United States General Accounting Office

GAO

Report to Congressional Requesters

June 2004

GULF WAR ILLNESSES

DOD's Conclusions about U.S. Troops' Exposure Cannot Be Adequately Supported

Accountability * Integrity * Reliability

GAO-04-159

GAO
Accountability•Integrity•Reliability

Highlights

Highlights of GAO-04-159, a report to Congressional Requesters

June 2004

GULF WAR ILLNESSES

DOD's Conclusions about U.S. Troops' Exposure Cannot Be Adequately Supported

Why GAO Did This Study

Since the end of the Gulf War in 1991, many of the approximately 700,000 U.S. veterans have experienced undiagnosed illnesses. They attribute these illnesses to exposure to chemical warfare (CW) agents in plumes—clouds released from bombing of Iraqi sites. But in 2000, the Department of Defense (DOD) estimated that of the 700,000 veterans, 101,752 troops were potentially exposed. GAO was asked to evaluate the validity of DOD, Department of Veterans Affairs (VA), and British Ministry of Defense (MOD) conclusions about troops' exposure.

What GAO Recommends

GAO recommends that the Secretary of Defense and the Secretary of Veterans Affairs not use the plume-modeling data for any other epidemiological studies of the 1991 Gulf War, since VA and DOD cannot know from the flawed plume modeling who was and who was not exposed. VA concurred with GAO's recommendation, but DOD did not concur.

GAO also recommends that the Secretary of Defense require no additional plume modeling of Khamisiyah and other sites. DOD concurred with GAO's recommendation.

The Central Intelligence (CIA) did not concur with the report, stating that it could not complete its review of the draft report in the time allotted.

www.gao.gov/cgi-bin/getrpt?GAO-04-159.

To view the full product, including the scope and methodology, click on the link above. For more information, contact Keith Rhodes at (202) 512-6412 or rhodesk@gao.gov.

What GAO Found

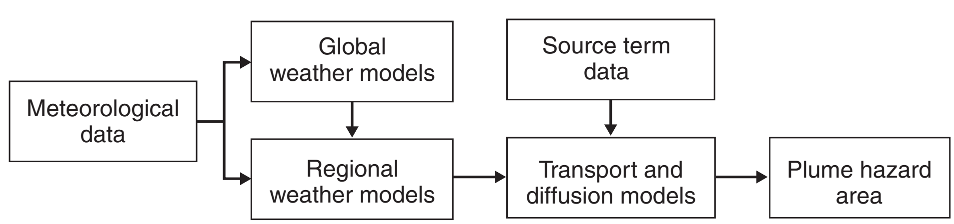
DOD's and MOD's conclusions about troops' exposure to CW agents, based on DOD and CIA plume modeling, cannot be adequately supported. The models were not fully developed for analyzing long-range dispersion of CW agents as an environmental hazard. The modeling assumptions as to source term data—quantity and purity of the agent—were inaccurate because they were uncertain, incomplete, and nonvalidated.

The plume heights used in the modeling were underestimated, and so were the hazard areas. Postwar field testing used to estimate the source term did not realistically simulate the actual conditions of bombings or demolitions. Finally, the results of all models—DOD and non-DOD models—showed wide divergences as to plume size and path.

DOD's and VA's conclusions about no association between exposure to CW agents and rates of hospitalization and mortality, based on two epidemiological studies conducted and funded by DOD and VA, also cannot be adequately supported because of study weaknesses. In both studies, flawed criteria—DOD's plume model and DOD's estimation of potentially exposed troops based on this model—were used to determine exposure. This may have resulted in large-scale misclassification.

Troops under the path of the plume were classified as exposed; those not under the path, as not exposed. But troops classified as not exposed under one DOD model could be classified as exposed under another DOD model. Under non-DOD models, however, a larger number of troops could be classified as exposed. Finally, as an outcome measure, hospitalization rate failed to capture the types of chronic illnesses that Gulf War veterans report but that typically do not lead to hospitalization

The Plume-Modeling Process



Source: GAO.

Contents

Tables

Figures

Abbreviations

ADPIC	Atmospheric Dispersion by Particle-in-Cell
AFTAC	Air Force Technical Applications Center
BSS	body-system symptom
CAPS	Clinician Administered Posttraumatic Stress
CDC	Centers for Disease Control and Prevention
CES	Combat Exposure Scale
CIA	Central Intelligence Agency
COAMPS	Coupled Ocean-Atmosphere-Mesoscale Prediction System
CW	chemical warfare
DIA	Defense Intelligence Agency
DMDC	Defense Manpower Data Center
DOD	Department of Defense
DOE	Department of Energy
DTRA	Defense Threat Reduction Agency
GDAS	Global Data Assimilation System
HPAC	Hazard Prediction and Assessment Capability
IDA	Institute for Defense Analyses
LLNL	Lawrence Livermore National Laboratory
MM5	Mesoscale Model Version 5
MOD	Ministry of Defense
NOAA	National Oceanic and Atmospheric Administration
NOGAPS	Naval Operational Global Atmospheric Prediction System
NUSSE	Non-Uniform Simple Surface Evaporation
OMEGA	Operational Multiscale Environmental Model with Grid Adaptivity
PON	paraoxonase/arylesterase enzyme
PTSD	Post-Traumatic Stress Disorder
RAMS	Regional Atmospheric Modeling System
SAIC	Science Applications International Corporation
SCIPUFF	Second-order Closure Integrated Puff
UNMOVIC	United Nations Monitoring, Verification, and Inspection Commission
UNSCOM	United Nations Special Commission
VA	Department of Veterans Affairs
VLSTRACK	Vapor, Liquid, and Solid Tracking
WMO	World Meteorological Organization

United States General Accounting Office
Washington, DC 20548

June 1, 2004

The Honorable Robert C. Byrd
Ranking Minority Member
Committee on Appropriations
United States Senate

The Honorable Christopher Shays
Chairman
Subcommittee on National Security, Emerging Threats,
and International Relations
Committee on Government Reform
House of Representatives

Many of the approximately 700,000 U.S. veterans of the Gulf War have experienced undiagnosed illnesses since the war's end in 1991. Some veterans fear they are suffering from chronic disabling conditions because of exposure to chemical warfare (CW) agents, as well as vaccines, pesticides, and other hazardous substances with known or suspected adverse health effects. They believe that their exposure may have been caused by the Coalition forces' bombing of several sites in Iraq, including storage and production facilities for nuclear, biological, or chemical warfare agents. DOD's estimates based on available bomb damage assessment during the war are that 16 of the 21 sites that were bombed were destroyed. Many U.S. and British troops were located near some of these sites. In addition, in March 1991, after the end of the war, U.S. troops conducted large-scale demolition operations to destroy munitions and facilities at Khamisiyah, a forward-deployed site in Iraq. These munitions were later found to have been filled with CW agents.

When the possible exposure of U.S. troops to low levels of CW agents first became an issue, during summer 1993, the Department of Defense (DOD) and the Central Intelligence Agency (CIA) concluded that no troops had been exposed because (1) there were no forward-deployed CW agent munitions and (2) plumes—clouds of CW agents released from the bombing that destroyed the chemical facilities—could not have reached the troops.

This conclusion was maintained until June 1996, when DOD publicly acknowledged that U.S. troops had destroyed stockpiles of chemical munitions at Khamisiyah after the war. Khamisiyah was a large ammunition storage depot that contained nearly 100 ammunition storage bunkers covering 25 sq km. Earlier, in October 1991, United Nations Special Commission (UNSCOM) inspectors had found evidence of munitions containing CW agents at Khamisiyah. Specifically, among the nearly 100 bunkers at Khamisiyah, remnants of 122 mm rockets were identified at Bunker 73. The rockets were found to have been filled with a combination of sarin and cyclosarin nerve agents. Several hundred 122 mm rockets containing the same nerve agents were also found at a pit area close to Bunker 73. It was not until 1996 that UNSCOM conclusively determined that CW agents were in Bunker 73.

Since DOD's 1996 recognition that the bombing and demolition of Iraqi facilities during the war did result in the release of plumes, DOD has conducted numerous investigations, studies, and analyses based on computer modeling. DOD has sought to determine the potential hazard area—the path of the plume—and the U.S. troops who may have been exposed to contamination from the bombing and demolition of storage facilities containing CW agents (see appendix. I).[1] In June 1996, DOD estimated that 300 to 400 U.S. troops participated in the demolition of Khamisiyah Bunker 73. In August 1996, the CIA, from the results of its computer modeling, stated that around Khamisiyah, an area as large as 25 km downwind and 8 km wide could have been contaminated.[2]

In September 1996, DOD estimated that 5,000 troops had been within a 25 km radius of Khamisiyah. In October 1996, DOD extended this radius to 50 km: It estimated that 20,000 U.S. troops had been within this zone. In July 1997, from the first plume analyses, DOD estimated that 98,910 U.S. troops had potentially been exposed. In 2000, additional analyses led DOD to reestimate that 101,752 U.S. troops had potentially been exposed.

In response to your request, we evaluated how well conclusions—about the extent of exposure of U.S. troops and the association between CW exposure and troops' hospitalization and mortality rates—are supported by available evidence. We presented our preliminary results to you in our

[1]Appendix I contains a detailed chronology of DOD's modeling events.

[2]Central Intelligence Agency, *CIA Report on Intelligence Related to Gulf War Illnesses* (McLean, Va.: Aug. 2, 1996).

testimony in June 2003.[3] In this report, we present our final results. Specifically, we have assessed the following:

1. How valid is the DOD and British Ministry of Defense (MOD) conclusion—based on CIA and DOD plume-modeling results—about U.S. and British troops' exposure to CW agents?

2. What were the total costs for the CIA's and DOD's various plume-modeling efforts?

3. How valid are the DOD and Department of Veterans Affairs (VA) conclusions from epidemiological studies, based on DOD's plume-modeling results, that there was no association between CW exposure at Khamisiyah and the troops' hospitalization and mortality rates?

Scope and Methodology

To determine the validity of DOD's conclusion, based on its plume-modeling analysis, that U.S. troops' exposure to CW agents was as low as it said it was, we examined the meteorological and dispersion models DOD used to model chemical agent releases from the U.S. demolition of Khamisiyah and Coalition bombings of Al Muthanna and Muhammadiyat in Iraq during the Gulf War. We evaluated the basis for technical and operational assumptions DOD made in the models and the specific data and information on source term, meteorological conditions, and other key parameters used for modeling chemical releases from Iraqi sites. We also evaluated the efforts of the CIA and DOD to collect and develop data on source term and other key parameters for the modeling.

We interviewed DOD and CIA modelers and officials involved with the modeling and obtained documents and reports from DOD's Deployment Health Support Directorate. We also interviewed and received documents from Department of Energy (DOE) officials who were involved with the modeling at DOE's Lawrence Livermore National Laboratory (LLNL). In addition, we interviewed officials and obtained documents from the

- Institute for Defense Analyses (IDA) concerning the IDA expert panel assessment of CIA's modeling of Khamisiyah,

[3] U.S. General Accounting Office, *Gulf War Illnesses: Preliminary Assessment of DOD Plume Modeling for U.S. Troops' Exposure to Chemical Agents*, GAO-03-833T (Washington, D.C.: June 2, 2003). www.gao.gov.

- U.S. Army Dugway Proving Ground in Utah to determine how CW agents in Iraq's rockets might have been released during the Khamisiyah pit area demolitions,
- U.S. Army Center for Health Promotion and Preventive Medicine to determine how specific troop unit exposures were identified, and
- United Nations Monitoring, Verification, and Inspection Commission (UNMOVIC) to obtain information on source term from UNSCOM's analyses and investigations on Khamisiyah, Al Muthanna, and Muhammidiyat.

To determine the validity of DOD's and VA's conclusions—based on epidemiological studies—that there was no association between Khamisiyah exposure and the rates of hospitalization or mortality, we reviewed published epidemiological studies in which hospitalization and mortality among exposed and nonexposed U.S. troops were analyzed. We also interviewed the study authors and researchers involved with the studies and examined the Gulf War Khamisiyah population databases DOD provided in support of these studies. We interviewed Veterans Benefits Administration officials and obtained documents and reports on their analysis of DOD's population databases. We did not examine whether plume modeling data were being used by VA to determine eligibility for treatment or compensation.

To identify total costs associated with modeling and analysis of CW agent releases during the Gulf War, we interviewed officials and collected cost data from various DOD agencies and contractors who supported the modeling efforts. However, total costs incurred could not be determined because DOD agencies and other organizations could provide only direct costs, not their indirect costs, associated with the modeling.

To determine the extent of British troops' exposure to CW agent releases during the Gulf War, we interviewed MOD officials in London and Porton Down, and we reviewed MOD reports concerning the effect of exposure to CW agent releases on British forces.

We conducted our work from May 2002 through May 2004 in accordance with generally accepted government auditing standards.

Results in Brief

DOD's and MOD's conclusions, based on DOD's plume modeling, about their troops' exposure to CW agents cannot be adequately supported.

Given the inherent weaknesses associated with the specific models they used and the lack of accurate and appropriate meteorological and source term data in their analyses, we found five major reasons to question their conclusions. First, the models were not fully developed for analyzing long-range dispersion of CW agents as an environmental hazard. Second, assumptions regarding source term data used in the modeling—such as the quantity and purity of the agent—were inaccurate, since they were based on (1) uncertain and incomplete information and (2) data that were not validated. Third, the plume heights from the Gulf War bombings were underestimated in DOD's models. Fourth, postwar field testing at the U.S. Army Dugway Proving Ground to estimate the source term data did not reliably simulate the actual conditions of either the bombings or the demolition at Khamisiyah. Fifth, there is a wide divergence in results among the individual models DOD selected, as well as in the DOD and non-DOD models, with regard to the size and path of the plume and the extent to which troops were exposed. Therefore, given these inherent weaknesses, DOD and MOD cannot know which troops were and which troops were not exposed.

The total costs for the various plume-modeling efforts to analyze the potential exposure of U.S. troops—from the demolition at Khamisiyah and the bombing of several sites in Iraq—cannot be estimated. DOD organizations and other entities involved with the plume-modeling efforts could provide only their direct costs (that is, contractors' costs), which totaled about $13.7 million. However, this amount does not include an estimate of the considerable indirect costs associated with the salaries of DOD, VA, and contractors' staff or costs of facilities, travel, and equipment. We requested, but DOD could not provide, this estimate. In addition, the CIA would not provide direct and indirect costs for Gulf War plume modeling because, in its view, our request constituted oversight of an intelligence matter and is beyond our scope of authority. The CIA's contractor, the Science Applications International Corporation (SAIC), also did not respond to our request.

DOD's and VA's conclusions—that there was no association between exposure to CW agents at Khamisiyah and U.S. troops' rates of hospitalization and mortality—also cannot be adequately supported. DOD and VA based their conclusions on two epidemiological studies they funded, one conducted by DOD researchers, the other by VA researchers.[4] In both of these studies, flawed criteria were used to determine the troops' exposure. For example, the studies' criteria were based on (1) the DOD plume model of exposure from postwar demolition of the Khamisiyah munitions depot and (2) DOD's estimation, using this modeling, of which troops might have been under the path of the plume. Troops under the path of the plume were classified as exposed, those not under the path as nonexposed. However, troops classified as nonexposed under one DOD model could be classified as exposed under another DOD model, thereby confounding the results. In the DOD models, a small area was identified as being under the path of the plume, resulting in a small number of troops identified as exposed. But in a model DOD used from LLNL, for example, a much larger area was identified under the path of the plume, resulting in a large number of troops exposed; in addition, these exposed troops included different troops from those in the DOD models.

These flaws may have resulted in large-scale misclassification of the exposure groups—that is, a number of exposed veterans may have been classified as nonexposed, and a number of nonexposed veterans may have been misclassified as exposed. In addition, in the hospitalization study, the outcome measure—number of hospitalizations—failed to capture the chronic illnesses of Gulf War veterans, which are commonly reported but typically do not lead to hospitalization. Some published scientific studies of exposure involving Gulf War veterans, studies of genetic anomalies, and animal exposure-response studies suggest an association between low-level exposure to CW agents and chronic illnesses.

We recommend that the Secretary of Defense and the Secretary of Veterans Affairs not use the plume-modeling data for future epidemiological studies of the 1991 Gulf War, since VA and DOD cannot know from the flawed plume modeling who was and who was not exposed.

We recommend that the Secretary of Defense require no further plume modeling of Khamisiyah and the other sites bombed during the 1991 Gulf War in order to determine troops' exposure. Given the uncertainties in the

[4] G. C. Gray and others, "The Postwar Hospitalization Experience of Gulf War Veterans Possibly Exposed to Chemical Munitions Destruction at Khamisiyah, Iraq," *American Journal of Epidemiology* 150 (1999); H. K. Kang and T.A. Bullman, "Mortality among U.S. Veterans of the Persian Gulf War: 7 Year Follow-Up, "*American Journal of Epidemiology* 154 (2001): 399-409.

source term and meteorological data, additional modeling of the various sites bombed would likely result in additional costs while still not providing DOD with any definitive data on who was or was not exposed.

