an alluvial plain. This point is also discussed in question 7. In think plume sediment settled out in the calmer pools to a greater degree than what the WASP model	particulate metals. We estimated the contribution of particulate metals coming in from the San Juan above the Animas by comparing the total particulates in the San Juan downstream of the confluence with the Animas and the total particulates in the Animas before the San Juan. 122) These suggested phrasings were updated in the final report. 123) The treatment of resuspension/deposition has been improved. See (103) and (119). 124) The EPA/ORD team agrees that an average reach velocity used by WASP does not represent the local scale variation in velocity that exists and would likely contribute to deposition of the plume particulates as it travelled. Undoubtedly, there was local settling of colloidal/particulate precipitates and particles in the low velocity zones that occur near the bed and banks, and locally within the riffle/pool morphology, as verified in photographs. WASP is not readily applied to finer scale channel morphology and streambed features that contribute to heterogeneities in roughness and velocity. We agree that it is important to highlight this in the final report.

Question 13 ***Did the application of assumptions and values in WASP modeling appropriately address particle transport and deposition of the acid mine drainage constituents? Please explain.***		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	would predict for homogeneous segments. This may, in part, explain why atypical erosion/deposition parameters were required to make the simulation match observations.	
Henk Haitjema	125) It offered a first approximation to these processes. I appreciate the use of a simple modeling approach as conducted here in view of the limited data availability and the limited study objectives as described by the EPA team leader and the modeler. In particular, I applaud that only the most fundamental processes have been included, while secondary processes that are more difficult to parameterize have been omitted. In this light I agree with the decision to ignore physical dispersion in this modeling exercise. However, some caution is needed to declare the omission of (physical) dispersion conservatively by declaring that the modeling results provide an "upper bound." It does for the concentrations (assuming that numerical dispersion in the model does not simply replace the physical dispersion in the river or even exceeds it), but it is not conservative in predicting early arrival, for instance. The fact that numerical dispersion has not been quantified relative to physical dispersion is a weakness in this study.	125) The EPA/ORD team agrees that the simple modeling approach is the best approach given our objectives and limited data sources. We further agree that numerical dispersion does not necessarily result in an "upper bound". There is insufficient data to adequately quantify the actual physical dispersion that occurred as the plume travelled or compare it to the numerical dispersion produced by the model. We will clarify the effects of dispersion and include a discussion of uncertainties associated with the modelling in the final report.
Kirk Nordstrom	126) As a more general comment – it would seem to me that putting the metals concentrations (dissolved and total) and loads in the perspective of the range of all data for low flow conditions (or similar flow and	126) We have increased use of historical data to help put our simulation results into context and to evaluate post-event water quality after the plume passed. While the plume concentrations were very high, easily observed and quantified,

Question 13 Did the application of assumptions and values in WASP modeling appropriately address particle transport and deposition of the acid mine drainage constituents? Please explain.		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	time of year as the GKM release) would help put the plume release in better perspective. This is where historical data could help considerably. A max and a min from historical data could show some kind of envelope around or near the plume results. 127) Also, it would be better to show the individual metals, especially Cu, Zn, Pb, and As when comparing the peak concentration with river distance and conservative (no settling) scenarios. It is also not clear why some samples that look like they were sampled nearly the same time had such different concentrations. This graph needs a lot better explanation.	at least in the Animas River, the historical comparisons are especially important in trying to discern if the river system has returned to pre-event levels. 127) In response to the reviewer's comments we have added simulations for individual metals including Cu, Zn, Pb, and As to our general treatment of summed metals. The observed concentrations are presented to provide context for the simulated results. For a given date, there may be multiple samples at the same locations, collected at different times on that date. The bulk of the plume mass moved through each site rather quickly and samples collected even a few hours apart often had very different concentrations. This is fully explored elsewhere in the project analyses.

Question 14 *Did the WASP modeling appropriately investigate the remobilization of metals during increased flow? Why or why not?*		
Reviewer Name	**Reviewer Comment**	**EPA Response**
Brian Caruso	128) Please see comment on question 13 above.	128) See response to comment (103) and (119)
Charles Fitts	129) Since the mechanisms for erosion/deposition in the model were using critical shear stress thresholds that are well outside normal ranges, these results must be viewed with skepticism. Like I said in the previous comment, this makes quantifying concentrations very uncertain and this analysis should probably be viewed as an example of how downstream concentrations could respond during a high flow period, not a prediction of how they will likely respond. That distinction should be made clear in the text. 130) On slide 27, label that these are scenario 2b-2 results, and correct the title of the lower graph so it says "Movement of Total Metals…" The resuspension scenario needs to be outlined clearly. What is the assumed event? Is it a hydrograph from a typical spring runoff period, a shorter duration storm event, or something else?	129) We have changed how we handle the mechanism of settling and resuspension. See response to comment (103) and (119). We agree and will present results as a hypothesis of what could happen rather than what will happen. The amount of resuspension and the nature of suspended materials in future runoff events remains uncertain. 130) The year 2010 – 2011 hydrology record was used for the simulation. The final report will discuss how this this year relates to other years.
Henk Haitjema	131) I am not sufficiently familiar with the WASP model to adequately evaluate this point. However, I do have an observation on the reconstruction of the total metals concentration in the release flow from the mine as presented on page 13. This calculation was done outside of WASP. In principle, the mass balance calculation as presented by the three formulas on page 13 is elegant due to its simplicity. What became apparent during the discussion, however, is that the assumption that the total metals concentration in the plume as measured at 4 p.m. at the 14th St. Bridge, which is after the peak of the plume passed, is the same as during the peak flow (peak of the plume) at 12:45 may be problematic. Reviewers pointed out that the higher turbulence in the peak flow more likely than not would have caused much more materials in suspension and thus a (much)	131) The analysis presented to the panel stated that the 4:00 pm (16:00 hr) sample was not equivalent to the peak and a method was offered to estimate (raise) the concentration to a more representative value. The panel and EPA/ORD team had considerable discussion on how to do that. The final report will rectify any figures that are unclear on this point. We have also corrected an error in the algorithm for total concentration and we now assure that both the WASP and empirical modeling of the plume begin in Cement Creek at the same location and that they use the same peak value. We feel that our analysis

Question 14 Did the WASP modeling appropriately investigate the remobilization of metals during increased flow? Why or why not?		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	higher total metals concentration than what was measured at 4 p.m. I concur with that observation and emphasize that the assumption on slide 13 (concentration at 4 p.m. same as in during peak flow) might significantly underestimate the total metals load that entered the Animas River! Thus the problem with this assumption is that it is *not conservative* as to the study objective, assessing the potential impact of the release. 132) However, the total metals load in the plume further down gradient in the Animas River has been independently estimated from plume size and concentrations there. Thus, the impact assessment of the release downgradient in the Animas River is not dependent on the estimate of the original total metals load in the plume while it was still in Cement Creek. It is more likely than not that the missing portion of the total metals load in Cement Creek (due to the underestimation discussed above) ended up as sediments in the very first kilometers of the Animas River. This is because the flow velocities were quickly reduced as soon as the plume entered the Animas River thus allowing settlement of the larger particulates that might have made up the higher peak flow concentrations. 133) In summary, while I agree that the peak flow concentrations in the plume in Cement Creek may have been underestimated, I do not believe that this underestimation affected the down gradient impact assessment, except perhaps for the sediment load in the first few kilometers of the Animas River.	choices result in "conservative" estimates, but we agree that cannot be known with certainty and estimates could be too high or too low. Therefore we will remove any text that suggests our modeling choice is clearly conservative. 132) We agree that some portion of the metals and sediments entrained in what was a moderate flood in Cement Creek would have certainly deposited within the first kilometers of the slower-moving Animas River. As noted, the implication of under-estimating the peak concentration in Cement Creek is that there would have been more actual sediment deposited in the Silverton reach than modeled. The final report will discuss the uncertainty in mass and how they affect estimates of where deposition occurred. 133) We agree that any underestimation of the peak concentrations in Cement Creek would result in underestimation of deposited material between A72 and Cement Creek, but that an error would not have affected the estimated load still carried in the river past Silverton (A72). We will fully discuss the implications of this in the final report.