We obtained comments on a draft of this report from VA, DOD, and CIA. VA concurred with our first recommendation (see appendix V). DOD did not concur with our first recommendation, indicating that it apparently represents a blanket prohibition of plume modeling in the future. The intent of our recommendation is only directed at epidemiological studies involving the DOD and CIA plume modeling data from the 1991 Gulf War and not a blanket prohibition of plume modeling in the future (see appendix VI). We have clarified the recommendation along these lines. DOD concurred with our second recommendation, indicating that despite enhancements in the modes, uncertainties will remain (see appendix VI). CIA did not concur with the report, indicating that it could not complete their review in the time allotted (see appendix VII).

Background

In March 1991, after the Gulf War had ended, U.S. Army demolition units destroyed munitions in Bunker 73 and in an open pit at the Khamisiyah ammunition storage depot in southeastern Iraq. In October 1991, it was discovered, from inspections conducted by UNSCOM that hundreds of 122 mm rockets destroyed at Khamisiyah had contained nerve agents sarin and cyclosarin.

U.S. and Coalition forces also bombed many other known or suspected Iraqi CW research, materiel, storage, and production sites. According to the CIA and DOD, Coalition aerial bombings resulted in damage to filled chemical munitions at only two facilities in central Iraq—Al Muthanna Bunker 2 and Muhammadiyat—and at one facility in southern Iraq—the Ukhaydir ammunition storage depot. During these aerial bombings, munitions were damaged at Al Muthanna containing an estimated 17 metric tons of sarin and cyclosarin and at Muhammadiyat containing an estimated 2.9 metric tons of sarin and cyclosarin and 15 metric tons of the chemical agent mustard.

According to DOD, the connection between the CW agent munitions UNSCOM found at Khamisiyah and U.S. demolition operations there had not been immediately made. However, in 1996, concerns that the Presidential Advisory Committee on Gulf War Illnesses raised prompted the CIA to examine this issue.[5] In early 1996, after linking Khamisiyah to the presence of CW agents, based on UNSCOM and other reporting, the CIA contracted with SAIC to conduct an initial analysis and modeling of the bombing of chemical munitions in Khamisiyah Bunker 73.

The CIA issued two reports. The first report, in August 1996, modeled a potential release of agents from Khamisiyah Bunker 73 and from Al Muthanna and Muhammadiyat.[6] However, when modeling of the pit area at Khamisiyah began, the CIA realized that the source term data—such as the quantity and the purity of the agent and data on meteorological conditions, including the wind and the weather patterns—were not available. Because of these uncertainties, DOD and the CIA asked IDA to convene an independent panel of experts to review the modeling. The IDA expert panel conducted its review from November 1996 to February 1997 and (1) reported that the initial analyses of the pit area were inadequate and (2) recommended taking different approaches to improve the modeling.

The CIA issued its second report jointly with DOD in September 1997, this time focusing on an open pit area at Khamisiyah.[7] This report combined the results of five different meteorological and dispersion models that the CIA and DOD used to project the size and path of the plume from the demolition operations. In 2000, DOD remodeled the Khamisiyah pit site, using updated CIA source assessments and troop location data, which changed the projected hazard area. The 1991–2000 timeline of major events at Khamisiyah is shown in figure 1.

[5]The Presidential Advisory Committee on Gulf War Veterans' Illnesses was established by Executive Order 12961 on May 26, 1995, to provide oversight for Gulf War illness investigations; it terminated in November 1997. http://www.gwvi.ncr.gov/ (Apr. 26, 2004).

[6]Central Intelligence Agency, *CIA Report on Intelligence.*

[7]Central Intelligence Agency and the Department of Defense, *Modeling the Chemical Warfare Agent Release at the Khamisiyah Pit* (McLean, Va.: Sept. 1997).

Figure 1: Timeline of Events Following the U.S. Demolition of CW Agents at Khamisiyah, 1991–2000

1991

March:
U.S. troops conduct large-scale demolition operations to destroy ammunition bunkers and warehouses at the Khamisiyah munitions storage site in southern Iraq.

October:
United Nations Special Commission (UNSCOM) inspection team find the presence of sarin-filled 122mm rockets at the Khamisiyah munitions storage site.

November:
U.S. intelligence and DOD become aware of the UNSCOM findings, but no action is taken.

1992

February and March:
UNSCOM destroy 463 sarin- and cyclosarin-filled 122 mm rockets at Khamisiyah.

1993

June:
U.S. Navy personnel testify before the U.S. Senate Committee on Armed Forces to possible chemical exposures following explosions and chemical agent alarm soundings on January 19, 1991, at the Port of Al Jubayl, Saudi Arabia.

July:
The Czech Minister of Defense announced that a Czechoslovak chemical decontamination unit detected sarin in areas of northern Saudi Arabia during the early phases of the 1991 Persian Gulf War.

1994

May:
In testimony before the U.S. Senate Committee on Banking, Housing, and Urban Affairs, DOD and intelligence community witnesses admit that USCOM found chemical weapons at Khamisiyah during its inspections.

1995

March and June:
The CIA begins reexamination of relevant intelligence, and DOD forms the Persian Gulf Illnesses Investigation Team.

October:
The DOD investigation team identifies some of the U.S. forces that had occupied the area around Khamisiyah.

1996

June:
DOD confirms that some of the munitions destroyed by U.S. troops in March 1991, at Khamisiyah, were filled with the chemical warfare agents sarin and cyclosarin.

November–February 1997:
The CIA and DOD task the Institute for Defense Analysis to establish an expert panel, so as to assess the models being used to estimate the potential hazard area from the destruction of the chemical warfare agent munitions at Khamisiyah.

1997

September:
The CIA and DOD present their first composite model, delineating the size, path, and exposure levels of the potential hazard area.

2000

January:
The CIA and DOD complete a remodeling and present a revised composite model that redefines the 1997 hazard area.

Source: GAO.

Modeling the Environmental and Health Effects of Fallout from CW Agents

According to the CIA, modeling is the art and science of using interconnected mathematical equations to predict the activities of an actual event. In this case, modeling was used to determine the direction and extent of the plume from CW agents. In environmental hazard modeling, simulations recreate or predict the size and path (that is, the direction) of the plume, including the potential hazard area, and potential exposure levels are generated.

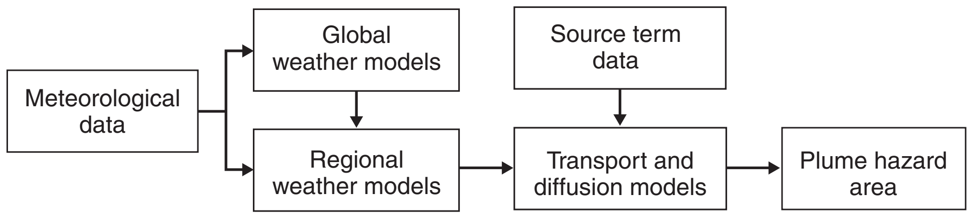
Modeling requires several components of accurate information:

- the characteristics or properties of the material that was released and the rate of release (for example, quantity and purity, the vapor pressure, the temperature at which the material burns, particle size, and persistency and toxicity),
- temporal information (for example, whether chemical agent was initially released during daylight hours, when it might rapidly disperse into the surface air, or at night, when a different set of breakdown and dispersion characteristics would pertain, depending on terrain, plume height, and rate of agent degradation),
- data that drive meteorological models during the modeled period (for example, temperature, humidity, barometric pressure, dew point, wind velocity and direction at varying altitudes, and other related measures of weather conditions), and
- data from global weather models to simulate large-scale weather patterns and from regional and local weather models to simulate the weather in the area of the chemical agent release and throughout the area of dispersion.

In addition, in modeling, information is required on the potentially exposed populations, animals, crops, and other assets that may be affected by the agent's release.

The process flow for chemical plume modeling, to estimate the plume hazard area, is shown in figure 2.

Figure 2: The Plume-Modeling Process



Source: GAO.

Meteorological Models

Various plumes during the 1991 Gulf War were modeled with global-scale meteorological models, such as the National Centers for Environmental Prediction Global Data Assimilation System (GDAS) and the Naval Operational Global Atmospheric Prediction System (NOGAPS). Regional and local weather models were also used, including the Coupled Ocean-Atmosphere Mesoscale Prediction System (COAMPS), the Operational Multiscale Environmental Model with Grid Adaptivity (OMEGA), and the Mesoscale Model Version 5 (MM5). Outputs from global weather models are mainly used as initial and boundary conditions for regional weather models.

Transport and Diffusion Models

Transport and diffusion models were also used during the Gulf War.[8] These models project both the path of a plume and the degree of potential hazard posed by the agents. Dispersion models used during the Gulf War included the Hazard Prediction and Assessment Capability (HPAC) model, along with the HPAC component, the Second-order Closure Integrated Puff (SCIPUFF) model; the Vapor, Liquid, and Solid Tracking (VLSTRACK) model; the Non-Uniform Simple Surface Evaporation (NUSSE) model; and the Atmospheric Dispersion by Particle-in-Cell (ADPIC) model.

The meteorological and dispersion models DOD used to model Khamisiyah are shown in table 1.

[8] We use dispersion in this report to refer to both transport and diffusion models.

Table 1: The Meteorological and Dispersion Models DOD Used to Model Khamisiyah

Model type	Developer or sponsor
Meteorological	
COAMPS: Coupled Ocean-Atmosphere Mesoscale Prediction System	U.S. Navy
MATHEW: Mass Consistent Wind Field[a]	Department of Energy, Lawrence Livermore National Laboratory
MM5: Mesoscale Model Version 5	National Center for Atmospheric Research
NOGAPS: Naval Operational Global Atmospheric Prediction System,	U.S. Navy
OMEGA: Operational Multiscale Environment Model with Grid Adaptivity	Defense Threat Reduction Agency
Dispersion	
ADPIC: Atmospheric Dispersion by Particle-in-Cell[a]	Department of Energy, Lawrence Livermore National Laboratory
HPAC/SCIPUFF: Hazard Prediction and Assessment Capability/Second-order Closure Integrated Puff	Defense Threat Reduction Agency
NUSSE4: Non-Uniform Simple Surface Evaporation, Version 4	U.S. Army
VLSTRACK: Vapor, Liquid, Solid Tracking	U.S. Navy

Source: Department of Defense, Technical Report: Modeling and Risk Characterization of U.S. Demolition Operations at the Khamisiyah Pit (Washington, D.C.: Apr. 2002).

[a]Institute of Defense Analysis used this model in its analysis of Khamisiyah modeling, an analysis done at DOD's request.

DOD's Conclusion about U.S. Troops' Exposure to CW Agents Cannot Be Adequately Supported

DOD's conclusion about the extent of U.S. troops' exposure to CW agents from the Gulf War, based on CIA and DOD plume models, cannot be adequately supported because of uncertainty associated with the source term data and meteorological data. The models are neither sufficiently certain nor sufficiently precise to draw definitive conclusions about the size or path (that is, the direction) of the plumes.

In particular, we found five reasons to question DOD's conclusion. First, the models DOD selected were not fully developed for analyzing long-range dispersion of CW agents as environmental hazards. Second, the assumptions about the source term data used in the models were inaccurate. Third, in most of the models, the plume height was significantly underestimated. Fourth, postwar field testing at the U.S. Army Dugway Proving Ground in Utah to estimate the source term data did not realistically simulate the actual conditions of the demolition operations at Khamisiyah or the effects of the bombings at any of the other sites in Iraq. And fifth, there are wide divergences among the individual models DOD selected with regard to the size and path of the plume and the extent to which troops were exposed.

DOD's Models Were Not Fully Developed for Analyzing Long-Range Environmental Hazards

CIA and DOD officials selected several in-house models for plume models. For Khamisiyah and the other Iraqi sites, DOD selected the COAMPS and OMEGA meteorological models and the HPAC/SCIPUFF and VLSTRACK dispersion models. However, at the time, these models were not fully developed for long-range environmental hazards. In particular, they were inappropriate for the long-range tracking of chemical agents.

COAMPS and OMEGA Meteorological Models

Although COAMPS was accepted by the DOD peer review panel, OMEGA was still under development. The DOD 1997 peer review panel that was reviewing the models chosen for the 1997 Khamisiyah analysis reported problems with OMEGA that resulted in major errors in its simulations. In the analyses of Khamisiyah, as well as Al Muthanna, according to an IDA technical review panel, OMEGA consistently underpredicted surface wind speeds by a factor of 2 to 3, compared with actual observations collected at five World Meteorological Organization (WMO) stations in the area.

VLSTRACK and HPAC Dispersion Models

VLSTRACK had been developed primarily to predict immediate health hazards (that is, lethal or incapacitating effects) associated with troops' direct exposure to CW agents. It was not developed for predicting long-term health effects from indirect exposure to low levels of these agents. According to modeling experts at the Naval Surface Warfare Center, the 2-month IDA panel reanalysis and modeling was a developmental effort because existing models did not have the capability to perform the required predictions. Considerations of dispersion areas associated with low-level exposure to CW agents released in Iraq, such as nerve and blister agents, required these experts to make many extensions and modifications to some of the methodology in VLSTRACK.

HPAC was developed jointly by the Defense Intelligence Agency (DIA) and the Defense Special Weapons Agency—now the Defense Threat Reduction Agency (DTRA)—and was specifically tailored for counterproliferation contingency planning. In a 1998 scientific review and evaluation of the SCIPUFF dispersion model (an integral component of HPAC), the National Oceanic and Atmospheric Administration's (NOAA) Air Resources Laboratory reported that SCIPUFF was better suited for short-range dispersion applications of about 10 km than for long-range transport modeling. One of HPAC's weaknesses is that it does not provide definitive answers because of uncertainties about agent transport, plume location, and weather.

It is also evident, from NOAA's review, that a group using VLSTRACK might receive significantly divergent results from a group using HPAC. Further, neither model is sufficiently accurate to permit a conclusion that the path of the plume is confined to the hazard area that the model predicts.

In a September 1998 memorandum, the Deputy to the Secretary of Defense for Counterproliferation and Chemical and Biological Defense cited a DOD panel study team finding that the VLSTRACK and HPAC models generate hazard predictions that differ significantly from each other: "This occurred even when the source terms and weather inputs are as simple and as identical as possible. In operational deployment, the average model user could obtain different answers for the same threat."

A former Modeling and Simulation Adviser to the Deputy for Counterproliferation and Chemical and Biological Defense told us that the reliability of these models was of extreme concern. Also, in 1998, at IDA, a panel study's initial comparison of the hazard-prediction models HPAC and VLSTRACK documented substantial differences—by factors between 5 and 1,000—between the models for the prediction of the same event. The most significant errors in the coding and the potential for misuse were found in HPAC and its subcomponent models. Given these problems with the analyses conducted up to 1998, HPAC could not be considered reliable.

In 1997, the director of NOAA's Air Resources Laboratory stated that DOD's model selection resulted in the use of in-house models that were not well known outside DOD. As to the meteorological models, he noted that three mainstream mesoscale models were available, such as MM5 and Regional Atmospheric Modeling System (RAMS), that were well accepted for deriving site-specific flow conditions from large-scale meteorological information. DOD used MM5 in both its 1997 and 2000 modeling.

DOD's Models Included Source Term Assumptions That Cannot Be Verified

The source term data DOD used in the modeling for sites at Khamisiyah, as well as Al Muthanna and Muhammadiyat, contain significant unreliable assumptions. DOD and the CIA based assumptions on field testing, intelligence information, imagery, UNSCOM inspections, and Iraqi declarations to UNSCOM. However, these assumptions were based on limited, nonvalidated, and unconfirmed data concerning (1) the nature of the Khamisiyah pit demolition, (2) meteorology, (3) agent purity, (4) amount of agent released, and (5) other CW agent data. In addition, DOD and the CIA excluded, for modeling, many other sites and potential hazards associated with the destruction of binary chemical weapons, CW agent precursor materials, and the potential release of CW agents and toxic byproducts from other sites.[9]

During the initial modeling of the demolition of CW agent munitions at the Khamisiyah pit, the CIA did not have the data—the number of rockets present, agent purity, amount of agent released into the atmosphere, agent reaction in an open-pit demolition, and prevailing meteorological conditions—on how rockets with chemical warheads would be affected by open-pit demolition compared with bunker demolition. A 1996–97 IDA panel found substantial uncertainty about the number of damaged rockets that might have released CW agents and how rapidly they were released. The panel also found that the CIA and SAIC had used what were, "essentially, guesses" to make up for the lack of data on how much agent was released and over what period of time. For example, the CIA based the number of rockets on the number known to have been there before the demolition and the number UNSCOM found during inspections. But, according to the IDA panel, the difference between what was estimated and what UNSCOM found varied by a factor of 5 or 6.

Meteorological data at Khamisiyah were lacking because relatively few observations had been made, according to DOD modelers, and those that had been made were far from the site. This lack of meteorological observation applied also to the modeling of other sites. Observations were few because Iraq stopped reporting weather station measurement information to WMO in 1981. As a result, data on the meteorological conditions in Iraq during the Gulf War were not available.

At Khamisiyah, the only data available were for the surface wind observation site, 80 to 90 km away, and for the upper atmosphere, about 200 km away. At other sites modeled, the nearest data were at even greater distances. The IDA panel recognized that wind patterns could contain areas of bifurcation—lines where winds move in one direction on one side and in another direction on another side—which migrate over time and are different at varying altitudes.

Assumptions about the purity of the CW agents sarin and cyclosarin established for Khamisiyah, Al Muthanna, Muhammadiyat, and Ukhaydir differed widely. In each case, agent purity was a key factor in the CIA and

[9] A binary weapon mixes two less toxic materials to create a toxic nerve agent within the weapon when it is fired or dropped.

DOD methodology for determining the amount of agent released. For example, for modeling purposes, 10 tons of agents with a purity of 18 percent would be represented as 1.8 tons of agent. The CIA relied on UNSCOM reporting on the amount of CW agents Iraq produced. But to establish these rates, UNSCOM relied primarily on Iraqi declarations and Iraqi production records, other UNSCOM testing, and assumptions about the extent of agent degradation.

For example, according to Iraqi production records UNSCOM obtained, the agent purity at Khamisiyah in early January 1991 was about 55 percent. The agent degraded to 10 percent purity by October 1991, when laboratory analysis had been completed on samples taken by UNSCOM from one of the rockets. On the basis of the initial sample purity and the projected degradation rate for sarin and cyclosarin, the CIA assessed that the ratio of sarin to cyclosarin when the munitions were blown up, in March 1991, was the same as that sampled in October 1991—3:1. According to the CIA, "assuming a conservative exponential degradation" of the sarin and cyclosarin, the purity on the date of demolition, 2 months after production, was calculated at about 50 percent.

At Al Muthanna, where sarin was stored in a bunker, the CIA estimated that it had deteriorated to approximately 18 percent purity by early February 1991, when Bunker 2 was destroyed, leaving the equivalent of about 1,600 kg (1.6 metric tons) of viable sarin. Iraqi records recovered by UNSCOM inspectors suggested that the agent in Bunker 2 was from 1988 production runs. UNMOVIC confirmed that in UNSCOM testing of other sarin samples, produced during 1998, the maximum purity of agent had degraded to a range of 18 percent to 2 percent by 1991.

According to a 2002 CIA assessment, only 2.5 percent of sarin was released from the demolition of rockets in Bunker 73 at Khamisiyah.[10] This assessment was based on comparisons with U.S. testing in the 1960s at Black Hills, South Dakota, on sarin-filled rockets.[11] The CIA considered the 2.5 percent estimate conservative because agent-heating conditions were

[10]Central Intelligence Agency, *Intelligence Update: Chemical Warfare Agent Issues during the Persian Gulf War* (McLean, Va.: Apr. 2002).

[11]U.S. Army Dugway Proving Ground, *Hazard Classification Tests of Igloo Storage of GB- and VX-Filled M55 Rockets*, RTD&E Project 1B650312D624, USATECOM Project 5-3-0135-02, DPGDR C-505 (Dugway, Utah: Apr. 1966).

harsher in Bunker 73 than in the Black Hills tests. The 2002 CIA assessment included this statement:

> Far less agent (a maximum of 0.01 percent) would have been released in the Al Muthanna bunker incident than the 2.5 percent indicated by 1960s US field tests at Black Hills, South Dakota. The Black Hills tests used simulated bunkers that had a wood slat, sand ceiling, and earth walls. Those bunkers did not allow heat to build up as rapidly as in an Iraqi bunker with thick reinforced concrete ceiling and walls. However, we have chosen 10 kilograms as the release amount to account for unmodeled releases from rocket flyouts or transients at the beginning of the fire.[12]

The CIA assessed that far less agent would have been released in the Al Muthanna bunker because, based on U.S. field testing with simulated bunkers, heat would build up rapidly in Iraqi bunkers made of thick reinforced concrete ceiling and walls, destroying most of the agent.

DOD's Postwar Field Testing Did Not Realistically Simulate Actual Conditions of Demolition Operations

During the 1960s Black Hills testing, rockets filled with sarin and VX nerve agents were intentionally ignited by thermite grenades—alone and with the addition of diesel fuel—as well as by fused initiation of the burster explosive charge. According to the testing report, on the first trial, the 11 crates of rockets were stacked 18 in. from the front wall on the right side of the igloo. Because of safety considerations, a centrally located burster was ignited with a 10-min. safety time fuse instead of an electric blasting cap. Once detonated, there were 104 major explosions during the 3 hrs. and 4 min. of observation. The maximum pressure recorded during the initial detonation was 19 lbs./sq. in. above atmospheric pressure. The maximum pressure during the trial was 621°C, which was recorded 15 min. after the ignition of the burster. The igloo sustained three wide cracks in the arch and two large holes on either side of the door. The only major debris found outside the igloo was the partial body of one M55 rocket, located approximately 200 ft. away. Because of the damage to the igloo, the test was deemed an official disaster.

During a second trial, crated rockets, with bursters removed, were stacked as in the first trial. Two thermite grenades were set on top of the crates at the motor end and were ignited with a 10-min. safety time fuse. The grenades functioned properly but failed to ignite the rocket motors or crate material. One day later, 12 thermite grenades were placed in three

[12] Central Intelligence Agency, *Intelligence Update*, p. 62.

groups of four, at the motor end of the crates, and 5 gal. of diesel fuel were poured over the 11 crates just before ignition with a safety fuse. From 2 crates, 12 rocket motors were ignited and hit the rear wall of the igloo, but the only damage found was a slight bulge in the bottom of the igloo door. The igloo withstood the fire and contained the rockets.

According to a later DOD study, based on the analysis in the 1960s of Dugway Proving Ground Trial C505, from a fire within an igloo containing GB-filled M55 rockets, the estimated amount of agent released was assumed to be 2.52 percent over a 60-min. interval.[13] This appears to be the basis for the conclusion that only approximately 2.5 percent of the agent would have been released.

However, the bunkers at Al Muthanna and other locations in Iraq were not deliberately set on fire with incendiary devices. They were targeted with high-explosive weapons such as Tomahawk missiles and laser-guided and nonguided bombs, which detonate and produce instantaneous and extreme blast forces, as well as shock and pressure waves and heat. A high explosive is one in which the speed of reaction (typically 5,000 to 8,000 m/s) is faster than the speed of sound in the explosive. High explosives produce a shock wave along with gas, and the characteristic duration of a high-explosive detonation is measured in microseconds (10^{-6} s). Further, if an explosion is confined within a chamber or room, the gas pressure increases rapidly to a sustained level and then decays by venting out. Under these conditions, shock reflections occur and the overall effect can be greater than that of the incident shock.[14] While the CIA analysts gave much credibility to blast heat, no consideration was given to either the shock blast effects of the munitions or the higher altitude plumes generated from the transfer of mass associated with the shock waves. Further, the CIA stated that its conclusion was supported by UNSCOM's physical inspections of Bunker 2. UNMOVIC, however, informed us that UNSCOM had not physically inspected this bunker for safety reasons relating to structural instability.

[13]U.S. Army Armament, Munitions, and Chemical Command, Chemical Research Development and Engineering Center, *Source Characteristics of a Fire within an Igloo Filled with M55 Chemical Munitions*, CRDEC-TR-87056 (Rock Island, Ill.: May 1987).

[14]National Academy of Sciences, Commission on Engineering and Technical Systems, *Protecting Buildings from Bomb Damage: Transfer of Blast-Effects Mitigation Technologies from Military to Civilian Applications* (Washington, D.C.: 1995), pp. 28–29.

At Muhammadiyat, DOD, using test data from Dugway Proving Ground, provided details about how source term characterizations for agent released were derived. However, the type and quantity of explosives used in the Dugway testing—and, therefore, the resulting effects—are not comparable with the type and quantity of munitions that were actually used at Muhammadiyat. At Dugway Proving Ground, small explosive charges were placed on boxed rockets; at Muhammadiyat, the munitions storage bunkers were targeted with multiple high-explosive bombs. Agent purity at Muhammadiyat was estimated at 15 percent. In addition, according to UNSCOM, there are many questions as to the accuracy of Iraqi records for the number of munitions filled with sarin at Muhammadiyat.

Major Unresolved Issues Concerning Modeling

The major unresolved issues concerning DOD's modeling include assumptions about (1) CIA's modeling of Khamisiyah Bunker 73; (2) repeated aerial bombings of storage facilities; (3) repeated aerial bombings of other storage, as well as research and production, facilities; (4) binary munitions and combustion by-products; and (5) detection of CW agents.

CIA's Modeling Assumptions about Khamisiyah Bunker 73

In July 1996, the CIA briefed the Presidential Advisory Committee on Gulf War Veterans' Illnesses on the results of its Bunker 73 model. In the August 2, 1996, report, identifying its modeling assumptions, the CIA concluded that any hazard area resulting from the demolition of Bunker 73 had moved east and northeast.[15] According to CIA reporting, Bunker 73 and a number of other bunkers were destroyed on March 4, 1991. The pit, warehouses, and most of the remaining bunkers were destroyed on March 10, 1991. The CIA revised some of the source term assumptions for Bunker 73 in its 2002 report, but Bunker 73 was never remodeled.[16] Among the CIA's more significant assumptions in the 1996 modeling of the demolition of Bunker 73 were these:

- Bunker 73 contained approximately 1,060 rockets filled with sarin (this figure was modified to 910 rockets in 2002).
- Each rocket was filled with 8 kg of a 2:1 ratio of sarin to cyclosarin.

[15]Central Intelligence Agency, *CIA Report on Intelligence.*

[16]Central Intelligence Agency, *Intelligence Update.*

- The demolition ejected an estimated 10 percent of the rockets from the bunker.

- U.S. tests showed that heat from the explosion, as well as burning motors and crates, degraded all but 2.5 percent of the agent in the bunker.

- Winds were light, northeast to east.

- The modeling did not include the effect of thermal energy released by the simultaneous burning and detonation of the other 32 to 37 bunkers at the site.

The models used to arrive at these conclusions, however, were not identified in the 1996 CIA report. According to the SAIC analyst who conducted this modeling, the models used, the NUSSE4 and DP2C, were dispersion models. (They were among the models that the IDA expert panel later evaluated as having been inadequate for assessing the pit.)

In addition, the potential for greater contamination than indicated by the initial models exists, given (1) UNMOVIC's assertion that UNSCOM did not physically inspect this bunker through 1996 for safety reasons related to contamination and structural instability and given DOD's conclusion that all but 2.5 percent of the agent was degraded and (2) the lack of correlation between the igloo testing studies and actual events.

Assumptions about Repeated Aerial Bombings of Storage Facilities

DOD reported that the Al Muthanna storage, as well as research and production, facilities for CW agents was repeatedly attacked. Despite its repeated bombing, however, on only one occasion did the CIA and DOD express any concern about agent release. According to DOD analysis of the destruction of Bunker 2 at Al Muthanna on February 8, 1991, no trace residue from CW agents was detected in or around the bunker during UNSCOM inspections. However, UNMOVIC told us that UNSCOM never inspected this bunker for safety reasons. A low-level vapor hazard, however, probably emanating from damaged and leaking 122 mm rockets stored in the open, was detected around some of the other facilities. UNSCOM also reported contamination in several other facilities at Al Muthanna that produced CW agents. These observations suggest that additional releases may have resulted from the repeated attacks at this site. However, DOD did not analyze or model additional direct or incidental destruction from the numerous bombings at the site.

There were 17 discrete Coalition aerial bombings of the Muhammadiyat munitions storage facility. For Muhammadiyat models, we identified two issues: time for agent dissipation and number of aerial bombings. For agent dissipation, DOD assumed that agent concentration would have been reduced to miniscule amounts during the first 24 hours. Therefore, for this model, DOD chose a duration of 24 hours.

For number of aerial bombings, DOD made the assumption that all sarin at Muhammadiyat was released at one time; therefore, DOD modeled each aerial bombing as if it were the only bombing that caused a release. According to DOD, each model produced a freeze frame of the largest hazard area. The models suggest that the hazard area grows until it reaches its maximum size, which is about 10 to 12 hours after the release. However, in these models, DOD focused on a single bombing, failing to consider the cumulative effects of exposure resulting from multiple aerial bombings.

Assumptions about Repeated Aerial Bombings of Other Storage, as Well as Research and Production, Facilities

CW agents could have been released from a number of suspected storage sites targeted for Coalition bombing, in addition to Al Muthanna and Muhammadiyat. As shown in table 2, the intelligence community identified numerous sites alleged to include research, production, or storage facilities for CW agents. Another element of uncertainty in the modeling site selection process results from the possibility that low-level chemical agent exposures might have originated from bombing these other sites during the Gulf War. Many of these sites were entirely discounted for the purposes of the CIA's and DOD's models because the intelligence community's assessment was that potential plumes from these sites could not have reached U.S. troops.

Table 2: Gulf War Suspected Chemical Weapon Sites

Site type	Site name and place	Site name and place
Facility	Al Muthanna (Samarra)	Fallujah III (Habbaniyah I)
	Fallujah I (Habbaniyah III)	Khamisiyah (Tall al Lahm)
	Fallujah II (Habbaniyah II)	Muhammadiyat (Qubaysah Storage Depot)
Airbase, airfield	Al Bakr Airfield (Samarra East Airfield)	Murasana Airbase (H3 NW Airfield)
	Al Taba'at Airstrip (H3 SW Airfield)	Qadisiyah Airbase (Al Asad Airfield)
	Al Taqaddum Airfield	Qayyarah West Airfield
	Al Tuz Airfield (Tuz Khurmatu Airfield)	Saddam Airbase (Qayyarah West Airfield)
	Al Walid Airbase (H3 Airfield)	Tallil Airfield
	K-2 Airfield	Tammuz Airbase (Al Taqaddum Airfield)
	Kirkuk Airfield	Ubaydah Bin al Jarrah Airfield
	Mosul Airfield	
Ammo	Ad Diwaniyah Ammo Depot	Kirkuk Ammo Depot West
	Al Fallujah Ammo Depot South	Qabatiyah Ammo Storage (Wadi al Jassiyah Ammo Storage)
	An Nasiriyah Ammo Storage Depot SW	Qayyarah West Ammo Storage Depot
	Ash Shuaybah Ammo Storage Depot	Tikrit Ammo Depot (Salahadin)
	Baghdad Ammo Depot Taji	Ukhaider (Karbala Depot and Ammo Storage)
Other	Al Qaim Superphosphate Fertilizer Plant	Fallujah Chem Proving Gnd (Habbaniyah CW Training Center)
	Dujayl/Awarah (Sumaykah SSM Support Facility SE)	

Source: Central Intelligence Agency, *CIA Report on Intelligence Related to Gulf War Illnesses* (McLean, Va.: Aug. 2, 1996).

Assumptions about Binary Munitions and Combustion By-Products

During our interview with UNSCOM and UNMOVIC officials, we were informed that Iraq had extensively deployed binary CW munitions components at a number of sites within range of the theater of operation. The chemical munitions were deployed unfilled but with chemical components that when mixed would make the nerve agent sarin. While there is no evidence that Iraq used these munitions offensively, it remains unknown whether these materials were mixed intentionally or inadvertently during combat engagements. In addition, Iraq declared that 823 metric tons of CW agent precursor materials, deployed throughout Iraq, were destroyed during the Gulf War. The hazards associated with environmental fallout from the destruction of these materials remain unknown. CIA assessments and DOD modeling specifically did not address potential exposure to binary CW agents, exposure to CW agent precursors, or the hazardous environmental consequences associated with the combustion by-products of sulfur mustard from the bombings at Muhammadiyat and other sites.[17]

Assumption about the Detection of CW Agents

Credible scientific evidence suggests that the reported detections of CW agents were reliable. These detections, during the early phases of the air war, included the use of an array of scientific methods. The CIA and DOD assert that the various detections are not valid because the source of the agents cannot be detected. Available evidence, however, can explain the detections and suggests that troop exposure may be far more widespread than that projected from the CIA and DOD modeling of the Khamisiyah release alone.

Coalition forces possessed a diverse array of CW detection and identification equipment. Czechoslovak chemical detection troops also used Warsaw Pact equipment, such as the GSP-1, GSP-11, and Czech mobile chemical agent laboratories. This broad array of equipment included various technologies to detect and confirm the presence of CW agents, as well as to identify specific agents. The different physical principles that were employed included wet chemistry, mass spectrometry, ion mobility spectrometry, chemical reaction, biochemical enzyme reactivity, flame photometry, and ionization. Detector (that is, sensor) and agent identification systems that Coalition forces deployed, reporting the detection of CW agents, are shown in table 3.