Part 5: Groundwater Modeling

Question 15 — Is the analysis as presented sufficient to evaluate the potential for impact of the acid mine release from the GKM on pumping wells located in the floodplain aquifers downstream of the spill?		
Reviewer Name	**Reviewer Comment**	**EPA Response**
Henk Haitjema	134) In principle it is. While a very basic groundwater code (WhAEM) was used, its limited capabilities are consistent with both the very limited data available and the limited objectives of this study. In other words, a more sophisticated model would require additional assumptions and would, therefore, not have offered more insight. To fully assess the sufficiency of the current analysis it is necessary to consider its objectives. The ultimate question to be answered is (from the presentation): "Could drinking water or irrigation wells drawing from river alluvium become impacted from the chemicals associated with the GKM release?" This question may be broken up into three interrelated questions: a) Which wells, if any, receive some of its water from the river? b) What are the travel times of water from the river to those wells? c) What is the dilution in the well of possible contaminants received from the river? Question (a) can be answered with capture zone analyses for the various wells. Question (b) can be answered by use of forward particle tracking starting at the river and ending in the well. Question (c) can be answered by tracing particles backward in time from the well, using a uniform distribution of particles around the well, and then comparing the number of path lines that reach the river to those that do not.	134) We have expanded the modeling approach and acquired additional data to better inform calibration. The refined approach creates a floodplain scale ("regional") model to set the boundary conditions for a local model capable of full 3-dimensional flow characterization applied at selected wells. The updated regional analytic element GFLOW model is calibrated to the observed static water levels at nearly 300 wells in the nearfield and baseflow estimated from USGS gages, giving effective hydraulic conductivity for the alluvium and surrounding rock, and an effective regional recharge rate. The local scale MODFLOW model with boundary conditions informed by the regional model are constructed for selected wells to better explore the local-scale complexities. This modeling combination is used to explore localized effects of factors such as heterogeneities in floodplain deposits (e.g. paleochannels), the effect of irrigation ditches in controlling floodplain water levels, as well as the impact of transient pumping and well interactions. In turn,

Question 15 Is the analysis as presented sufficient to evaluate the potential for impact of the acid mine release from the GKM on pumping wells located in the floodplain aquifers downstream of the spill?		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	WhAEM is EPA's standard model for well capture zone delineation in the context of wellhead protection and can address all three questions. As such it is a logical choice for this analysis. However, it is necessary to consider both the limitations of WhAEM and the limitations in field data, and document how these might impact the outcome of the analysis with the above research questions in mind. I will discuss these limitations in arbitrary order below. **Dupuit-Forchheimer flow** WhAEM falls in the class of codes that solves "two-dimensional flow in the horizontal plane," at least that is how these types of models are routinely referred to. Regretfully, this is misleading terminology! WhAEM is a *Dupuit-Forhheimer model*, which is a model in which resistance to vertical flow is being ignored, thus not vertical flow itself. While the underlying partial differential equation in WhAEM involves only the horizontal coordinates (x and y), flow into the vertical direction can and is being approximated using conservation of mass considerations. Consequently, path lines in WhAEM are being traced in three dimensions. For a Dupuit-Forchheimer model to offer a good approximation to the actual three-dimensional flow regime, its application must be limited to groundwater flow systems in which the horizontal distances traveled by groundwater are much larger than the vertical distances traveled. In practice, this translates into groundwater flow systems in which the distances *L* between boundary conditions (e.g. distance of the well from the river) is larger than five times the aquifer thickness. This is for isotropic aquifers. In case the aquifer is anisotropic, with a lower	insights from the local model are used iteratively to inform the Animas River floodplain-scale model. The revised modeling approach using GFLOW and MODFLOW will be fully described in the final report. The report will also contain sensitivity and uncertainty analysis.

Question 15 *Is the analysis as presented sufficient to evaluate the potential for impact of the acid mine release from the GKM on pumping wells located in the floodplain aquifers downstream of the spill?*		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	vertical hydraulic conductivity than the horizontal conductivity, the following criterion may be used (Haitjema 2006): $$L \geq 5H\sqrt{k_h/k_v}$$ Where *H* is the aquifer thickness, k_v is the vertical hydraulic conductivity, and k_h is the horizontal conductivity. Dr. Fitts suggested using a k_h *to* k_v ration of 10 for the Animas alluvium, which seems reasonable to me. The condition in the displayed formula above is not meant for wells that are relatively close to the Animas River, and unfortunately these are the wells of most interest (most likely to receive river water). What is the consequence of violating the Dupuit-Forchheimer criterion for wells near the river? In reality the well – river interaction is influenced by possible (bottom) resistance to flow between the river and the aquifer as well as resistance to vertical flow inside the aquifer. Neither is included in the model presented, although bottom resistance could have been applied. By not including any of these resistances, the flow potential for drawing water from the river that flows into the well is *overestimated*. In other words, the model as constructed *is conservative* with respect to the objectives of this study. To keep the analysis conservative in nature I recommend <u>not</u> adding bottom resistance to the line-sinks representing Animas River. <u>135) Still to be done</u>: While the analysis as conducted and presented is sound regarding this issue it must be fully documented, including calculations for representative wells to show whether they satisfy the Dupuit-Forchheimer criterion or not. This is currently missing from the	135) We have obtained better data on aquifer thickness based on a gravimetric survey of the mid Animas alluvium. As suggested by the reviewers, we re-examined the Dupuit-Forcheimer assumption for the modeled wells using improved estimates for aquifer thickness and the assumption of anisotropy in hydraulic conductivity. For one of the modeled wells, the DF assumption was violated, confirming the reviewer's assessment. The modeling approach employed in the study has been expanded to include models that can address fully 3-D flow and improve regional water table calibration (GFLOW). See (134).

Question 15 Is the analysis as presented sufficient to evaluate the potential for impact of the acid mine release from the GKM on pumping wells located in the floodplain aquifers downstream of the spill?		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	analysis! A sketch of the expected three-dimensional flow patterns and those modeled in WhAEM, as suggested by Dr. Fitts, would be useful for full disclosure of this issue. Remember, that the path lines in WhAEM are also 3D, but approximate and thus somewhat different from the “real” 3D path lines in cases where Dupuit-Forchheimer does not apply. What the sketch would not show is the fact that vertical resistance to flow is ignored in the Dupuit-Forchheimer model, which must be made clear in the figure caption. 136) Single homogeneous aquifer with horizontal base WhAEM represents the alluvium near the Animas River as a single homogenous aquifer, which means that it lumps the various depositional layers in the alluvium into a single homogenous layer. Furthermore, it assumes a horizontal aquifer base below which no flow is considered. The question is how these simplifications affect the modeling results. Specifically, what effect does this simplification have on the potential well – river interaction? This was not discussed in the presentation, but I will address this below. The actual aquifer base is unknown, but at or below the depth of the wells in the alluvium. In the absence of data it has been assumed in the current analysis that the aquifer base occurs at the bottom of the well under consideration. I agree with this choice! This will generally lead to an underestimation of the aquifer thickness, but does not affect the flow regime as much since the transmissivity in the model does not depend on this assumption because it has been based on a pump test. Assuming for a moment that the transmissivity is accurate (or reasonable) an underestimation of the aquifer thickness will result in an overestimation	136) The EPA/ORD team agrees with comments regarding the significance of the assumption of a homogeneous aquifer with horizontal base on the groundwater capture zones and particle tracking. Estimates of the base and thickness of the floodplain aquifer has been improved based on an Animas valley gravity survey (Hasbrouk Geophysics, 2003). The actual aquifer base and elevation for the local scale groundwater models of the community wells remains an estimate that will be explored through sensitivity analysis. These data and the sensitivity analysis will be described in the final report.