[17]Precursors are chemicals and other materials used in producing CW agents

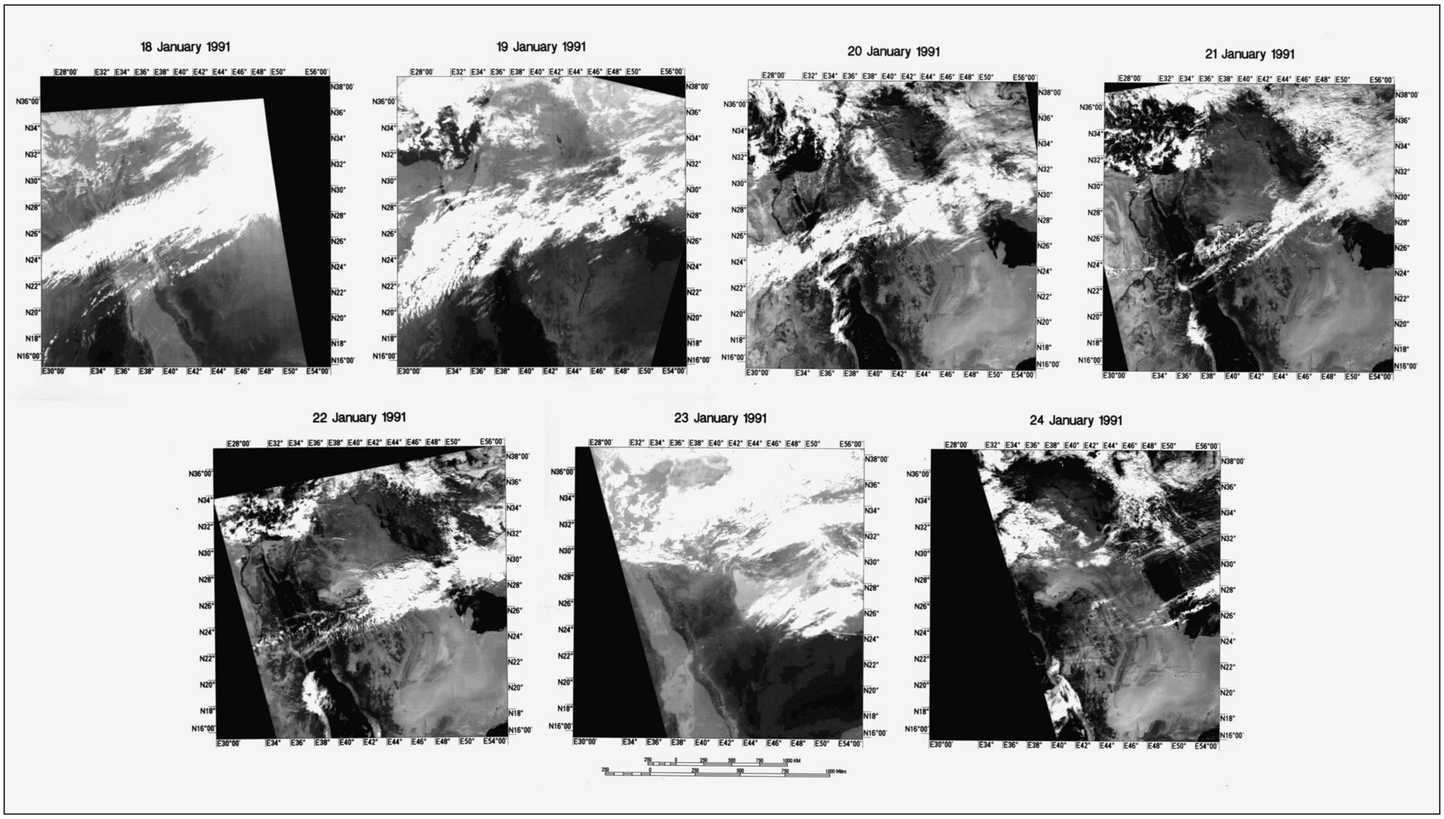
Table 3: Detector and Agent Identification Systems Deployed by Coalition Forces Reporting the Detection of CW Agents

Nation	System	Chemical agent	Sensitivity	Method or technology
Czech Republic	Automatic Nerve Agent Detector Alarm GSP-11	G- and V-series nerve agents	0.05 mg/m3	Air sampling, biochemical enzyme (cholinesterase reactivity)
	Mobile Laboratory CHP-71, PCHL-90, and PCHL-90	Most CW agents	Agent identification through wet chemistry analysis	Field portable chemical agent laboratory, chemical reagents, wet chemistry analysis
France	Detalac F1	G- and V-series nerve agents	Not available	Biochemical enzyme detector
	Chemical Detection Control Kit (TDCC)	Tabun/sarin (GA/GB) Hydrogen cyanide (AC) Cyanogen chloride (CK)	1 mg/m3 350 mg/m3 2000 mg/m3	Chemical biochemical detector
	Chemical Detection Unit for Fixed Installations (ADLIF)	Sarin/soman (GB/GD)	Not available	Flame spectrometry
United Kingdom	Chemical Agent Monitor	G- and V-series nerve agents Blister agents (H)	0.1 mg/m3 2.0 mg/m3	Ion mobility spectrometry (quantitative feature)
	Nerve Agent Immobilized Enzyme Alarm and Detector (NAIAD)	G-series nerve agents V-series nerve agents	0.05mg/m3 0.005 mg/m3	Biochemical enzyme detector (cholinesterase reactivity)
	MARK I	G-series nerve agents Blister agents (H)	Not available	Biochemical chemical reactivity
United States	M8A1 Alarm	G-series nerve agents V-series nerve agents	0.1 mg/m3 0.2 mg/m3	Ionization Automatic alarm
	M8 Paper; M9 Paper	G- and V-series nerve agents Blister agents (H)	Yes/no Yes/no	Chemical reactivity Color interpretation
	Chemical Agent Detector Kit M256 and M256A1	G-series nerve agents V-series nerve agents Blister agents (H)	0.05 mg/m3 0.15 mg/m3 3.0 mg/m3	Biochemical enzyme detector (cholin-esterase reactivity) Chemical reactivity
	Chemical Agent Monitor	G- and V-series nerve agents Blister agents (H)	0.1 mg/m3 2.0 mg/m3	Ion mobility spectrometry (quantitative feature)
	MM1 FOX NBC vehicle	G-series nerve agents 60 other preprogrammed agent spectra	Several mg/m3 M8A1 (M43) ionization backup unit (early warning)	Quadrapole gas chromatography C-mass spectrometryS Full GC-MS

Sources: DOD FM Series 3; Chemical Research, Engineering, and Development Command, Aberdeen Proving Grounds, Md.; manufacturers; Jane's NBC Protection Equipment (1991–92 and 1995–96); CIA, Chemical Warfare Agent Issues during the 1991 Persian Gulf War (Apr. 2002); video footage of Czech chemical detection unit activity during the 1991 Gulf War; Czech Chemical Warfare Report, Intelligence Assessment of Chemical and Biological Warfare in the Gulf, prepared by U.S. Army Foreign Science and Technology Center and DIA for the Defense Science Board investigating the Desert Storm Syndrome.

Shortly after the air war began, the period between January 18 and January 23, 1991, was marked by a stationary frontal pattern and the development of low-level cloud and fog activity directly over the areas occupied by Coalition forces. Figure 3, based on NOAA-11 AVHRR 1B meteorological imagery, shows contemporaneous visual evidence of this anomaly.

Figure 3: NOAA-11 Meteorological Imagery of Areas Occupied by Coalition Forces, January 18–24, 1991



Source: NOAA.

Note: NOAA-11 AVHRR 1B meteorological imagery, processed by Earth Satellite Corporation, Rockville, Md., was obtained from the National Center for Atmospheric Research, Ashville, N.C.

During this prolonged period, a significant number of reported CW agent detections were reported in northern Saudi Arabia where Coalition forces were deployed (see table 4).

Table 4: Principal Reported Detections of Chemical Agents in Saudi Arabia, January 17–23, 1991

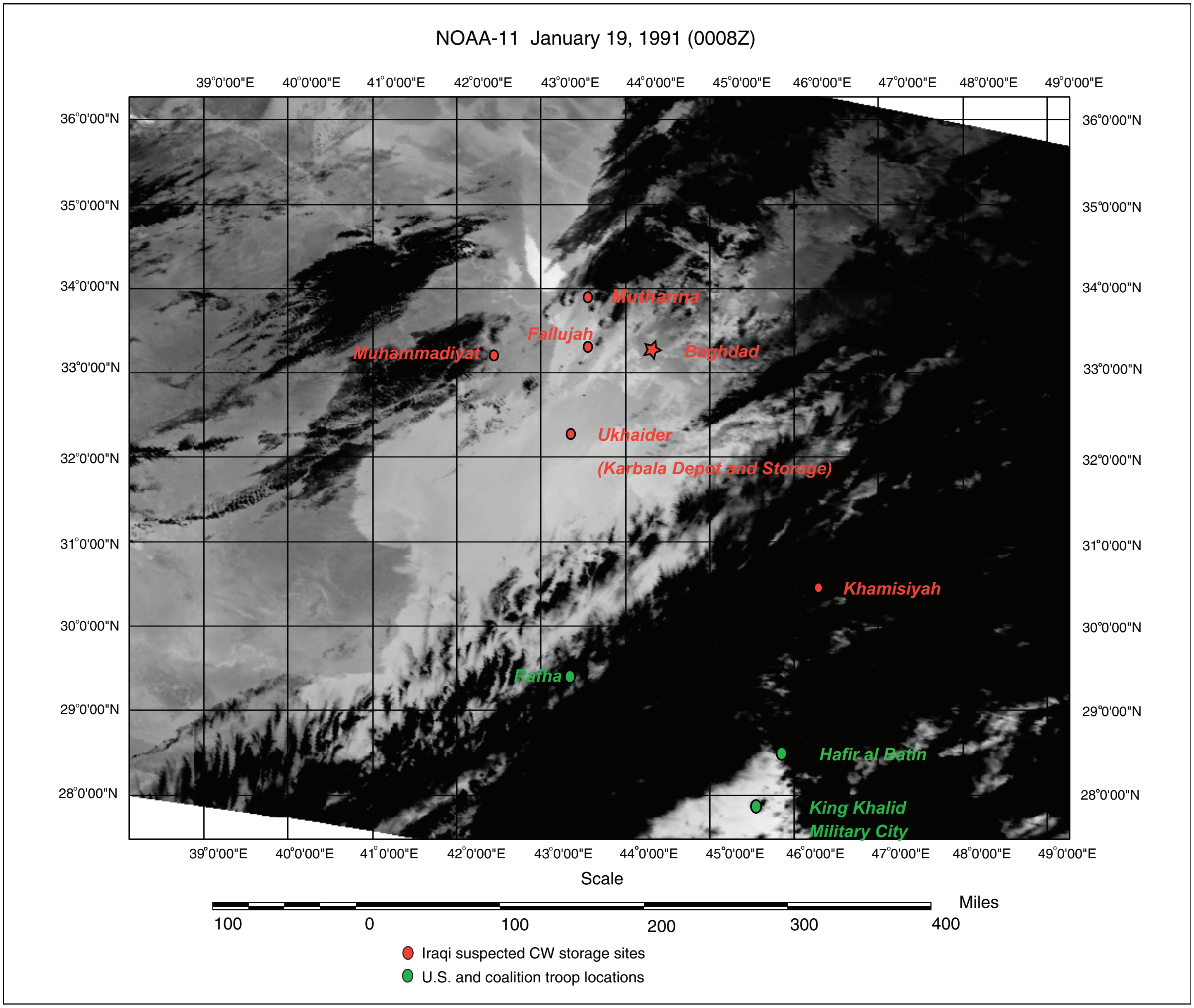
Date	Nation and unit	Place	Agent detected	Method or technology
Jan. 17	United States 2/5TH SFG	NW Hafir Al Batin	Unknown nerve agent	Ionization, biochemical reaction; ion mobility spectrometry; M8A1, M256, CAM
Jan. 19	Czech Republic Chemical Detection Unit	N Hafir Al Batin	Sarin (GB)	Biochemical reactivity; wet chemistry; GSP-1(11), mobile lab
		NE Hafir Al Batin	Sarin (GB)	Biochemical reactivity; wet chemistry; GSP-1(11), mobile lab
		King Khalid Military City (KKMC)	Unknown nerve agent	Biochemical reactivity; wet chemistry; GSP-1(11), mobile lab
	France	30 km from KKMC	Unknown nerve agent	Biochemical reactivity
	Czech Republic Chemical Detection Unit	30 km from KKMC	Confirm French detection	Wet chemistry mobile lab
		KKMC	Sulfur mustard (HD)	Wet chemistry mobile lab
	United Kingdom	Jubayl	Unknown blister (after unexplained explosions)	Chemical reactivity; ion mobility spectrometry; M-9, CAM
	United States 24th Naval Reserve Construction Battalion (Seabees)	Jubayl	Unknown blister (after unexplained explosions)	Chemical reactivity; M-256 (2/3 tests)
Jan. 20	Czech Republic Chemical Detection Unit	Near KKMC	Sulfur mustard (HD) for 2 hours	Wet chemistry mobile lab
		French sector KKMC	Sarin (GB)/tabun (GA)	Biochemical reactivity; wet chemistry mobile lab
	France	Near KKMC	Unknown nerve agent	Biochemical reactivity
	United States 800th MP BDE	NW of KKMC	Unknown nerve agent	Ionization, biochemical reactivity; M8A1, M256
	United Kingdom	Dharan	Unknown nerve agent (after SCUD attack)	Biochemical reactivity (separate devices); NAIAD, MARKI
Jan. 21	Czech Republic Chemical Detection Unit	KKMC	Sarin (GB)/tabun (GA), sulfur mustard (HD)	Biochemical reactivity; wet chemistry mobile lab
	France	KKMC	Unknown nerve agent	Biochemical reactivity
		KKMC	Unknown CW agent	Chemical or biochemical reactivity
Jan. 22	United States	RAFHA	Unknown nerve agent	Ionization, biochemical reactivity; M8A1, M256
Jan. 23	Czech Republic Chemical Detection Unit	KKMC	Unknown CW agent	Wet chemistry mobile lab
		Near KKMC	Patch of sulfur mustard (HD)	Wet chemistry mobile lab

Sources: CENTCOM CCJ3-X log (partially declassified 1995), Defense Science Advisory Board report (June 1994), Czech government reports on detection activity during the Gulf War, and declassified Defense Intelligence Agency reports on chemical detection activity.

This period of reporting on chemical agent detection is attributable to the use of many of the instruments and methodologies cited in tables 3 and 4 and coincides directly with the initial Coalition bombings of confirmed and suspected Iraqi CW research, production, and storage facilities during January 17–24, 1991.

That these detections were reported raises continued and unresolved concern that U.S. troop exposures may have been more frequent and more widespread than the CIA and DOD believe. In addition, these exposures may have involved different populations than can be assumed from the limited simulations performed with the CIA and DOD models we refer to in this report. For example, figure 4 shows NOAA-11 collected infrared satellite imagery data for January 19, 1991, that NOAA published, indicating a line of instability and an inversion coincident with, and possibly precipitating, the detection of CW agents in the vicinity of Hafir al Batin and King Khalid Military City, reported by Czech chemical detection units and other Coalition units in the region (as shown in table 4).

Figure 4: Meteorological Satellite Photography, Iraq, January 19, 1991



Source: GAO analysis of NOAA data.

The lighter areas in the image represent warmer air, the darker areas colder air. In the image, a collision is shown between cold and warm air masses, as well as what appears to be evidence of inversion activity. That is, colder upper air moves below the warmer air, which is normally closer to the surface, possibly precipitating the detections.

The assessment by the CIA that the detections of the Czech chemical detection units are not credible—despite earlier DOD technical assessments that Czech detections of sarin and mustard agent on two occasions were credible—are unsound because these assessments do not refute the underlying scientific evidence, according to DOD and DIA experts.[18]

The presence of both sarin and mustard agents at subacute levels, and in such close proximity, could reasonably be explained as the result of fallout from Coalition bombings of Iraqi weapons facilities and storage bunkers. This possibility is supported not only by the atmospheric data but also by observing that the Czech nerve agent detections were made, according to DIA reporting, by multiple teams over a range of 20 to 50 km, within 30 minutes of one another, during a strong weather inversion.[19] According to U.S. reporting, during the period immediately before the detections, the following known Iraqi CW agent research, production, and storage sites (with geocoordinates and dates) were bombed:

- Al Muthanna (3350 N, 04348 E, January 17, 1991),
- Al Nasiriyah (3058 N, 04611 E, January 17, 1991), and
- Qabatiyah (3353 N, 04239 E, January 19, 1991).

In addition, production sites for CW agent precursors were bombed during this period.[20] The potential remains that sites that may have been incorrectly assessed as not containing CW agents were also bombed

[18]See Czech chemical warfare report, *Intelligence Assessment of Chemical and Biological Warfare in the Gulf*, prepared by the U.S. Army Foreign Science and Technology Center and DIA for the Defense Science Board investigating Desert Storm Syndrome.

[19]Defense Intelligence Agency, *Detection of Chemical Warfare Agents by Czechoslovak Unit during Desert Storm*, Part III (U), IIR 6 284 0008 94 (Washington, D.C.: Oct. 14, 1993).

[20]U.S. Army Operations Group, UNSCOM, *Inspection of Chemical Warfare Facilities*, IIR 2 201 0022 92 (Washington, D.C.: Oct. 3, 1991, declassified 1995).

during this period. These activities may have resulted in additional releases of CW agents.

DOD's Modeling Efforts Were Flawed

In this section, we discuss our findings about DOD's modeling efforts in more detail, explaining the ways in which we believe them to be lacking in credibility.