Question 15 Is the analysis as presented sufficient to evaluate the potential for impact of the acid mine release from the GKM on pumping wells located in the floodplain aquifers downstream of the spill?		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	of the hydraulic conductivity, since the product of the two is the (known) transmissivity. So while the discharge rates in the aquifer, including the flow component from the river if present, are not affected (question a), the *specific discharges* and associated average groundwater flow *velocities* are. An underestimation of the aquifer thickness will result in an underestimation of the groundwater travel times (question b). This is *conservative* in view of the model objective since actual early arrival of contaminants may be later than predicted by the model. The actual aquifer heterogeneity offers the potential for preferential pathways from the river to the well. The WhAEM model assumes a homogeneous aquifer that lacks preferential flow. Consequently, the assumption of homogeneity is **not** conservative in view of the model objectives. Preferential pathways would shorten the travel times from the river to the well (question b). While a multi-layer model may be able to capture this effect to some degree, data on aquifer stratification near the wells or between the wells and the river are absent. <u>137) Still to be done</u>: The above discussion must be integrated into the description of the modeling analysis to fully disclose the impact of the simplifications and assumptions. It should be pointed out that predicted early arrival times in the wells of chemicals released from the river may not preclude that some (small) portion of the chemicals arrive even earlier due to preferential flow. This is true in spite of the fact that the actual aquifer thickness may be larger than assumed and thus result in slower groundwater velocities, hence later early arrival than predicted	137) The final report will include a discussion of the implications of preferential flow pathways on breakthrough curves. A preferential flow pathway associated with a buried stream channel was introduced to the model of a community well to demonstrate the point.

Question 15 ***Is the analysis as presented sufficient to evaluate the potential for impact of the acid mine release from the GKM on pumping wells located in the floodplain aquifers downstream of the spill?***		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	by the model. Preferential flow may well outweigh the effect of the aquifer thickness on the groundwater velocities. **138) Steady state flow** WhAEM models steady state flow, ignoring water that may go into storage or is released from storage due to temporal changes in the water table (unconfined flow) or head (confined flow). For the purpose of capture zone delineation (in the context of wellhead protection), a steady state model is considered adequate. In fact, producing capture zones that change over time seems impractical for the purpose of defining wellhead protection areas. However, replacing the actual transient flow system by a steady state one raises the question what the steady state model actually represents. Haitjema (1995, 2006), using a study by Townley (1995), presents a dimensionless response time τ: $$\tau = \frac{SL^2}{4TP}$$ where S [-] is the aquifer storage coefficient, L [m] the distance between head specified boundaries, T [m^2/day] the aquifer transmissivity (product of aquifer thickness and hydraulic conductivity), and P [days] the period of a periodic forcing function. This formula differs slightly from the one presented on slide 12 due to a different definition of the distance L. When considering seasonal variations in flow in the alluvial aquifer, the definition of L on slide 12 is more convenient where it is the distance between the river and the valley boundary (rock outcrop). Haitjema (2006) offers the following rules of thumb: $\tau < 0.1$ treat transient flow in the aquifer as successive steady state.	138) The EPA/ORD agrees that the rule-of-thumb analysis using the dimensionless response time parameter, τ, provides guidance on how best to use the steady state model with the available data, or when a fully transient model is more appropriate. The formula for dimensionless response time, τ, as well as an error in pump test data for storativity were corrected. This brought τ to a range of 0.29 to 0.36, which is more in line with what would be expected in unconfined floodplain deposits. The updated analysis shows that τ is most sensitive to the assumed periodicity. If P is assumed to be 365 days, which represents the annual snowmelt forcing, then τ is sometimes < 0.1 and other times between 0.1 and 1. When τ < 0.1 this suggests putting in actual pumping rates as successive steady states. When τ is between 0.1 and 1, this suggests putting in both actual pumping rates and averaged pumping rates to bound the response. Because the community well falls in the mid-range, where transient flow may not be well represented by a steady state model, the

Question 15 Is the analysis as presented sufficient to evaluate the potential for impact of the acid mine release from the GKM on pumping wells located in the floodplain aquifers downstream of the spill?		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	$0.1 \leq \tau \leq 1$ transient flow cannot be meaningfully represented by a steady state model. $\tau > 1$ represent transient flow by a steady state model using average boundary conditions. These guidelines are approximate in that values just below 0.1 or just above 1 are to be considered transitional from the aquifer responding relatively fast or slow to transient forcing, respectively. The analysis offered on slide 12 is incomplete and partially incorrect! First of all, successive steady state simulations requires τ to be smaller than **0.1**, not 1 as stated on the slide. Incidentally, I missed this issue during the presentation. Secondly, the parameters used in the equation on slide 12 must be reassessed. Specifically, the storage coefficient *S* is too small for an unconfined aquifer as present in the alluvium surrounding the Animas River. Instead of *S*-values between 0.003 and 0.006 as shown on slide 12, I expect the *S*-value to be more in the order of 0.1. This will increase the dimensionless response time τ by almost two orders of magnitude! Note: The reason for the measured low *S*-value is unclear, although Dr. Fitts suggested that it may be an artifact of an imperfect pump test and I concur. Secondly, while the periodicity of *P*=365 days is appropriate to assess the response of the flow system to seasonal variations in recharge (in this case inflow into the aquifer near the rock outcrop) and seasonal variations in river stages, it is not suitable to assess the response of the flow system to short term variations in pumping and short term	final report will include an investigation of the steady-state vs transient assumption on capture zones for the community well that has pump test data.

Question 15 ***Is the analysis as presented sufficient to evaluate the potential for impact of the acid mine release from the GKM on pumping wells located in the floodplain aquifers downstream of the spill?***		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	variations in river stage (e.g. storm surges). For that purpose a periodicity *P*=1 day would be a better choice. This reduction in the value of *P* would further increase the value of τ indicating that the aquifer responds rather slowly to storm events and pumping variations. The current analysis does not distinguish between these long term (seasonal) and short term effects (storm events and pumping variations). Therefore, the current analysis mistakenly suggests that successive steady state is always a good approximation of this transient flow system, while in fact it most likely is not. <u>Still to be done</u>: The analysis on slide 12 has to be redone with appropriately chosen parameters and for different combinations of parameters that apply to different (well) locations *and* include bracketing values for those parameters that are not fully known from field data. None are, of course! It is likely that most of these new calculations of τ will result in values larger than 1 or at least larger than 0.1. I recommend that for values larger than 1 average pumping conditions and average river stages are used in delineating well capture zones. In addition, I suggest repeating the analysis for the actual river stage during the passing by of the plume. I suggest that for values of τ smaller than 1, but larger than 0.1, both average pumping rates and actual pumping rates are used (pumping rates when the wells were turned on). Finally, I suggest producing capture zones for bracketing values of uncertain parameters. The resulting suite of capture zones would account for (1) representing transient flow as steady state and (2) for data uncertainty.	

Question 15 ***Is the analysis as presented sufficient to evaluate the potential for impact of the acid mine release from the GKM on pumping wells located in the floodplain aquifers downstream of the spill?***		
	139) Groundwater levels and calibration In almost all cases a groundwater flow model is being calibrated using observed potentiometric heads (confined flow) or water table elevations (unconfined flow). Ideally, base flows in streams are also included as calibration targets. Calibration leads to the determination of most likely hydrogeological parameters such as hydraulic conductivities, aquifer recharge due to precipitation, and perhaps stream bottom resistances. In the current study area (or areas) almost no water level data were shown to be available. I wonder, however, if the many domestic wells in the alluvial aquifer may be on record with the state (well logs). If so, many of them might have static water levels that can be used as calibration targets. Similarly, the high capacity wells (irrigation wells and public water supply wells) may have logs that include static water levels as indeed are shown on slide 11. In the absence of domestic well static water levels, the static water levels in high capacity wells could be used as calibration targets by excluding them one-by-one from the model. This would mean conducting several calibration runs each with one of the high capacity wells replaced by a "test point" (calibration target). Currently, hydraulic gradients toward the Animas River are generated in the model by defining head specified boundaries away from the river. The water released by these head-specified boundaries presumably comes from the surrounding mountains. A common approach in modeling flow in alluvial valleys is to apply so-called "mountain range recharge" along the valley boundaries at the bottom of the surrounding mountains. In WhAEM this would be done using discharge-specified line-sinks along the base of the mountains or boundary of the alluvium. This, of course, is only possible if there are reasonable estimates available for the mountain range recharge rate. However, the measured baseflow increase along the Animas River could offer some insight into	139) The EPA/ORD team agrees with the comments on the issue of model calibration. We have expanded modeling to include regional calibration using GFLOW (see 134) using observed static water elevations reported by almost 300 wells. The GFLOW model and its conjunctive groundwater-surface water solution technique allows the additional calibration to the estimated baseflow gain between the USGS gages on the Animas River that bracket the study area. The regional model set the boundary conditions for the local capture zone models (MODFLOW) where the sensitivity analysis was expanded to include porosity and local hydraulic conductivity. We will discuss the uncertainties in groundwater flow rates towards the Animas, including accuracy of elevation data, and we will seek data that may improve estimates. A recent synoptic well water level study with LiDAR-based altimetry is available for the lower Animas River floodplain study area that will help overcome some of these limitations when modeling this area.