DOD's Models Significantly Underestimated Plume Height

Actual plume height might have been significantly higher than the height DOD estimated from its models of demolition operations and bombings, given the assumptions DOD's modeling used for Khamisiyah and the other Iraqi storage facilities. The plume height estimates that the CIA provided for demolition operations at the Khamisiyah pit were 0 to 100 meters. However, neither the CIA nor DOD conducted testing to support estimated plume height associated with the bombings of Al Muthanna, Muhammadiyat, or Ukhaydir. According to DOD modelers, neither plume height nor any other heat or blast effects associated with these bombings were calculated; instead, these data were taken from DOD's Office of the Special Assistant for Gulf War Illnesses.

In addition, according to a principal DTRA modeler, DOD's data on plume height were inconsistent with other test data for the types of facilities bombed. This expert cited test studies conducted at White Sands Proving Ground demonstrating that plume height would range from 300 to 400 m. The CIA maintains, however, that the plume occurred near ground level because there was no altitude, burst of munitions, or any kinetic force such as bomb blast to force the agent to become airborne initially. DOD maintains that during the bombing of chemical agents, liquid agent absorbs the energy of the blast, greatly reducing plume heights. We asked the CIA and DOD to provide test data in support of their assertions but neither agency provided any evidence.

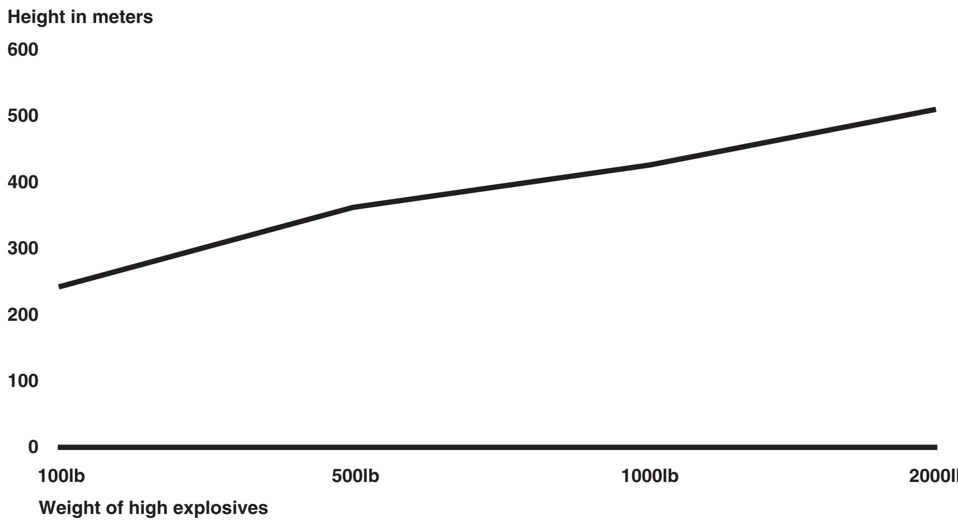
Modeling experts from LLNL who participated in only the initial modeling of the Khamisiyah pit site also said that they questioned the plume height estimates. In a prewar analysis, LLNL projected that the plume immediately following the bombing of Iraqi storage facilities for CW agents would be a surface-based plume with a horizontal radius of 54 m (177 ft.) and a height of 493 m (1,617 ft.). A 1969 Sandia National Laboratories empirical study established a power-law formula for calculating plume heights attributable to high-explosive detonations (see appendix II). By this formula, a conventional MK-84 or GBU-24 bomb (containing 942.6 lb.

of high explosives) of the type used to bomb sites other than those at Khamisiyah would generate a plume of approximately 421 m.

Impact by Weight of Explosive

Figure 5 shows the trend line for a plume height predicted on the basis of the formula for calculating plume height resulting from detonating high explosives ranging from 100 to 2,000 lbs. in weight.

Figure 5: Plume Height Trend Line by Weight of Explosive



Source: GAO.

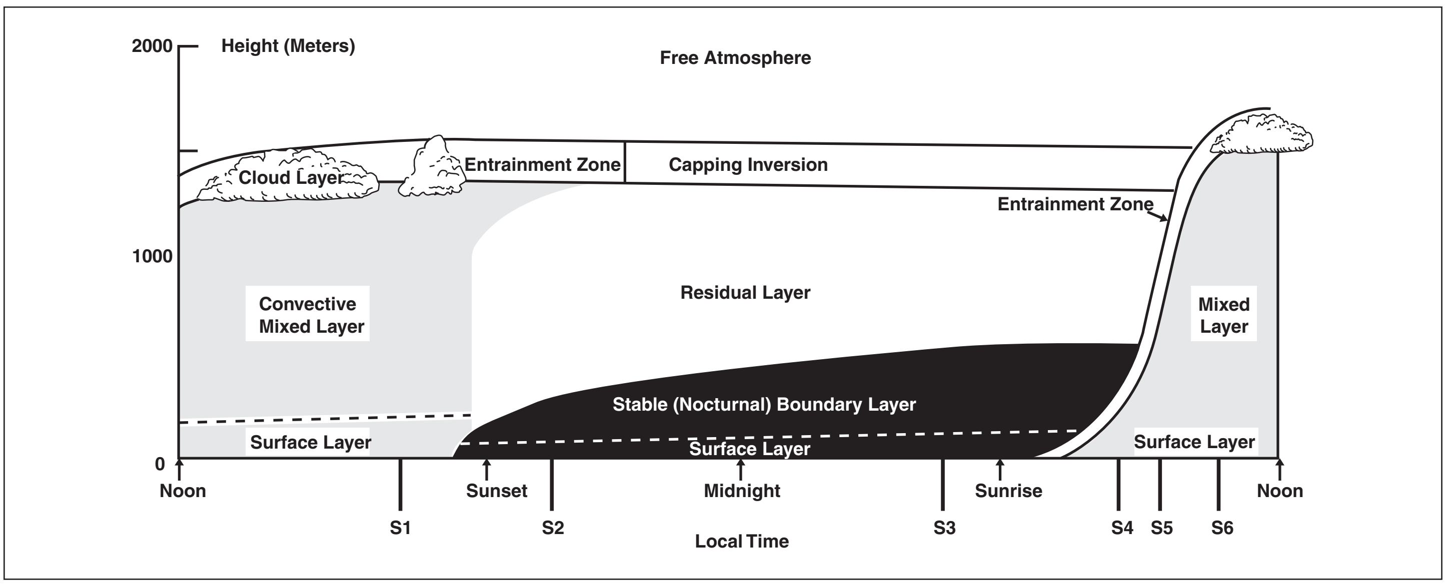
According to DOD officials, the Sandia National Laboratories study is not accurate or scientifically valid because it did not account for weather effects. Further, they said, it based plume height on the detonation of conventional explosives, but the liquid agent of chemical-filled munitions would have absorbed much of the energy of the blast if these had been bombed; therefore, plume height would have been greatly reduced. However, DOD could not provide us with any data, studies, or testing of the explosive aspects and buoyancy of chemical agents to corroborate these observations.

At Muhammadiyat, DOD established a plume height of 0.5 m (roughly half the bomb height) for nerve agent and a plume height of 1.0 m (roughly half the median height of the various bomb stacks) for mustard agent destroyed at this location. Moreover, according to one internal DOD memorandum, an “initial cloud size” of 10 m, in both lateral and vertical directions, was “arbitrarily” established. According to DOD, no effort was made to validate these estimates by analyzing video images that showed some of the plume data, such as (1) those taken from ground level at Khamisiyah and (2) the available footage from the aircraft and munitions used to bomb the other sites.

Impact by Wind Speed

Figure 6 shows that disparity in source term data for plume height could lead to vastly divergent results as to how far the plume travels and disperses. This observation is particularly relevant during nighttime periods, when a stable nocturnal boundary layer emerges.

Figure 6: Boundary Layer Characteristics



Source: Roland B. Stull, *An Introduction to Boundary Layer Meteorology* (Boston, MA: Kluwer Academic Publishers, 1988), p. 11. Used with kind permission of Kluwer Academic Publishers.

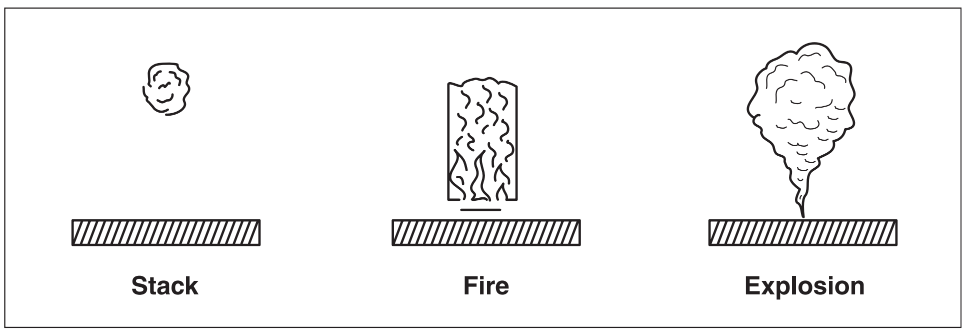
Figure 6 shows the stable nocturnal boundary layer where winds often accelerate to higher speeds, in a phenomenon referred to as the low-level or nocturnal jet—that is, winds are aloft in the nighttime hours. At altitudes on the order of 200 m above ground, winds may reach 10 m to 30 m per second (22 to 67.5 mph) in the nocturnal jet. Higher plumes than those DOD postulated, coupled with this phenomenon, could result in chemical agents being transported rapidly until disturbed by turbulence or the return of the mixed layer, some time after dawn. However, this possibility was not taken into consideration in any of the modeling DOD performed. Consequently, the models may have dramatically underestimated the extent of plume coverage.

According to DOD, such winds are not known to be present in Iraq. DOD confirmed, however, that it has no available data on prevailing wind conditions in the region, over the varying terrain, or during the time period in question to rule them out. DOD also stated that data are not available for determining the presence of a low-level or nocturnal jet at the time of the Khamisiyah demolition. However, DOD acknowledged uncertainty as to whether a low-level jet was present on any specific date or at any specific location.

Impact by Plume Geometry

According to plume geometry, the majority of plume mass associated with high-explosive discharges is located toward higher altitudes (see figure 7). This suggests that the majority of the plume mass would move to higher altitudes to be transported by higher wind speeds.

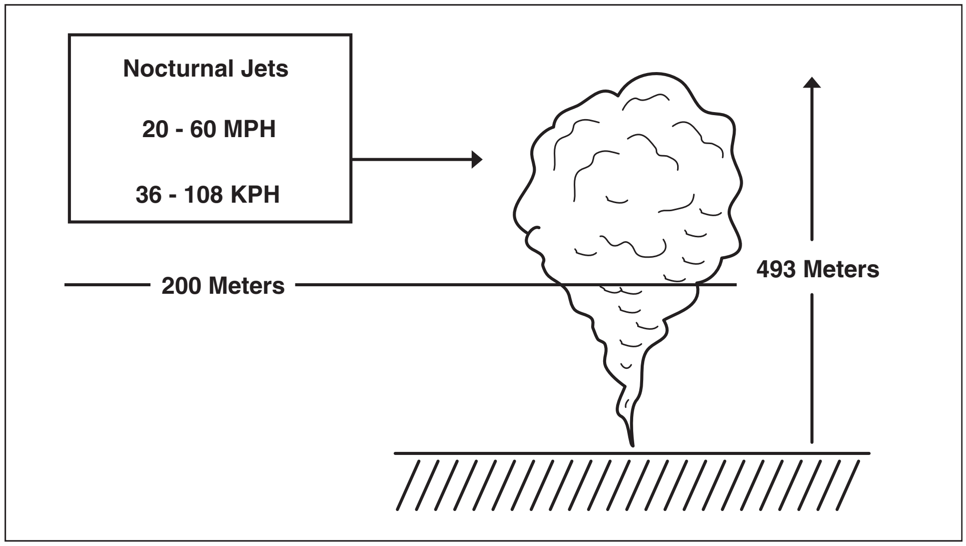
Figure 7: Three Types of Plume Geometry



Source: Department of Energy, Lawrence Livermore National Laboratory.

Similarly, the distribution of plume geometry may be affected by nocturnal jets, as shown in figure 8.

Figure 8: The Impact of Nocturnal Jets on a Plume at Higher Altitudes



Source: Department of Energy, Lawrence Livermore National Laboratory.

Empirical studies and observed events tend to refute the assumptions with which the CIA and DOD discounted the alternative assumption that the plume was transported by low-level jets. First, empirical testing suggests that the plume heights were much higher than postulated in the source term data. Second, no massive casualties or effects were claimed, reported, or observed in areas immediately surrounding the Iraqi CW research, production, and storage facilities bombed by Coalition forces. In the absence of an alternative explanation, acute effects should have been observed in areas near the bombed sites.

Third, since many of the bombings were at night, the explosive effects—coupled with higher-altitude plumes and a nocturnal boundary layer capable of moving hazardous materials hundreds of miles—could easily account for the phenomena reported above. These effects could also account for the numerous reports of CW agents detected in sites occupied by U.S. and Coalition forces. Fourth, these effects may account for reported nighttime detections of low levels of CW agents, associated with turbulence—resulting from aircraft-related sonic booms and incoming missiles and artillery—mixing the upper-level and lower-level atmospheric layers.

DOD's Field Testing Did Not Realistically Simulate Actual Conditions

To simulate the effects of demolition on chemical nerve agent stockpiles, the CIA and DOD conducted postwar field testing at Dugway Proving Ground. They explored what percentages of agent would be deposited on the ground as a liquid, consumed by the demolition, and aerosolized. To obtain these source term data for their models of Khamisiyah, the Dugway Proving Ground testing center conducted seven field tests and two laboratory studies from May 1997 through November 1999.

Demolition and Bombing Simulations Were Not Realistic

For field testing to be effective, conditions have to be as close to the actual event as possible, but these tests did not provide more definitive data for the CIA's and DOD's models. The tests did not realistically simulate the conditions of the demolition of 122 mm chemical-filled rockets in Khamisiyah. The simulations took place under conditions that were not comparable with those at Khamisiyah. There were differences in meteorological and soil conditions; the construction material of munitions crates; rocket construction (including the use of concrete-filled pipes as rocket replacements to provide inert filler to simulate larger stacks); and the numbers of rockets, using far fewer rockets and, therefore, explosive materials. Additionally, the tests used an agent stimulant whose physical properties differed from those of the actual agent.

For example, of the 32 test rockets with simulant-filled warheads, a small sample was used to conduct all seven field tests: five tests were single-rocket demolitions and two involved multiple-rocket demolitions. One multiple-rocket field test demolition used 9 functional rockets plus 3 dummy rockets; the other multiple-rocket field test used 19 functional rockets and 5 dummy rockets. In contrast, at the Khamisiyah pit, stacks of 122 mm rockets were detonated, estimated at about 1,250 rockets. Dugway officials acknowledged that detonating a larger number of rockets would have made a significant difference to the testing. In addition, aerial bombing with a heavy explosive load, such as had occurred at the sites other than Khamisiyah, would have had a far greater effect than was achieved with the Dugway testing.

Extrapolation Data Contained Acknowledged Limitations

According to DOD officials, SAIC developed projection data, in support of the Dugway tests, to extrapolate the results to the larger stockpiles identified at Khamisiyah. However, in our review of these data, we found that the SAIC analysts acknowledged limitations to their study. They noted in their report:

> This is the first attempt to develop an understanding of these processes and we lack sufficient data to either validate or completely calibrate our models. The models in many cases have diverged from first principles models and have become "engineering" models in the sense that explicit calculations have been based on the observations of others and these have not been well documented as yet. In some cases, it may be that further calculations and/or tests will determine that for different configurations than those calculated/tested, some of the understandings developed from examining the existing data may have to be modified.[21]

Finally, SAIC acknowledged that calculations were performed in two dimensions, stating that full, three-dimensional calculations were not feasible.[22]

SAIC's conclusion was that the test explosions did not produce gases likely to endure long enough to cook off—that is, ignite—a motor, even in a large stack. But SAIC pointed out that "the real world often holds surprises in chaotic situations of this sort." For example, if a rocket motor were to ignite and burn in place, its energy might sensitize adjacent motors or cause cook-off initiation: "these processes could walk their way through the aisles [stacks] with minutes between dramatic events."[23] The potential for real world events like these points to the inadequacy of attempting to extrapolate from the small-scale controlled testing at Dugway Proving Ground to the large-scale and relatively uncontrolled conditions at Khamisiyah.

Testing Did Not Use Iraqi Soil and Wood

According to DOD and CIA analysts, the type of soil and wood can have a significant effect on the dispersion of the agent. Their estimates of the evaporation and retention rates of the chemical agent spilled on the soil may not be similar to what actually evaporated from and was retained in the pit sand at Khamisiyah. Although Iraqi soil was available and used in the laboratory testing, it was not used in field testing. Similarly, assessments of DOD and CIA estimates of the amount of spilled agent evaporated from and retained in wooden crates are uncertain. This is because Dugway testing officials could not obtain actual wood from the Khamisiyah pit site. The aged and possibly damp wood at Khamisiyah

[21] G. I. Kent and others, *Interim Report: Calculations and Measurements in Support of the Rocket Testing at Dugway Proving Ground* (Dugway, Utah: June 17, 1997), p. 2.

[22] Kent and others, *Interim Report: Calculations and Measurements*, p. 2.