Question 15 Is the analysis as presented sufficient to evaluate the potential for impact of the acid mine release from the GKM on pumping wells located in the floodplain aquifers downstream of the spill?		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	this mountain range recharge rate. This baseflow increase has already been considered in the current study as shown on slide 8. Still to be done: The final report on the WhAEM model study should address the uncertainty in groundwater flow rates toward the Animas River and the lack of calibration targets. Possible data sources, as mentioned above, should be discussed and used if data is indeed available. The data uncertainty should be resolved through sensitivity testing. Different possible groundwater flow rates toward the river should be tested in the model and their effect on the capture zones (question a), early arrival times of chemicals in the wells (question b), and dilution of chemicals in the well (question c) be shown and discussed. Note: The hydraulic gradient toward the river in combination with the aquifer transmissivity defines the groundwater flow rate toward the river. This groundwater flow rate is what really matters for the shape and orientation of the capture zones, addressing questions (a) and (c). The groundwater flow rates toward the river may be bounded by considering base flow increases along the Animas River as already done on slide 8. The actual hydraulic conductivity as well as the aquifer porosity, however, affect the groundwater flow velocities and thus travel times (question b). A sensitivity analysis on hydraulic conductivity and porosity are thus also in order. References Haitjema, H.M. (1995) *Analytic Element Modeling of Groundwater Flow*. Academic Press, San Diego. Haitjema, H.M. (2006) The Role of Hand Calculations in Ground Water Flow Modeling. *Groundwater*, Vol. 44, No. 6, pages 786 – 791.	

Question 15 ***Is the analysis as presented sufficient to evaluate the potential for impact of the acid mine release from the GKM on pumping wells located in the floodplain aquifers downstream of the spill?***		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	Townley, L.R. (1995) The Response of Aquifers to Periodic Forcing. *Advances in Water Resources* 18: 125 - 146	
Charles Fitts	140) I think the analysis is sufficient to conclude that certain wells farther from the river were not susceptible to drawing river water (wells 575m71km, 650m71km, 1000m70km), and that certain wells close to the river likely do draw in river water (well 35m66km and the 5 NM wells). The analysis of well 75m71km was a closer call, and for that well a more sophisticated analysis could shed better light on the extent to which it draws in river water. The analysis conservatively estimates whether wells have potential for drawing river water. By *conservatively* I mean that the potential for drawing river water and the fraction of water drawn from the river is probably overestimated by the WhAEM models. This is discussed in more detail under question 16.	140) We agree with the assessment of the likelihood of community wells drawing source water from the Animas River. To address the issues described here, as well in Question 16, the groundwater modeling approach has been expanded to include models that can address local 3-diminesional flow paths (see 134).

Question 16 ***Were the assumptions informing the choice and construction of the groundwater flow model appropriate for the intended use? Please explain.***		
Reviewer Name	**Reviewer Comment**	**EPA Response**
Henk Haitjema	141) In general they were, but additional analyses and discussions are needed as indicated in my answer to question 15.	141) See the EPA/ORD response to Question 15 (134).
Charles Fitts	142) As stated above, the potential for communication between well and river is probably overstated in the WhAEM model results. The main reason is that the models are two-dimensional, not three dimensional. The models neglect the resistance to vertical flow and only account for resistance to horizontal flow. The models treat the river boundary condition as a fully-penetrating vertical curtain of specified head extending from the surface to the bottom of the aquifer. In reality, the river bed is a nearly horizontal constant head boundary atop the aquifer, and the well screen elevations are unknown but probably in the lower portion of the aquifer. For wells closer to the river, vertical resistance is a significant part of the total resistance to flow between the river bank and the well screen. Vertical resistance is even greater in stratified sediments where the vertical hydraulic conductivity (Kv) is much smaller than the horizontal hydraulic conductivity (Kh). For example, consider well 35m66km which is 35 meters from the river (horizontal) with a well screen that is about 25 meters below river. If Kh/Kv = 10 (ratios of 5 to 50 are common), the vertical distance in an equivalent isotropic medium would be about 80 m vertical distance (scale the vertical axis by the square root of Kh/Kv to make an equivalent isotropic medium). In this case, the vertical resistance between river and well screen would likely be greater than the horizontal resistance. Neglecting the vertical resistance in the WhAEM models overestimated the communication between well and river, and underestimated the travel time for flow from river to well. 143) The models presented are very large scale and may be omitting some important features closer to the wells. The alluvial plain has irrigation ditches, old braids, and other surface water features that likely connect to the	142) We recognize the issue of neglecting resistance to vertical flow in the analytic element model (WhAEM). We conclude based on the input of both reviewers (Haitjema, question 15) that we cannot assume 2-D flow as represented in WhAEM to investigate communication of wells near the river without independent verification. We have expanded the modeling approach to include a regional scale calibration (GFLOW) and local 3-D representation at selected wells (MODFLOW). See (134). We agree that the floodplain deposits are likely stratified with vertical hydraulic conductivity less than horizontal conductivity. We will use MODFLOW to explore the issue of aquifer stratification and anisotropy in hydraulic conductivity using a community well case study.

Question 16 *Were the assumptions informing the choice and construction of the groundwater flow model appropriate for the intended use? Please explain.*		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	underlying aquifer. Just NE of well 75m71km there is an old swampy braid that still contains water and had visible discharge into the Animas River (see close-up of Google Map image). Also, there is a braid (Coon Creek) about 1200 ft E of well 35m66km. 144) The inclusion of vast areas of rock in the far-field of this model seems unnecessary and although the model calibration process took care of critical inputs near the well fields, the largely unknown inputs for the rock and far-field may draw distracting scrutiny. I did not hear of a sound basis for these inputs: • horizontal K of the rock = 1/100th the horizontal K of alluvium • calling the rock unconfined with a horizontal base so its saturated thickness grows from around 30 meters near the alluvial aquifer to over several hundred meters to the NW (slide 15) • The no-flow boundary around the whole modeled area • The specified head boundaries to the N and W near the model limits • Zero overland flow from the steep rock areas added to the margins of the alluvium These poorly-grounded inputs could be avoided by using local scale models of just the alluvial aquifer, one model for each cluster of wells. The outermost boundary conditions could be established as specified-head boundaries where heads of surface waters or wells are known up-gradient in the alluvium, and by no-flow boundaries along estimated flow lines. 145) Slide 4 shows numerous other wells in the alluvial plain. The domestic well discharges are not likely to be large enough to have much effect, especially since most of their discharge returns to the subsurface in leaching	143) We have expanded the modeling approach (see 134) and created local models for three of the community wells to further explore the major areas of uncertainty, including, 3-dimensional flow, floodplain heterogeneities and transient pumping. The regional analytic element models provide boundary and initial conditions for the refined local models. These analyses will be included in the final report. 144) We have expanded the modeling approach to include a regional model to set the boundary conditions for the local models. The GFLOW model is calibrated to the observed static water levels at nearly 300 wells in the nearfield and baseflow estimated from USGS gages, giving effective hydraulic conductivity for the alluvium and surrounding rock, and an effective regional recharge rate. The no-flow boundary surrounding the catchment enforces the water balance. A local scale MODFLOW model set with boundary conditions informed by the regional model is then used to assess local wells. (Also see response 134.)

Question 16 *Were the assumptions informing the choice and construction of the groundwater flow model appropriate for the intended use? Please explain.*		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	fields. Irrigation wells, on the other hand, could have significant discharges, and some of these are close to municipal wells. 146) Slides 8 and 9 discuss groundwater inflows into the Animas River. The slides and the notes that were distributed did not explain how this information was incorporated into the model. Were the modeled groundwater discharges to the Animas calibrated to values from these analyses? This should be explained in the text. 147) I'm not sure the analysis of slide 12 is relevant – that applies to cyclic recharge, and recharge was not included in these models. I have no objection to omitting recharge, since at the scale of these well-river distances in late summer, lateral aquifer flows probably far outweigh recharge flows. I also think that a steady-state analysis is warranted given the limited data, roughly continuous use of municipal wells, and the close proximity of key wells to the river. 148) The 2D models presented are capable of only crudely estimating the fraction of flow that is river water and the timing of river water arrival at the wells. If there is a great need to quantify simulated concentration vs. time at a well, a more accurate approach would be to construct a more localized 3D model that includes all the closer hydrologic boundaries in the alluvial plain. A 3D flow model with enough vertical discretization to capture the resistance between the river bed and the well screen could be created with MODFLOW, AnAqSim, or any number of other 3D groundwater flow codes. This could be done at one or a few of the wells to quantify the impacts of the 2D assumption.	145) The steady-state models for the community wells were based on average pumping for the period August through December (non-snowmelt period) and assumed that the irrigation wells were not active (non-growing period). 146) We will fully describe the calibration process and data in the final report. 147) The updated regional groundwater (GFLOW) model includes areal recharge (mountain) as a source of the baseflow in the Animas River. The local model (MODFLOW) does not include areal recharge. 148) We have expanded the modeling used for this assessment to include regional flow calibration (GFLOW) and local groundwater flow at selected wells (MODFLOW). See (134).

Question 17 Were the assumptions informing the capture zone and particle tracking analysis appropriate for the intended use? How so?		
Reviewer Name	**Reviewer Comment**	**EPA Response**
Henk Haitjema	149) In general they were, but additional analyses and discussions are needed as indicated in my answer to questions 15. Specifically, I have suggested more sensitivity analyses be conducted, along the lines as to what has been shown on slide 16 and 18. The analysis as presented suggests that the aquifer responds quickly to transient forcing, allowing for successive steady state solutions or instantaneous steady state solutions. I found that analysis to be in error!! In fact, the aquifer probably responds rather slowly and a steady state model is more suitable for representing average conditions. In areas (near wells) where the aquifer may respond neither fast nor slowly, bounding solutions may be offered assuming both a fast and slow responses to transient forcing. The issue is discussed in more detail as part of my answer to question 15.	149) See response (134). The modeling approach has been expanded to better accommodate aquifer conditions near rivers, and transient vs steady state pumping. Both reviewers concluded that there were considerable uncertainties in applying WhAEM to model groundwater movement in river alluvium, especially when trying to detect a short-term interaction. The expanded modeling approach performs a floodplain-scale (regional) calibration using GFLOW and local 3-D modelling at selected wells using MODFLOW. The modeling will evaluate the effects of floodplain heterogeneities and transient forcing. The modeling will be fully described in the final report.
Charles Fitts	150) As noted under question 16 above, these 2D simulations probably overestimate the fraction of well discharge coming from the river, and probably underestimate the travel time a conservative tracer takes from river to well. The assumed porosity of 0.20 is probably low for these sediments. Such values are common in poorly sorted (widely graded) unconsolidated materials like glacial till, but for alluvial sands and gravels $0.25 < n < 0.35$ is a more common range. The impact of using a higher porosity in the simulations would be slower velocities and increased travel times between river and well.	150) A sensitivity analysis representing a range of porosities will be included in the final report. We agree with the suggested values.

Question 17 Were the assumptions informing the capture zone and particle tracking analysis appropriate for the intended use? How so?		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	151) The graphs in slides 25 and 26 would be much more striking with an arithmetic, not logarithmic Y-axis. 152) The presentation did not show models of the RK 171 or RK 179 wells in NM, so it is not possible to comment on those analyses and their conclusions. I assume the final report will provide details on those.	151) The graphs of metals concentrations will be presented with arithmetic y-axes in the final report. 152) Modeling of a second study area on the lower Animas River near Aztec, New Mexico was not completed at the time of the peer review. The same modeling approach described in (134) will be applied to this area and will be incorporated into the final report.

Question 18 ***Did the method for calibration of the local scale groundwater flow model performance to the observed drawdown reported in the driller's log serve as an effective method? Please explain.***		
Reviewer Name	**Reviewer Comment**	**EPA Response**
Henk Haitjema	153) There was a lack of calibration targets (only one USGS monitoring well), but that well has been used in this study to arrive at calibrated values of hydraulic conductivity in the alluvium, see slide 15. In my answer to question 15 I described additional *potential* calibration targets that should be explored.	153) The expanded modeling will include the regional GFLOW model of the mid Animas River floodplain with calibration to almost 300 static water levels as test points (piezometers).
Charles Fitts	154) The key factors in estimating the fraction of well discharge from the river are aquifer transmissivity (T), ambient hydraulic gradients, well discharge (Q), and well location relative to the river. The models based input T and Q values on yield test data and river/well locations on maps, which are reasonable approaches. The estimation of ambient gradients was probably the greatest source of uncertainty. Using the USGS well near well near RK 70 was helpful for constraining modeled gradients in that area. However, there are other data that could help constrain the modeled gradients elsewhere. The available aerial images and maps should be examined to determine the location and elevations of other surface waters in the vicinity of the simulated municipal wells. The modeled background gradients should be consistent with observed features, and some surface waters may need to be included as boundary conditions. For example, modeled heads should not be well above ground surface or above surface water elevations, and surface water elevations should not be far above simulated heads. This would be easier to do with smaller, local-scale models. 155) There is uncertainty in many model input parameters (e.g. T distribution, saturated thickness of the aquifer, well screen elevations, vertical conductivities), so a more thorough approach would be to run a suite of simulations (realizations) that investigate the range of possible input variations. However, if the aim of the modeling is to only determine which municipal wells have a chance of pumping a significant amount of river	154) We have refined the calibration of the regional mid Animas River floodplain model to include the static water level and baseflow calibration targets. The regional model provides boundary conditions to the local models applied to the selected community wells. 155) The EPA/ORD team agrees that the use of the simple models needs further justification. More complex local models will be used to address the uncertainties of modeling groundwater movement in floodplain alluvium (see 134). The expanded approach will be included in the final report.

Question 18 ***Did the method for calibration of the local scale groundwater flow model performance to the observed drawdown reported in the driller's log serve as an effective method? Please explain.***		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	water, the WhAEM simulations with just a single realization probably suffice to draw the broad conclusions shown in the first two "Model Results" columns in the table on slide 28, except in the case of well 75m-71km, which was a close call and could use a more sophisticated analysis to draw conclusions.	

Part 6: EnviroAtlas Modeling

Question 19 ***Are the sources of the data included in the maps valid, complete, and adequately documented?*** ***Are there any points of confusion, gaps, or suggestions for improvement?***		
Reviewer Name	**Reviewer Comment**	**EPA Response**
Brian Caruso	156) It seems as though the purpose of the Atlas modeling, and Atlas itself, were only briefly described in one slide. The purpose of the modeling and data sources were not entirely clear. For example, how do the data sources used in Atlas differ from those in any other publicly available national geospatial databases, such as USGS databases including their 'National Map'? Or do they use some of the same data sources? This all needs more explanation.	156) The EPA/ORD team utilized the EnviroAtlas tool to access background data as well as to animate study results, such as the plume travel and time series of monitoring data throughout the 600-km reach of river as a product of analysis. The input data to the animated map of the movement of the plume was the WASP modeling results. This information on data sources will be fully discussed in the final report.
Charles Fitts	157) I understand that in this section we are to comment on the animations that were presented. I can't comment on whether data sources were valid or complete, since only limited information was included on the animation slides. I think the slides were reasonably labeled and clear about what they were presenting. I found both animations to be quite helpful to visualize the plume's migration and dispersion, and to visualize the spatial variability of concentration data and the response of the system to localized precipitation events. I'm sure these animations would be even more useful to a lay person than a scientist with more background.	157) The source of the animated data will be made clearer in the maps and final report. We appreciate the comments regarding EnviroAtlas as an animation tool for depicting some of the analysis results. References to data sources will be provided in the final report and animated results will be available to the public on the web.

Question 19 Are the sources of the data included in the maps valid, complete, and adequately documented? Are there any points of confusion, gaps, or suggestions for improvement?		
Reviewer Name	**Reviewer Comment**	**EPA Response**
William Stubblefield	158) It is assumed that the sources of the data that are included in the Atlas modeling maps are the same as those described in the overview presentation and listed on page 64 of the Empirical Analyses presentation. No additional explanation for the data contained in the specific maps was provided therefore it is difficult to address the validity, completeness, and adequacy of the documentation.	158) The animated maps show results from the WASP plume modeling in one instance, and the spatial and temporal distribution of EPA water quality samples in other cases. The EPA/ORD team will ensure that the source of the data depicted in these products is clearly described in the final report.