[23] G. I. Kent and others, *Interim Report: Rocket Motor Ignition/Explosives Issue* (Dugway, Utah: June 17, 1997), p. 15.

might have absorbed less agent than the new wood used at Dugway. DOD and CIA determined that only about 32 percent of the agent was released and that most leaked into the soil and wood, with 18 percent of the leakage becoming part of the plume (2 percent through aerosolization and 16 percent through evaporation).

Field tests were also conducted at a time of year and time of day different from the actual Khamisiyah pit event. According to Dugway officials, testing was done in May and in the early morning hours, when drainage conditions prevail.[24] When the March 10, 1991, U.S. demolition operations took place at the Khamisiyah pit, it was late afternoon with a mixed layer. The bombings of the other modeled storage and production sites often took place during evening and nighttime hours, when a stable nocturnal boundary layer emerges.

Despite the uncertainties in approximating the conditions at Khamisiyah, DOD and the CIA used these data not only for modeling Khamisiyah but also for modeling other sites. At the other sites, the CW munitions would have been destroyed by aerial bombings with much greater quantities of high-explosive charges and under significantly different meteorological conditions.

All Models, Including DOD's Composite Model, Showed Divergent and Unresolved Results

The models DOD used to predict the fallout from Khamisiyah and the other sites showed great divergence, even with the same source term data. While the models' divergences included plume size and path, DOD made no effort to reconcile them (see appendix III). The IDA expert panel determined that the results were so divergent that it would not be possible to choose the most exposed areas or which U.S. troops might potentially have been exposed. IDA therefore recommended a composite model, which DOD adopted.[25] However, this approach masked differences in individual model projections with respect to plume size and path.

[24] Drainage winds, also called mountain or gravity winds, are caused by the cooling air along the slopes of a mountain. Drainage wind periods were intentionally chosen for Dugway testing to minimize the dispersion of the test materials to the surrounding areas. According to Dugway officials, these winds are common to Dugway Proving Ground in the morning before the development of the mixing layer.

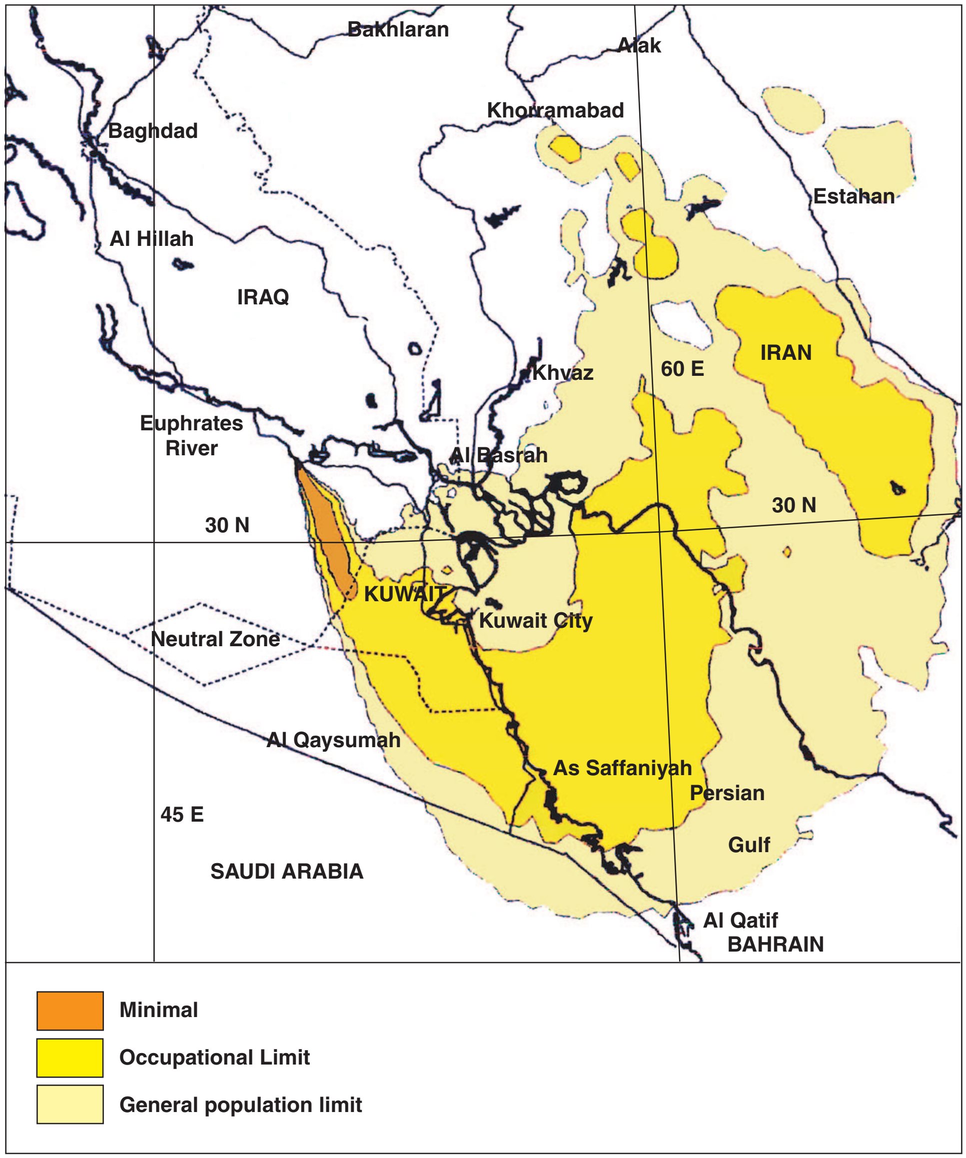
[25] The composite approach DOD used is also known as the ensemble approach.

In addition, DOD chose not to include in the composite model the results of the LLNL model, created at the IDA expert panel's request, which showed a different and larger plume size and path than DOD's models showed. The IDA panel regarded the LLNL model as less capable than other models because it modeled atmospheric phenomena with less fidelity. Finally, modeling of Khamisiyah that the Air Force Technical Applications Center (AFTAC) had done also showed differences from DOD's composite model.

LLNL's Model

To determine plume size and path, LLNL conducted analyses using DOE's MATHEW meteorological model with the ADPIC dispersion model. During LLNL's presentations to the IDA panel in November 1996 and February 1997, LLNL provided a 72-hour projection, assuming an instantaneous release of the contents of 550 rockets containing sarin (see figure 9). The results of LLNL's analyses show the plume covering an area extending south southeast from the release point to the Persian Gulf, then turning eastward at the Gulf coast, and then turning northeast over the Gulf and extending northeastward across central Iran.

Figure 9: Lawrence Livermore National Laboratory Projection



Source: Department of Energy, Lawrence Livermore National Laboratory.

LLNL's modeling assessment showed that the 72-hour exposure projection—for the instantaneous release of sarin from 550 rockets—covered a large hazard area. According to LLNL, 2,255 sq km were covered by agent concentration in excess of the dosage expected to cause a person minimal effects.[26] This area extended approximately 130 km south southeast from the release point. Dosages in excess of the amount that would be allowed for a worker exposed to sarin in the workplace, the occupational limit, were predicted over 114,468 sq km—including Kuwait City, an area approximately 200 km wide across the Persian Gulf, and the higher elevations of the Zagros Mountain range of western Iran.[27] The remaining area was determined to be at the general population limit.[28]

DOD's Composite Model

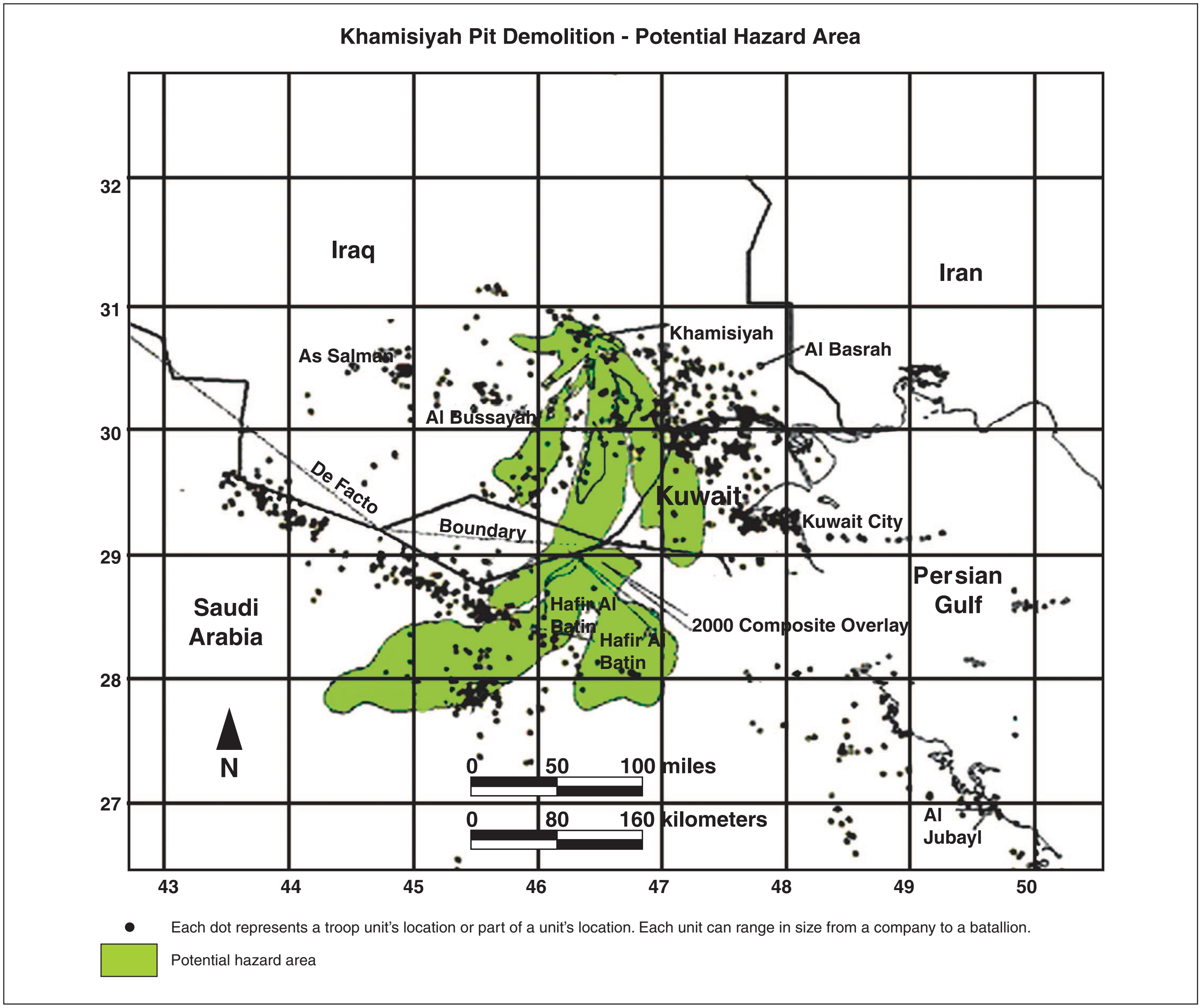
DOD's composite model was based on OMEGA and COAMPS meteorological models and HPAC/SCIPUFF and VLSTRACK dispersion models. In contrast to LLNL's modeling simulations, it showed the plume moving first southerly and then turning to the west southwest. The 72-hour plume overlay for DOD's composite model, resulting from using the VLSTRACK and HPAC/SCIPUFF dispersion models with COAMPS and OMEGA meteorological models, is shown in figure 10.

[26]"Minimal effect" is the lowest concentration expected to have noticeable effects on human beings.

[27]"Occupational limit" is about one-tenth of the minimal effects value and the maximum concentration level that would be allowed for a worker who could become exposed to sarin in the course of performing job duties.

[28]"General population limit" represents the limit below which any member of the general population could be exposed—by, for example, exhaling—7 days a week, every week, for a lifetime without experiencing any adverse health effects.

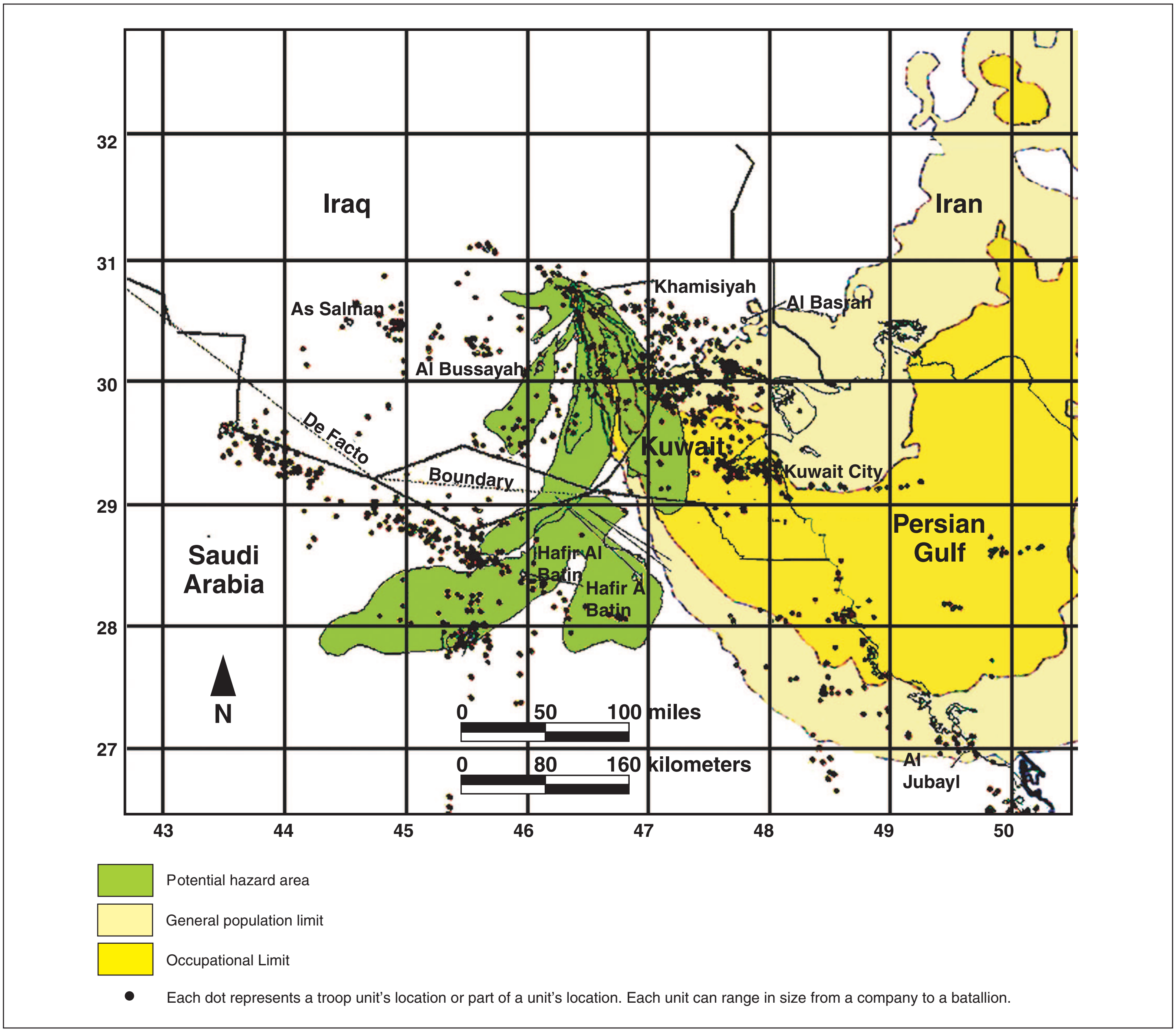
Figure 10: DOD Composite Projection



Source: Department of Defense, Office of the Special Assistant for Gulf War Issues.

Composite modeling may be an appropriate methodology, but DOD's composite model understated the number of troops potentially exposed by not including the LLNL model. If LLNL's model were included, a far larger number of forces would potentially be shown as having been exposed (see figure 11). DOD's exclusion of this model seriously skewed the outcome of any epidemiological studies done thus far.

Figure 11: DOD Composite Projection and Lawrence Livermore National Laboratory Projection



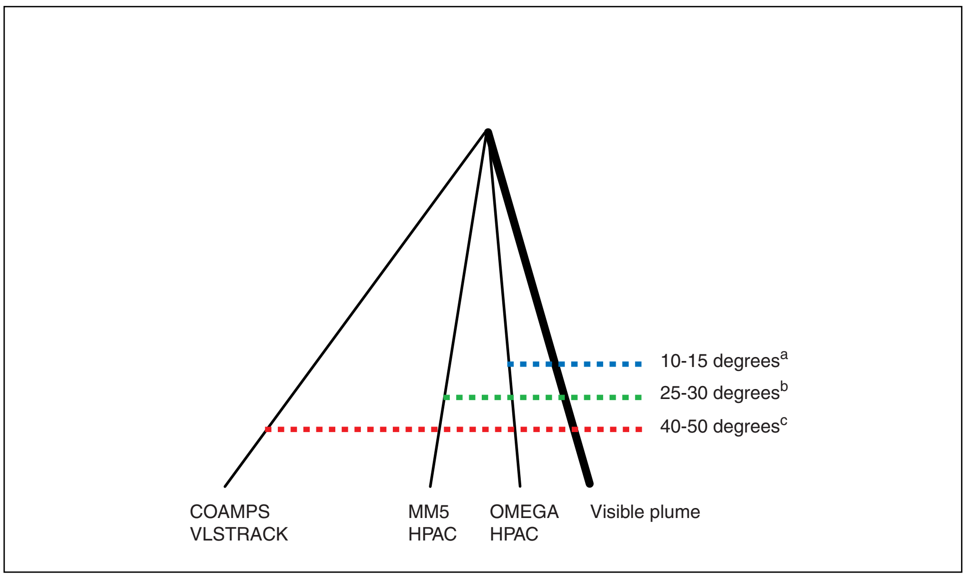
Source: GAO analysis of Department of Defense and Department of Energy, Lawrence Livermore National Laboratory, models.

Explaining the Models' Divergent Results

Further research was conducted to determine whether the models' divergent results could be explained. It was observed that the divergence in the modeling outcomes could be explained by a directional split—a line of diffluence—in the independently modeled 10 m wind field data near Khamisiyah in the first 2 days of the modeling period. While the precise location of this line is critical in determining which way the wind would have transported the CW agent, its precise location cannot be resolved with any degree of certainty. (Appendix IV illustrates this diffluence with three different data sets.)

In addition, DTRA officials said at the time of the modeling that they had conducted data-validation analyses of the various models against visible smoke plumes from the oil well fires in Kuwait. These analyses indicated a definite directional bias, shown in figure 12. This validation demonstrates that the actual area covered could have been from 10 to 50 degrees to the east and could have affected a different population from that indicated by the model results.

Figure 12: Validation Runs of Various Models



Source: GAO.

[a]East bias compared with OMEGA/HPAC.

[b]East bias compared with MM5/HPAC.

[c]East bias compared with COAMPS/VLSTRACK.

In addition, in the 1997 DOD peer review panel report on the Khamisiyah models, a panel of experts in meteorological and turbulent diffusion modeling stated that the VLSTRACK and SCIPUFF/HPAC results were complicated by the use of significantly different source term inputs.[29] According to DOD, the 2000 modeling used a consistent source term for SCIPUFF and VLSTRACK. For VLSTRACK, internal source algorithms had been used, while for SCIPUFF, source term inputs from the Dugway experiments had been used. These differences would lead to significant divergences in the dosage contours the two models predicted, which were then used to generate the composite.

In addition, the panel noted, while using a composite model is a valid method, successfully applied to other atmospheric modeling problems, using a composite model to reconstruct the dosage requires more advanced, state-of-the-art, high-resolution models, with the fewest physical limitations and assumptions. Furthermore, in 1998, the Air Force Human Effectiveness Directorate at Wright-Patterson Air Force Base reviewed the Muhammadiyat modeling and found that the work may have been flawed. According to the Chief of the Human Effectiveness Directorate, the protocol was not correct in that it constituted a "plume of plumes, rather than a plume based upon data."

AFTAC Modeling

AFTAC modeling of Khamisiyah also showed differences from DOD's composite model. After DOD's Khamisiyah models were published in 1997, the Senate Committee on Veterans' Affairs asked AFTAC to conduct an analysis of the event.[30] AFTAC is a principal modeling agency for DOD and, according to experts, the quality of its modeling system is among the highest. To conduct the analysis, AFTAC meteorologists used four models

[29] Richard A. Anthes and others, "Comments by Review Panel on Khamisiyah Modeling Report and Presentations on November 4–5, 1997," report for the Directorate for Deployment Health Support of the Special Assistant to the Under Secretary of Defense (Personnel and Readiness) for Gulf War Illnesses, Medical Readiness, and Military Deployments, Fairfax, Virginia, December 11, 1997.

[30] In early 1997, the Senate Committee on Veterans' Affairs established a Special Investigation Unit on Gulf War Illnesses to conduct a bipartisan review of the U.S. government's response to the unexplained illnesses suffered by veterans of the Gulf War. The year-long effort produced a detailed report on the actions by DOD before and during the war and by VA in its aftermath, relating to the current health of Gulf War veterans. See Arlen Specter, William F. Tuerk, and James R. Gottlieb, *Report of the Special Investigation Unit on Gulf War Illnesses*, U.S. Senate, Committee on Veterans' Affairs (Washington, D.C.: U.S. Government Printing Office, 1998).

from their suite of atmospheric models, including two primary transport and diffusion models and the RAMS meteorological model.

While AFTAC's analysis was meteorologically consistent with the modeling efforts of the CIA and DOD, the results showed that some additional areas might have been exposed to at least low-level exposure dosages from the Khamisiyah plume. In particular, it showed the plume differing in several aspects from the plumes generated by the DOD and CIA analyses. Most significantly, the AFTAC plume is shown drifting across Kuwait and the northwestern Persian Gulf coast, an area not covered by the DOD and CIA plume. AFTAC's analysis was published in a report to the Senate Committee on Veterans' Affairs and was provided to DOD's Office of the Special Assistant to the Deputy Secretary of Defense for Gulf War Illnesses.[31] The report recommended that AFTAC's results be integrated into the DOD and CIA modeling results to depict these additional areas of potential exposure.

However, in 1998, a DOD expert panel reviewed the AFTAC modeling simulations and recommended that AFTAC's modeling results not be included in DOD's composite plume model, given that (1) AFTAC's simulations were generally consistent with DOD's composite model results and that the effect of any differences would only have resulted in the additional notification of a small number of individuals, (2) continuing to refine DOD's modeling process rather than including AFTAC's modeling results would be the best use of DOD's resources, and (3) the decay capability of the agent in AFTAC's model was still immature and would have limited any efforts to identify potentially exposed individuals.

[31]U.S. Air Force, Technical Applications Center, *Report on Atmospheric Modeling of the 10 Mar 91 Chemical Warfare Agent Release at the Khamisiyah (Iraq) Munitions Pit* (Patrick Air Force Base, Fla.: Dec. 15, 1997).

The MOD Relied on U.S. Plume Modeling to Determine Its Troops' Exposure to CW Agents

According to British officials, the MOD did not collect any source term or meteorological data during the Gulf War. It also did not independently model the plume from Khamisiyah, relying instead on the 1997 DOD and CIA modeling of Khamisiyah. According to British MOD officials, however, they were reassessing the extent of British troops' exposure, based on DOD's revised 2000 remodeling of Khamisiyah. We requested from the British MOD but did not receive information on the findings from this reassessment.

Responding to parliamentary concerns and questions raised in 1997, the British MOD reviewed the U.S. modeling of demolition operations at Khamisiyah, publishing a report in December 1999.[32] The MOD concluded from the 1997 DOD and CIA composite model of the Khamisiyah demolitions that the maximum concentration of agent that British troops might have been exposed to was below the level that the most sensitive British warning device could have been expected to detect. Moreover, according to the MOD, the highest theoretical dosage troops received would have been 3.6 times lower than the level at which the first noticeable symptoms occur. Finally, the MOD said, this level of exposure would have had no detectable effect on health.

The MOD also determined that a number of British troops were within the boundary of the plume in the DOD and CIA composite model, and it estimated that the total number of British troops potentially exposed was about 9,000. The total number of troops definitely within the path of the plume, however, was estimated to be about 3,800. In addition, of 53,500 British troops deployed, at least 44,000 were definitely not within the path of the plume. However, since the MOD relied exclusively on DOD's modeling and since we found that DOD could not know who was and who was not exposed, the MOD cannot know the extent of British troops' exposure.

Total U.S. Plume-Modeling Costs

The CIA and DOD were the primary agencies in the modeling and analysis of U.S. troops' exposure from the demolition at Khamisiyah and bombing of chemical facilities at Al Muthanna, Muhammadiyat, and Ukhaydir, but several other agencies and contractors also participated. Funding to

[32]"Review of Events Concerning 32 Field Hospital and the Release of Nerve Agent Arising from U.S. Demolition of Iraqi Munitions at the Khamisiyah Depot in March 1991," December 1999.

support the modeling efforts was provided to various DOD agencies and organizations, the military services, and non-DOD agencies and contractors. We collected data on the direct costs these agencies incurred or funds they spent. Table 5 shows direct costs to the United States for modeling the Gulf War of about $13.7 million.

Table 5: U.S. Direct Costs for Modeling Gulf War Illnesses

Agency or contractor	Direct costs[a]	Work done
BAHR Inc.	$11,796	Reviewed (1) processes and technology used to produce estimates of U.S. forces potentially exposed and (2) draft reports on Khamisiyah
Central Intelligence Agency	[b]	Computer-modeling analysis
Chemical Biological Defense Command, Aberdeen Proving Ground	140,000	Wood-surface evaporative modeling and environmental data support efforts
Defense Threat Reduction Agency	870,000	Computer-modeling analyses with HPAC/SCIPUFF dispersion and OMEGA weather models
Institute for Defense Analyses	149,429	Convened a panel of experts to review Khamisiyah pit modeling analyses
Lawrence Livermore National Laboratory	60,000	Computer-modeling analyses with ADPIC dispersion and MATHEW weather models
National Center for Atmospheric Research	308,000	Computer-modeling simulations using MM5 weather model
Naval Research Laboratory	1,090,000	Meteorological analysis to identify downwind hazard assessment with NOGAPS and COAMPS weather models.
Naval Surface Warfare Center	522,000	Computer-modeling analyses with VLSTRACK dispersion and COAMPS weather models
Office of the Special Assistant to the Deputy Secretary of Defense for Gulf War Illnesses	7,980,000	Internal costs for producing case narratives for Al Muthanna, Khamisiyah, Muhammadiyat, and Ukhaydir
Science Applications International Corporation	[c]	Computer-modeling analysis
U.S. Army Center for Health Promotion and Preventative Medicine	731,000	Exposure assessment and environmental modeling to determine U.S. troops' exposed to chemical releases from multiple incidents during the Gulf War
U.S. Army Dugway Proving Ground	1,861,950	Field trials and laboratory testing using 122 mm chemical-simulant filled rockets to develop source term data for modeling
White Sands Missile Range	2,600	Missiles for testing at Dugway Proving Ground
Total	**$13,726,775**	

Sources: Agency and contractor responses provided to GAO on their modeling and analysis costs.

[a]Direct costs for agencies includes funding for contracts provided by the Office of the Special Assistant to the Deputy Secretary of Defense for Gulf War Illnesses.

[b]The CIA denied our request for its costs for modeling chemical releases from Khamisiyah, as well as Al Muthanna, Muhammadiyat, and Ukhaydir.

[c]SAIC did not respond to our requests for information.

Indirect costs were much more difficult to obtain because the modeling efforts involved many people, and some were full-time while others were part-time. However, these estimates do not include, and DOD could not provide, an estimate of the considerable indirect costs involved for salaries of DOD and VA staff, contractors, facilities, travel, and equipment. In addition, the CIA would not provide us with the direct and indirect costs for modeling Gulf War plume and determinations of source term because, in its view, our request constituted oversight of an intelligence matter and is beyond our scope of authority. The CIA's contractor, SAIC, also would not respond to our request for cost information.

DOD's and VA's Epidemiological Conclusions on CW Exposure and Hospitalization and Mortality Rates Cannot Be Adequately Supported

DOD and VA each funded an epidemiological study on CW exposure—DOD's on hospitalization rates and VA's on mortality rates. From the hospitalization study conducted by DOD researchers and the mortality study conducted by VA researchers, DOD and VA each concluded that there was no significant difference in the rates of hospitalization and mortality between exposed and nonexposed troops. These conclusions, however, cannot be adequately supported by the available evidence. These studies contained two inherent weaknesses: (1) flawed criteria for classifying exposure, resulting in classification bias, and (2) an insensitive outcome measure, resulting in outcome bias. In addition, several other studies of Gulf War veterans, genetics, and animals found a strong association between exposure and illnesses.

DOD and VA Used Flawed Criteria for Determining Troops' Exposure

In the two epidemiological studies, DOD and VA researchers used DOD's 1997 plume model for determining which troops were under the path of the plume—that is, were exposed—and which troops were not—that is, were not exposed. However, this classification is flawed, given their inappropriate criteria for inclusion and exclusion.

In the hospitalization study, the DOD researchers included 349,291 Army troops "coded" as being in the Army on February 21, 1991. However, the researchers did not report a cutoff date for inclusion in the study—that is, they did not indicate whether these troops were in the Persian Gulf between January 17, 1991, and March 13, 1991, the period during which bombing and demolition took place. Although we requested this date, DOD researchers failed to provide it. Finally, the total number of 349,291 troops is misleading because many troops left the service soon after returning from the Persian Gulf. Moreover, the researchers did not conduct any analyses to determine whether those who left were in the exposed or nonexposed group (including uncertain low-dose exposure or estimated subclinical exposure). Given all the methodological problems in this study, it is not possible to accurately determine the total number of the exposed and nonexposed groups.

In the mortality study, the VA researchers included 621,902 Gulf War veterans who arrived in the Persian Gulf before March 1, 1991. Troops who left before January 17, 1991—the beginning of the bombing of Iraqi research, production, and storage facilities for CW agents—were included in the study. This group was not likely to have been exposed. Therefore, including them resulted in VA's overestimation of the nonexposed group.

Troops who came after March 1, 1991—the period during which Khamisiyah demolition took place—were excluded from the VA study. The Defense Manpower Data Center (DMDC) identified 696,000 troops deployed to the Persian Gulf, but the mortality study included only the 621,902 troops deployed there before March 1, 1991. This decision excluded more than 74,000 troops, approximately 11 percent of the total deployed. However, these troops were most likely to have been exposed. Therefore, excluding them resulted in underestimation of the size of the exposed group. In addition, 693 troops who were in the exposed group were excluded because identifying data, such as Social Security numbers, did not match the DMDC database. But VA researchers did not conduct follow-up analysis to determine whether those who were excluded differed from those who were included in ways that would affect the classification.

DOD and VA Used an Insensitive Outcome Measure for Determining Hospitalization Rate

Hospitalization rate—the outcome measure used in the hospitalization study—was insensitive because it failed to capture the chronic illnesses that Gulf War veterans commonly report but that typically do not lead to hospitalization. Studies that rely on this type of outcome as an end point are predetermined to overlook any association between exposure and illness.

Based on DOD's 1997 plume model, DOD's hospitalization study compared the rates for Gulf War veterans who were exposed with the rates for those who were not exposed. This study included 349,291 active duty Army troops who were deployed to the Persian Gulf. However, DOD researchers did not determine the resulting bias in their analyses, because they did not account for those who left the service.

The Institute of Medicine noted that the hospitalization study was limited to Army troops remaining on active duty and to events occurring in military hospitals. Conceivably, those who suffered from Gulf War–related symptoms might leave active duty voluntarily or might take a medical discharge. Hospitalization for this group would be reflected in VA or private sector databases but not in DOD databases. The health or other characteristics of active duty troops could differ from those of troops who left active duty and were treated in nonmilitary hospitals. Moreover, economic and other factors not related to health are likely to affect the use of nonmilitary hospitals and health care services.[33]

This limiting of the study to troops remaining on active duty produced a type of selection bias known as the healthy warrior effect.[34] It strongly biased the study toward finding no excess hospitalization among the active duty Army troops compared with those who left the service after the war.

Some Studies Suggest an Association between CW Exposure and Gulf War Illnesses

We found some studies that suggest an association between CW exposure and Gulf War illnesses. They are Gulf War veterans studies, genetics studies, and animal studies. Each of these studies, described below, has strengths and limitations.

Gulf War Veterans Studies

In a privately funded study, Haley and colleagues reported an association between a syndromic case definition of Gulf War illnesses, developed to model the ill veterans' symptomatic complaints, with exposure to CW agents.[35] In this study, the authors developed questionnaires on symptoms and environmental exposure identified in pilot studies of ill Gulf War veterans, similar to epidemic investigations by the Centers for Disease Control and Prevention (CDC).[36] These questionnaires were given to 249 troops from a U.S. Navy Mobile Construction Battalion that participated in the Gulf War. Factor analysis of the data on symptoms was used to derive

[33]Institute of Medicine, *Gulf War Veterans: Measuring Health* (Washington, D.C.: National Academy Press, 1999), p. 36.

[34]R. W. Haley, "Point: Bias from the 'Healthy-Warrior Effect' and Unequal Follow-Up in Three Government Studies of Health Effects of the Gulf War," *American Journal of Epidemiology* 148 (1998): 315–38.

[35]R. W. Haley and T. L. Kurt, "Self-Reported Exposure to Neurotoxic Chemical Combinations in the Gulf War," *JAMA* 277 (1997): 231–37.

[36]See Michael B. Gregg, ed., *Field Epidemiology*, 2nd ed. (New York: Oxford University Press, 2002).

a case definition identifying six syndrome factors.[37] Three syndrome factor variants found to be the most significant were (1) impaired cognition, (2) confusion-ataxia, and (3) arthro-myo-neuropathy.