Question 20 ***Do all of the maps and charts communicate the analysis methods and results in such a way as to be readily understood by stakeholders with interest in the impacts of the Gold King Mine spill (e.g., First Nations; NGO's; news media; and State water, recreation, public health, and wildlife managers)? Are there points of confusion, gaps or suggestions for improvement?***		
Reviewer Name	**Reviewer Comment**	**EPA Response**
Brian Caruso	159) No. The maps and charts for the Atlas modeling need more explanation as to their purpose and what the results are attempting to show and what they are actually showing. The usefulness of these maps is not yet clear.	159). The EPA/ORD team utilized the EnviroAtlas tool to access ancillary data to help put the area into context and highlight potential additional sources of metals and potential exposure. These will be clarified in the final report.
Charles Fitts	160) As noted throughout my comments, there is a need for editing, clarifying, and polishing. We were reviewing early draft figures and tables without the benefit of much additional text. The final report will have much more text to qualify and explain the tables and figures, which will be a big help.	160). We are continuing to edit and clarify the background information slides and the animations. We will explain and qualify the figures and tables in the final report.
Henk Haitjema	161) I trust that the ATLAS maps referenced here are the animations of total metals and arsenic migration along the Cement Creek, Animas River, and San Juan River. I felt that these animations were an effective way of communicating the plume migration through the system and thus quite informative for the stakeholders. Other ATLAS maps that have been used in the various presentations were also quite helpful in communicating the spatial and temporal relationships under discussion. Overall, I find ATLAS an impressive communication tool.	161) The animations of metals concentrations along the river systems were provided during the peer review. These products will be available to the public as part of the report products.

Part 7: Bioaccumulation

Reviewer Name	Reviewer Comment	EPA Response
	Question 21 ***Given the limitations of the BASS model, how appropriate is the simulation of bioaccumulation of As, Cd, Cu, Pb, and Zn in the Animas River trout fishery? What are the strengths and weaknesses of using this approach?***	
Charles Fitts	162) I am not an expert in this area, but from what I understood of the discussion, it seems that the partitioning coefficient approach that was used in this section may not be appropriate for metals that have highly regulated concentrations within organisms.	162) BASS does not use chemical partitioning (i.e., BCFs, BAFs etc.) to simulate the chemical exchanges between a fish and its food, feces, or gill water. BASS is a differential equation model that simulates the growth, population dynamics, and bioaccumulation dynamics of age-structure fish communities. Chemical uptake from food and gill water and chemical excretion to feces and gill water are simulated as individual kinetic processes. BASS uses, chemical partitioning to describe the internal distribution of chemicals between the water, lipid, and non-lipid organic matter of a fish's whole body and feces but this is not the level of partitioning discussed during the review. Also see response to comment (198). Although aquatic biota certainly regulate their internal metal concentrations by a variety of physiological processes, these concentrations are also strongly determined by the diffusion gradients that determine the excretion and uptake of metals to and from the organism's ambient water and its food/feces. These later processes are represented in detail in BASS. BASS, however, also models the physiological storage and sequestration of metals as partitioning and complexation reactions between the fish's internal water and its lipid and nonlipid organic matter. Since the Peer Review, BASS's algorithms have been updated to accommodate

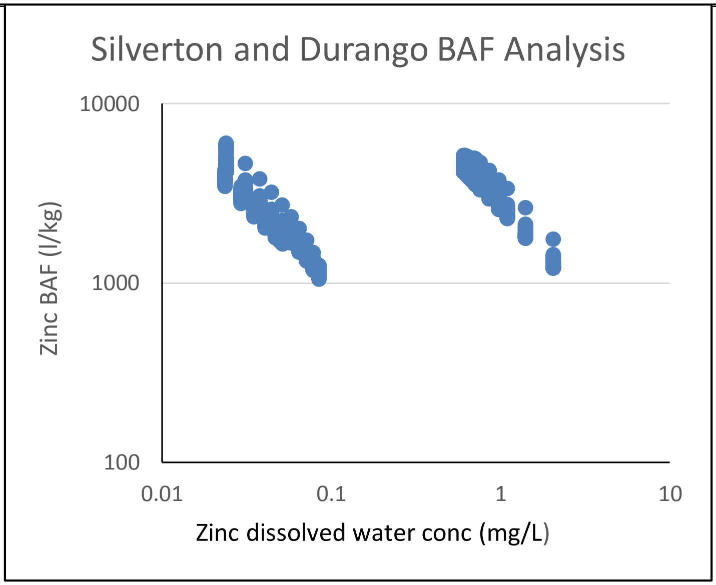
Question 21 Given the limitations of the BASS model, how appropriate is the simulation of bioaccumulation of As, Cd, Cu, Pb, and Zn in the Animas River trout fishery? What are the strengths and weaknesses of using this approach?		
Reviewer Name	**Reviewer Comment**	**EPA Response**
		saturation kinetics and have been tested using fish data collected in the Animas River just prior to the GKM release. The EPA/ORD team will improve presentation of the analytical system integrated into BASS to clarify how bioaccumulation of metals is determined.
Kirk Nordstrom	163) I was considerably puzzled by the presentation on the BASS code for several reasons. I found the model that the code is based on to be interesting and theoretically appropriate, but it seemed to me to require so much empirical information that it could be decades before it might be useful. This was confirmed by some of the discussion because there clearly is some debate in the scientific literature as to the practical application of the model. It does seem useful for certain groups of organic contaminants for which it was originally designed but not for metals toxicity. More importantly, I failed to see how this model, with parameters that would come from longer term experiments than the lifetime of the GKM plume, had any relevance. The GKM was a rapid transient event and only similarly transient experiment data with fish would be similar in application. Hence, the caged fish study and the lack of observable fish kills would seem to be the only relevant information to gage toxicity.	163) We do not agree with the conclusion that BASS requires “so much empirical information that it could be decades before it might be useful”. Currently, BASS’s auxiliary parameterization software provides users with complete default data sets of bioenergetic and ecological parameters for 691 species of North American fish. This software also can configure a default food web for any combination of these species. The input that must be provided by a user is actually rather modest. Generally, user supplied data is limited to providing 1) a limited number of chemical properties, 2) time series of dissolved water concentrations, and 3) physical features of the water body of interest (e.g., water temperature and depths). Although BASS outputs can be used to evaluate/gauge potential residue-based toxicity for fish, BASS was originally developed to assess dietary exposures to predatory fish, piscivorous wildlife, and humans for ecological and human health risk assessments; BASS was not developed to be a toxicity model per se as was the Biotic Ligand Model (BLM). Unfortunately, this fundamental difference between the BASS and the BLM was lost during much of the ensuing discussion during the peer review.

Question 21 *Given the limitations of the BASS model, how appropriate is the simulation of bioaccumulation of As, Cd, Cu, Pb, and Zn in the Animas River trout fishery?* *What are the strengths and weaknesses of using this approach?*		
Reviewer Name	**Reviewer Comment**	**EPA Response**
		We agree that observations of toxicity such as the caged fish study and measurements of body burdens are very important to understanding the acute toxicity of the GKM plume as it passed through the system. However, we do not agree that "the lack of observable fish kills would seem to be the only relevant information to gage toxicity". As with the other natural processes that were modeled, the team feels that bioaccumulation modeling adds insights not possible from fish observations alone. The modeling enables estimation of potential dietary exposures to wildlife and humans eating fish from the Animas River following the GKM release and evaluates the assumption that the plume passed too quickly for fish to have taken up metals. To this end, we used BASS to the bioaccumulation of As, Cd, Cu, Pb, and Zn before, during, and after the GKM plumes at Silverton and Durango based on BASS's tested and peer reviewed status with respect to methylmercury, which also binds to fish proteins containing sulfhydryl groups.

Question 21 ***Given the limitations of the BASS model, how appropriate is the simulation of bioaccumulation of As, Cd, Cu, Pb, and Zn in the Animas River trout fishery? What are the strengths and weaknesses of using this approach?***		
Reviewer Name	**Reviewer Comment**	**EPA Response**
William Stubblefield	164) As stated on the EPA web site, the BASS model was developed to "predict the population and bioaccumulation dynamics of age-structured fish assemblages exposed to hydrophobic organic pollutants and class B and borderline metals that complex with sulfhydryl groups (e.g., cadmium, copper, lead, mercury, nickel, silver, and zinc)." In the scope of the GKM effort, the BASS model was used to: 1) predict tissue metals concentrations in trout resulting from estimated dissolved metal concentrations in the Animas and San Juan Rivers and 2) using the estimated tissue concentrations predicted from BASS, make an assessment of "short-term impacts" to trout populations by comparing these values to residue-based tissue concentrations reported in the review by Jarvinen and Ankley (1999). However, this approach seems to be somewhat lacking for a number of reasons. 165) First, the BASS model was developed to predict tissue concentrations using a BCF/BAF approach. This assumes that the concentration of a chemical in the tissues of exposed organisms is a function of waterborne (or waterborne and food) concentrations and that the uptake of the material into tissues is a function of exposure concentration. Steady state concentrations are reached when the rate of uptake is	164) The EPA/ORD team agrees with this description of goals and objectives of the bioaccumulation modeling are what was intended. 165) There is a significant difference between a BCF/BAF approach and the bioaccumulation modeling used in this study. The BCF/BAF approach to predicting chemical concentrations in aquatic biota assumes that the ratio of an organism's whole body chemical concentration [$C_o(t)$] to the freely dissolved concentration of that chemical [$C_w(t)$] at any time t is a constant (i.e., BCF if the organism accumulates the chemical only from water, or BAF if it accumulates the chemical from both water and food).

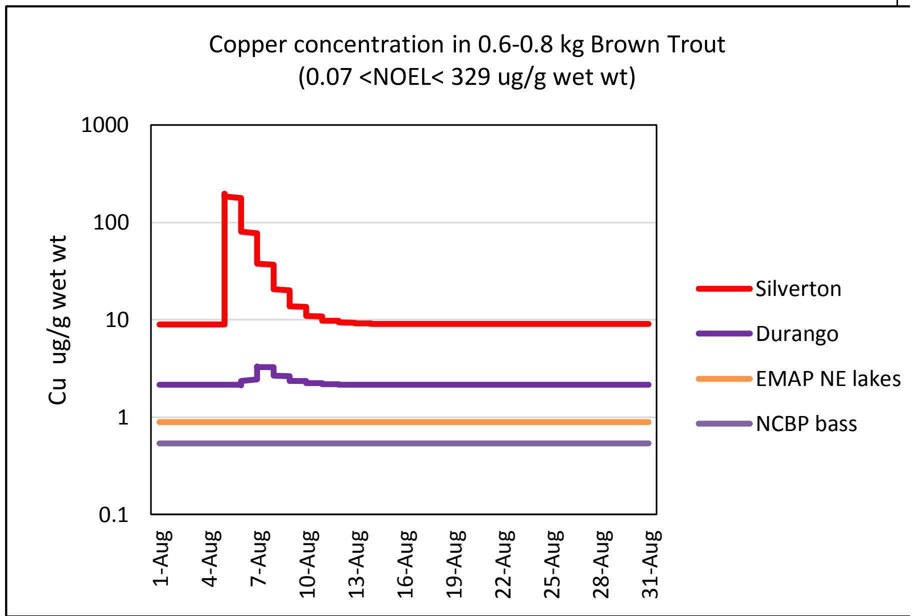
Question 21 ***Given the limitations of the BASS model, how appropriate is the simulation of bioaccumulation of As, Cd, Cu, Pb, and Zn in the Animas River trout fishery? What are the strengths and weaknesses of using this approach?***		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	equivalent to the rate of elimination. Most, if not all, of the materials that the model has been used with in the past are neutral, lipophilic organic compounds (e.g., DDT, PCB) or organometallic compounds (e.g., methyl-mercury) that follow these kinetics (i.e., concentration dependent kinetics). Metals, on the other hand, do not follow this model. Many metals are "essential" for life processes (e.g., Co, Cu, Zn) and their concentrations in the body are homeostatically controlled to maintain "constant" concentrations necessary for life processes. BCF values are calculated as the quotient of the internal tissue concentrations divided by the exposure concentration. Thus, with a metal when external concentrations are low, the body actively "concentrates" metals to maintain necessary internal concentrations resulting in extremely high BCFs; in situations when metals exposure concentrations are elevated but tissue concentrations are maintained at homeostatic levels, BCFs are low. A recent review by DeForest et al (2007) states that "Results indicate that field BAFs, like laboratory BCFs, tend to be significantly ($p \leq 0.05$) inversely related to exposure concentration" and "Data presented indicate that for metals and metalloids, unlike organic substances, no one BAF or TTF can be used to express bioaccumulation and/or trophic transfer without consideration of the exposure	Mathematically, this relationship is described by the following simple differential equation $$\frac{dC_o}{dt} = \frac{d\left(\mathrm{BAF} \times C_w\right)}{dt} = \mathrm{BAF} \times \frac{dC_w}{dt}$$ This is the simplest model used to describe the bioconcentration or bioaccumulation of chemicals in aquatic organisms and is most commonly used only for screening level risk assessments for which time dynamics can be safely ignored. BASS, however, does not use this approach to simulate chemical concentrations in fish. Rather, BASS simulates these concentrations by solving the following system of equations $$C_f = \frac{B_f}{W_f}$$ $$\frac{dB_f}{dt} = J_g + J_i - J_m$$ $$\frac{dW_f}{dt} = F - E - R - EX - SDA$$ where • C_f is the fish's chemical concentration; • B_f is the fish's chemical body burden; • J_g is the net chemical exchange across the fish's gills to and from water;

Question 21 Given the limitations of the BASS model, how appropriate is the simulation of bioaccumulation of As, Cd, Cu, Pb, and Zn in the Animas River trout fishery? What are the strengths and weaknesses of using this approach?		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	concentration." McGeer et al (2003) conclude from their review on the topic of BCFs that: • "The accumulation of Zn, Cd, Cu, Pb, Ni, and Ag in aquatic biota were, in general, remarkably consistent, particularly for Zn, where total body/tissue concentration varied little over a wide range of exposure concentrations, exposure conditions, and species. However, mean BCF values for the six metals were characterized by high variability, and there was an inverse relationship between BCF and exposure concentration. Therefore, using the weight of evidence available, it is virtually impossible to derive a meaningful BCF value that one could say is representative of the BCF for each of the metals. Even when BCFs are limited to the exposure range where chronic toxicity might be expected (based on water-quality guidelines), it is not possible to derive a precise and accurate BCF value To correctly assess potential hazards, it would be necessary to distinguish between essential nutritional accumulations, that which is sequestering and stored, and accumulation that causes adverse effects. Because BCFs are	• J_i is the net exchange across the fish's intestine to and from food/feces; • J_m is the chemical's rate of biotransformation; • W_f is the fish's body weight; • *F*, *E*, *R*, *EX*, and *SDA* are fish's rates of feeding, egestion, respiration, excretion, and specific dynamic action, respectively. The reviewer noted that several recent papers demonstrate that BCFs and BAFs are generally inversely related to exposure concentrations, and questioned BASS's ability to reproduce such relationships. A detailed analysis of the time varying BAFs [i.e., $BAF(t)=C_f(t)/C_w(t)$ where $C_f(t)$ and $C_w(t)$ denote the fish's whole body concentration and dissolved water concentration at time *t*, respectively] predicted by BASS for the Silverton and Durango plumes demonstrate that BASS indeed predicts an inverse relationship between BAF(*t*) and $C_w(t)$. (See the inserted figure below.) For example, the slope of the simple linear regression log $BAF(t)=a+b*\log C_w(t)$ for the Silverton plume was –1.01 [P(b=0) $<10^{-6}$; n=1009]. Similarly, the slope of the same regression model for the Durango plume was –1.08 [P(b=0) $<10^{-6}$; n=2663]. Similar results were obtained for Cd, Cu, and Pb at both Silverton and Durango. For arsenic, however, only the Silverton demonstrated a strong inverse relationship between BAF(*t*) and $C_w(t)$ since there was no arsenic detected as the plume passed Durango.

Reviewer Name	Reviewer Comment	EPA Response
Question 21 ***Given the limitations of the BASS model, how appropriate is the simulation of bioaccumulation of As, Cd, Cu, Pb, and Zn in the Animas River trout fishery? What are the strengths and weaknesses of using this approach?***		
	based on the whole-body concentration, the BCF model does not distinguish between these different forms of bioaccumulation and therefore it would seem unlikely that the criterion would be correlated to adverse effects such as chronic toxicity.”	 The EPA/ORD team agrees with the concerns about using a BCF or BAF approach to assess metal bioaccumulation. The problems associated with attempting to use a single BCF of BAF without reference to the fish’s aqueous exposure conditions are real and not restricted to metals; the same problems can be manifested by hydrophobic organic chemicals, especially those having “high” log K_{ow}. To overcome this problem, we used BASS to simulate the bioaccumulation dynamics of metals in Animas River fish before, during, and after the GKM release as a part of the exposure analysis. The approach will be fully documented in the final report and will be subject to another peer review at that time.

Question 21 Given the limitations of the BASS model, how appropriate is the simulation of bioaccumulation of As, Cd, Cu, Pb, and Zn in the Animas River trout fishery? What are the strengths and weaknesses of using this approach?		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	166) The second major concern with the approach employed stems from the estimation of adverse effects based on tissue concentrations (i.e., based on comparison of tissue residue levels to those reported by Jarvinen and Ankley (1999)). This approached is built upon the Critical Body Residue (CBR) concept. CBR is the concentration of chemical bioaccumulation in an aquatic organism that corresponds to a defined measure of toxicity (e.g., mortality, reproductive impairment). Rainbow, in his 2002 article, *"Trace Metal Concentrations in Aquatic Invertebrates: Why and So What?"* concludes: "Toxicity is related to a threshold concentration of metabolically available metal and not to total accumulated metal concentration." Finally, Adams et al. 2010, concluded that "Available information suggests that it is not possible to develop universally applicable whole-body CBRs for metals (except for Se, methylmercury or other organo metals[sic]). Aquatic organisms differentially handle accumulated metals with respect to storage, detoxification, and excretion. As a result, measuring total metals in an organism provides limited information on the metal concentration associated with the biologically active pool. However, the benefits of monitoring for contaminant trend and exposure assessment are acknowledged."	166) The residue-based toxicity approach is based on the simple assumption that chemical concentrations which are internal to an organism (e.g., critical body residues, CBRs) are inherently more accurate for assessing onset of toxic responses in organisms than are external environmental concentrations or benchmarks (e.g., LC_{50}, etc.). Although we firmly believe in the residue-based toxicity approach, we do not assert that there is one CBR or CBR-like threshold for each metal and fish species. Rather, depending on the endpoint of concern, there is a range of CBRs which are useful in making objective and well defined ecological risk assessments. These CBRs, however, will vary not only by the species of concern but also on that species life history (i.e., age, body weight, trophic position etc.) and exposure history. We acknowledge that it is important to discuss and quantify the uncertainties/variabilities of using this approach and we will do so in the final report. For this study, we simply assumed that if the range of BASS-predicted whole body metal concentrations of Animas River fish during and immediately after the GKM release plumes significantly overlapped with the range of CBRs reported by Jarvinen and Ankley (1999), then further analysis and study could be warranted. The BLM is a residue-based toxicity model that uses metal concentrations in the fish gill to assess acute metal toxicity. During the Peer Review, the reviewer strongly argued for its use in assessing the acute impacts of the GKM release. Although the EPA/ORD team is certainly open to investigating the utility of the

Question 21 *Given the limitations of the BASS model, how appropriate is the simulation of bioaccumulation of As, Cd, Cu, Pb, and Zn in the Animas River trout fishery? What are the strengths and weaknesses of using this approach?*		
Reviewer Name	**Reviewer Comment**	**EPA Response**
		BLM, we also remain committed to the whole body CBR approach since whole body and tissue concentrations of metals and organic chemicals in fish should be significantly correlated with one another. This assertion is certainly true for whole body and fillet concentrations (e.g., Bevelhimer, M. S., et al. 1997. Estimation of Whole-Fish Contaminant Concentrations from Fish Fillet Data. Oak Ridge, TN, Oak Ridge National Laboratory), whole body and liver concentrations (e.g., Goldstein and DeWeese 1999, JAWRA 35:1133-1140), and fillet and blood concentrations (e.g., Schmitt et al. 2009, Arch. Environ. Contam. Toxicol. 56:509-524).
	167) Finally, based on the presentations made it seems clear that the waterborne exposures following the Gold King Mine incident were reasonably short-term in nature (hours not days) and were characterized by an initial spike in concentration that dissipated rapidly returning to pre-spill conditions. Kinetics of such an exposure would suggest that steady-state whole body tissue concentrations would not have been achieved given the duration of the exposure and its variable nature. Initial impacts to organisms would likely have been acute in nature due to the initial pulse exposure. Whole body tissue concentrations would not reflect possible effects to organisms.	167) The goal of the modeling was to investigate whether the assumption expressed by the reviewer would be supported by a bioaccumulation kinetics analysis of exposure to the plume concentrations. We did not start with this assumption but set out to investigate it. The very high concentrations that characterized the plume lasted for hours but the full plume duration was more like 48 hours at the downstream locations. We will assess exposure to acute levels of metals defined by water quality criteria. However, the EPA/ORD team feels that the analysis adds additional insights gained from comparing metals and can be also be used to discuss potential human dietary exposure. We note that the bioaccumulation modeling suggested that metals would have been taken up by trout during the relatively short exposure time and also expelled quickly. The modeling also suggested that critical levels were not achieved and mortality would not be expected given this rapid uptake and depuration, consistent with the reviewer's general interpretations.
	168) If the analysis objective of the bioaccumulation and residue-based effects evaluation was to "assess the expected implications of the Gold King Mine release on Animus River biota" it would seem that a more traditional and straightforward approach to evaluating the potential impacts could be achieved by comparing estimated exposure concentrations for individual	

Question 21		
Given the limitations of the BASS model, how appropriate is the simulation of bioaccumulation of As, Cd, Cu, Pb, and Zn in the Animas River trout fishery? What are the strengths and weaknesses of using this approach?		
Reviewer Name	**Reviewer Comment**	**EPA Response**
	metals to appropriate state standards or US EPA Ambient Water Quality Criteria. References: Adams WJ, Blust R, Borgmann U, Brix KV, DeForest DK, Green AS, Meyer JS, McGeer JC, Paquin P, Rainbow PS, Wood CM. 2010. Utility of Tissue Residues for Predicting Effects of Metals on Aquatic Organisms. Integrated Environmental Assessment and Management. 7:75–98. DeForest DK, Brix KV, Adams WJ. 2007. Assessing metal bioaccumulation in aquatic environments: The inverse relationship between bioaccumulation factors, trophic transfer factors and exposure concentration. Aquatic Toxicology 84:236–246. Jarvinen AW, Ankley GT. 1999. Linkage of effects to tissue residues: Development of a comprehensive database for aquatic organisms exposed to inorganic and organic chemicals.: Society of Environmental Toxicology and Chemistry Pensacola, FL McGeer JC, Brix KV, Skeaff JM, Deforest DK, Brigham SI, Adams WJ, Green A. 2003. Inverse Relationship Between Bioconcentration Factor and Exposure Concentration for Metals: Implications for Hazard Assessment of Metals in the Aquatic Environment. Environmental Toxicology and Chemistry, 22:1017–1037 Rainbow PS. 2002. Trace metal concentrations in aquatic invertebrates: why and so what? Environmental Pollution 120:497–507	 168) We have added more analysis of the exposure for various water uses during and following the GKM event by comparing water and sediment concentrations to water quality criteria. These will be included in the final report.

References

Versar, 2016. Peer review of EPA's Gold King Mine analysis of fate and transport in the Animas and San Juan Rivers. Versar. Inc. 6850 Versar Center, Springfield, VA 22151.

SCIENCE

United States
Environmental Protection
Agency

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

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

Official Business
Penalty for Private Use
$300

Recycled/Recyclable Printed on paper that contains a minimum of
50% postconsumer fiber content processed chlorine free