Impaired cognition (syndrome 1) was associated with troops' having worn flea collars that contained chlorpyrifos.[38] Confusion-ataxia (syndrome 2), the most severe clinically, was associated with three risk factors.[39] The first was likely CW exposure; the second was the geographic location near the Saudi-Kuwaiti border around the fourth day of the air war, conducted January 18–23, 1991, when Czech chemical detection units detected sarin and mustard in ambient air near the Saudi-Kuwaiti border; and the third was side effects experienced after taking pyridostigmine. There was also a significant synergistic association between likely exposure to CW agents and the number of side effects from pyridostigmine.[40] Arthro-myo-neuropathy (syndrome 3) was associated with the amount of exposure to 95 percent DEET in ethanol insect repellent and with the number of side effects of pyridostigmine.[41]

[37]R. W. Haley and others, "Is There a Gulf War Syndrome? Searching for Syndromes by Factor Analysis of Symptoms," *JAMA* 277 (1997): 215–22. The six syndrome factors were impaired cognition, confusion-ataxia, arthro-myo-neuropathy, phobia-apraxia, fever-adenopathy, and weakness-incontinence.

[38]Impaired cognition is characterized by problems with attention, memory, and reasoning, as well as insomnia, depression, daytime sleepiness, and headache. (Study results showed relative risk 8.2, 95 percent, CI 2.9–23.5, $p = 0.001$.)

[39]Confusion-ataxia is characterized by problems with thinking, disorientation, balance disturbances, vertigo, and impotence.

[40](1) CW exposure, relative risk 7.8, 95 percent, CI 2.3–25.9, $p < 0.0001$; (2) geographic location, relative risk 4.3, 95 percent, CI 1.9–10.0, $p = 0.004$; (3) pyridostigmine side effects, dose-response trend up to relative risk 32.4, 95 percent, CI 7.8–135.0, $p < 0.0001$; (4) synergistic association, Rothman synergy statistic 5.3, 95 percent, CI 1.04–26.7, $p < 0.05$. See Jonathan B. Tucker, "Evidence Iraq Used Chemical Weapons during the 1991 Persian Gulf War," *The Nonproliferation Review* 4:3 (Spring–Summer 1997): 114–22. Center for Nonproliferation Studies, Monterey Institute of International Studies, http://cns.miis.edu/pubs (Apr. 28, 2004); and U.S. Department of Defense, Office of the Special Assistant for Gulf War Illnesses, *Coalition Chemical Detections and Health of Coalition Troops in Detection Area* (Washington, D.C.: Aug. 5, 1996). http://www.gulflink.osd.mil/czech_french/czfr_refs/n08en011/coalitn.html (Apr. 28, 2004).

[41]Arthro-myo-neuropathy is characterized by joint and muscle pains, muscle fatigue, difficulty lifting, and paresthesias of the extremities. (Results showed for exposure, dose-response effect to relative risk 7.8, 95 percent, CI 2.4–24.7, $p < 0.0001$; for side effects, dose-response effect to relative risk 3.9, 95 percent, CI 1.3–12.1, $p < 0.0001$.)

The inference that these risk-factor associations represent causal effects is supported by (1) the large, highly significant relative risks; (2) the dose-response effects; and (3) the synergistic effect of the risk factor associations with the syndromic case definition. Risk factors found not to be significantly associated with the case definition include environmental pesticides, pesticides in uniforms, antibiotic or antimalarial prophylaxis, multiple immunizations, smoke from oil well fires, fumes from jet fuel, fumes from burning jet fuel in tents, petroleum in drinking water, depleted uranium munitions, smoking, alcohol use, and combat exposure.

Another study of Gulf War veterans by Nisenbaum and colleagues, funded by CDC, examined the risk factors in 1,002 Air Force reservists.[42] They found, first, that the case definition of Fukuda and colleagues of "multisymptom illness" was strongly associated with at least one of the three chronic symptom groups fatigue, mood/cognition, and musculoskeletal pain. And, next, they found that reported exposure to CW agents was most strongly associated with the "severe illness" case definition of Fukuda and colleagues and less strongly associated with their "mild–moderate illness" case definition.[43]

Both case definitions were less strongly associated with the use of insect repellent ($p = 0.006$), the use of pyridostigmine ($p = 0.01$), and having an injury requiring medical attention ($p = 0.03$). But neither case definition was associated with smoke from oil well fires, coming under attack, seeing casualties, or having adverse health events in the family. The findings were attributed to the effects of stress but offered no empirical support for the explanation.

In a study that VA funded, Proctor and colleagues compared the exposure histories of 186 Gulf War veterans from Fort Devens, Massachusetts, and 66 from New Orleans, including 48 who deployed only to Germany. Collectively, the 252 veterans are known as the Massachusetts–New

[42]R. Nisenbaum and others, "Deployment Stressors and a Chronic Multisymptom Illness among Gulf War Veterans," *Journal of Nervous and Mental Disease* 188 (2000): 259–66.

[43]Association with "severe illness," adjusted OR 3.46, 95 percent, CI 1.73–6.91, $p < 0.0001$; association with "mild–moderate illness," adjusted OR 2.25, 95 percent, CI 1.54–3.27, $p < 0.0001$. See K. Fukuda and others, "A Chronic Multisymptom Illness Affecting Air Force Veterans of the Persian Gulf War," *JAMA* 280 (1998): 981–88.

Orleans cohort.[44] The case definition was a set of eight body-system symptom scores (BSS, distributed from 0 to 4), each constructed by summing the 5-point frequency-of-occurrence scales (0 = occurs never, 4 = occurs almost every day) for three symptoms typical of a particular body system. The eight body systems were cardiac, dermatological, gastrointestinal, musculoskeletal, neurological, neuropsychological, psychological, and pulmonary. Post-traumatic stress disorder (PTSD) was diagnosed with the structural clinical interviews, Clinician Administered Posttraumatic Stress (CAPS) disorder scale, or a Mississippi Scale score of >89. The symptoms were obtained from the 52-item Expanded Health Symptom Checklist, the exposure measures from an environmental exposure questionnaire and an Expanded Combat Exposure Scale (CES) questionnaire. Multiple regression analysis—controlling for age, sex, education, study site, Expanded CES score, and PTSD status—was used to develop a risk-factor model for each BSS scale.

Exposure to CW agents and debris from SCUD missiles was associated with four BSS scales; exposure to smoke from tent heaters, with three BSS scales; exposure to pesticides, vehicle exhaust, and burning human waste, with two BSS scales; the Expanded CES, with only one BSS scale; and exposure to pyridostigmine bromide (antinerve gas pills) and smoke from oil well fires, with no BSS scale. Controlling for depression scores and excluding veterans diagnosed with PTSD did not substantially affect the associations.

Three additional studies conducted with VA and DOD funding extended the risk-factor research for the Massachusetts–New Orleans cohort. The association of self-reported CW agent (nerve agent) exposure was tested with different formulations of the case definition. White and colleagues used psychological and neuropsychological tests to define illness. They found that exposure to CW agents was associated with abnormal measures of mood, memory, and attention or executive function.[45] Associations remained significant after controlling for age, sex, race, years of education, repeated grade in school, head injury, medication use, diagnosis of current

[44] S. P. Proctor and others, "Health Status of Persian Gulf War Veterans: Self-Reported Symptoms, Environmental Exposures, and the Effect of Stress," *International Journal of Epidemiology* 27 (1998): 1000–10.

[45] R. F. White and others, "Neuropsychological Function in Gulf War Veterans: Relationships to Self-Reported Toxicant Exposures," *American Journal of Industrial Medicine* 40 (2001): 42–54.

PTSD (by CAPS), diagnosis of current depression (by structural clinical interviews), active duty versus Reserve or Guard status, seeking disability rating, and Vietnam service.

Lindem and colleagues developed multiple regression models for neuropsychological test measures as case definitions of Gulf War illnesses.[46] Chemical warfare agent exposure was found to be associated with attention and executive function (continuous performance test), delayed verbal recall (California Verbal Learning Test and Visual Reproduction Test), and confusion and fatigue (Profile of Mood States). These associations remained significant when controlling for age, education, and PTSD diagnosis (by CAPS).

Wolfe and colleagues, studying 945 troops from the Massachusetts–New Orleans cohort, found that the CDC case definition of multisymptom illness was most strongly associated with having smelled a chemical odor, having taken up to 21 antinerve gas pills, or having experienced up to 10 formal alerts for CW agent attack.[47]

Kang and colleagues conducted a random sample mail survey that VA funded. Obtaining responses from 11,441 Gulf War veterans and 9,476 nondeployed Gulf War era veterans, they developed a case definition by factor analysis of symptoms measured by their questionnaire.[48] The first three syndrome factors closely resembled those that Haley and others derived (noted earlier). Finding that syndrome 2 was unique to the sample that had been deployed in the Gulf War (found in the deployed, but not the nondeployed, sample) and that the component symptoms were neurological in character, the researchers termed their syndrome 2 a possible unique Gulf War neurological syndrome. Four symptoms—blurred vision, loss of balance or dizziness, tremor or shaking, and speech difficulties—were associated with syndrome 2 only in the deployed sample. Consequently, Kang and colleagues established their case

[46]K. Lindem and others, "Neuropsychological Performance in Gulf War Era Veterans: Traumatic Stress Symptomatology and Exposure to Chemical-Biological Warfare Agents, *Journal of Psychopathology and Behavioral Assessment* 25:2 (2003): 105–19.

[47]Chemical odor, OR = 6.2, 95 percent, CI 3.9–9.9; antinerve gas pills, OR = 3.7, 95 percent, CI 2.4–5.6; formal alerts for CW attack, OR = 2.7, 95 percent, CI 2.0–3.7. See J. Wolfe and others, "Risk Factors for Multisymptom Illness in U.S. Army Veterans of the Gulf War," *Journal of Occupational and Environmental Medicine* 44:3 (2002): 271–81.

[48]H. K. Kang and others, "Evidence for a Deployment-Related Gulf War Syndrome by Factor Analysis," *Archives of Environmental Health* 57:1 (2002): 61–68.

definition as having all four of these symptoms. In the deployed sample, 277 met the case definition and 6,730 who had none of the four symptoms constituted the control group. Of a large number of risk factors analyzed, only nine were associated with the case definition, with an odds ratio greater than 3.0. Of these, perceived exposure to nerve agent had the strongest association (odds ratio 15.1, 95 percent, CI 11.5–19.9, $p < 0.000001$). This finding—a neurological syndrome appearing as the second factor in a factor analysis and being the most strongly associated risk factor, 15 times more common in ill veterans meeting the case definition than in controls—closely parallels the findings of Haley and colleagues. The finding received little notice, however, because the VA-funded mail survey did not (1) provide the odds ratio values in the table reporting the risk factor analysis results and (2) describe the finding in the text or abstract of the paper. When we noticed the finding, we manually calculated the odds ratios from the raw data in the table.

Smith and colleagues showed that hospitalization rates for several ICD-9 diagnoses were higher in veterans categorized in the Khamisiyah 2000 plume than in those not in the plume, and the association for cardiac arrhythmias was statistically significant. However, this study suffers from the same deficiencies as the earlier study that we cited: use, inappropriately, of hospitalization outcome measures rather than measures of Gulf War illness, which usually do not result in hospitalization, and use of plume modeling based on flawed data.[49]

The 2002 Kang and Bullman study has not been published in a peer-reviewed journal and therefore should not have been included in a review of the scientific epidemiologic literature. The DOD studies were invalid for two reasons: (1) Hospitalization and mortality were inappropriate outcomes because they do not measure Gulf War illnesses, which often do not lead to hospitalization, and (2) The DOD studies, no matter how powerful their techniques, could not control for the selection bias that resulted from the disproportionate early discharge of the ill Gulf War veterans soon after the Gulf War. Including only DOD hospital records of service members remaining on active duty resulted in the exclusion of veterans who left service for poor health. No amount of sophisticated

[49]T. C. Smith and others, “Gulf War Veterans and Iraqi Nerve Agents at Khamisiyah: Postwar Hospitalization Data Revisited,” *American Journal of Epidemiology* 158 (2003): 457–67.

techniques can correct for this selection bias toward finding no difference.[50]

Genetics Studies

In one genetics study, Haley and colleagues found an association between the case definition of Gulf War illnesses in U.S. Gulf War veterans and low blood levels of the Q-type isoenzyme of the paraoxonase/arylesterase enzyme group (PON).[51] The PON group of enzymes is a potentially important predisposing factor in Gulf War illnesses because one of its major functions in normal body physiology is to protect the nervous system from organophosphate chemical toxins, such as pesticides and nerve agents. This finding was remarkable because the only function of Q type of the PON enzyme group is to protect the nervous system from nerve agents sarin, soman, tabun, and VX. The R-type isoenzyme has as its main function protection from organophosphate pesticides, such as diazinon, malathion, and parathion. Thus, an association between Gulf War illnesses and blood levels of only the Q-type isoenzyme of PON points specifically to nerve agent exposure. In addition, the total PON level—that is, the sum of the Q and R isoenzyme levels—was not associated with the illnesses. And the genotype (QQ, QR, or RR) was only marginally associated with them, as expected, because the level of the Q-type isoenzyme is a more important determinant of susceptibility to nerve agents than the genotype.

In another genetics study, Mackness and colleagues reported lower blood levels of total PON in ill British Gulf War veterans than in civilian controls in a previously published study; however, they did not measure the blood levels of the Q and R isoenzymes of PON, needed for a definitive study of Haley's hypothesis.[52] This finding could indicate that ill British Gulf War veterans represented a mixture of some with low Q-type PON and others with low R-type PON. In some veterans, the illness would be associated with exposure to nerve agents; in others, with exposure to pesticides. Alternatively, the difference in total PON levels may have resulted from differences in the assays or in the veterans, since (1) the enzyme assays in

[50]H. K. Kang and T. A. Bullman, *Mortality among U.S. Gulf War Veterans Who Were Potentially Exposed to Nerve Gas at Khamisiyah, Iraq* (Washington, DC: Department of Veterans Affairs, May 2002).

[51]R. W. Haley and others, "Association of Low PON1 Type Q (Type A) Arylesterase Activity with Neurologic Symptom Complexes in Gulf War Veterans," *Toxicology and Applied Pharmacology* 157 (1999): 227–33.

[52]B. Mackness and others, "Low Paraoxonase in Persian Gulf War Veterans Self-Reporting Gulf War Syndrome," *Biochemical and Biophysical Research Communications* 276 (2000): 729–33.

the controls were performed years before those for the ill veterans and (2) the controls were civilians studied in an entirely different setting.

In yet a third genetics study, Hotopf and colleagues reported results of tests for total PON levels in blood samples—obtained in a study by Unwin and colleagues—for four groups of British troops: (1) ill veterans of the Gulf War, (2) well veterans of the Gulf War, (3) ill nondeployed veterans of the Gulf War era, and (4) ill veterans of the Bosnian conflict.[53] The case definition of illness was a score below 72.2 on the SF-36 Physical Status questionnaire. Again, the researchers did not measure the levels of the Q and R isoenzymes of PON, making the findings difficult to interpret. The researchers found a low mean level of total PON in both ill and well groups deployed to the Gulf War and higher levels in the Gulf War era and ill Bosnian groups.

The depressing of the total PON level, the researchers suggested, might be the result of some deployment-related exposures. However, instead of looking at exposure to CW agents, the researchers investigated the possible effect of multiple immunizations on total PON levels and found no evidence for it. An alternative explanation is that total PON level in both ill and well deployed veterans was the result of misclassification of veterans by the case definition. A score of 72.2 on the SF-36 scale is not a very low score, particularly in ill Gulf War veterans, and it is a nonspecific measure of illness, given that a low score indicates illness from any cause.[54] Consequently, many veterans ill from causes unrelated to the war would be misclassified as cases of Gulf War illness and, conversely, many ill from the war but with less disability would be misclassified as controls. This conclusion is supported by a nonsignificant trend showing that ill veterans who had been deployed to the Gulf War had a lower median total PON level than well veterans who had also been deployed to the Gulf War.

The many flaws of design and methodology in both British studies of PON levels do not contribute to an understanding of the PON hypothesis and leave the finding of Haley and colleagues in need of better replication.

[53] See Matthew Hotopf and others, "Paraoxonase in Persian Gulf War Veterans," *Journal of Occupational and Environmental Medicine* 45 (2003): 668-75, and C. Unwin and others, "Health of UK Servicemen Who Served in the Persian Gulf War," *Lancet* 353 (1999): 169–78.

[54] R. W. Haley and others, "Severely Reduced Functional Status in Veterans Fitting a Case Definition of Gulf War Syndrome," *American Journal of Public Health* 92 (2002): 46–47.

Animal Studies

A series of laboratory studies with animals have established the biological plausibility that brain cell damage results from low-level exposure to sarin. Husain and colleagues demonstrated in two studies at the Division of Pharmacology and Toxicology at the Defense Research and Development Establishment in Gwalior, India, that repetitive administration of low-dose sarin (approximately 0.25 LD50) daily for 10 days caused delayed onset damage to neurons in the spinal cords and brains of mice exposed by inhalation and of hens exposed by subcutaneous injection.[55]

Privately funded studies by Abou-Donia and colleagues demonstrated that combinations of organophosphates and similar cholinesterase-inhibiting chemicals in hens produce greater neurotoxic effect on brain and nerve tissue than any of the agents alone.[56] Abou-Donia's subsequent work, funded by DOD, extended the findings to synergistic combinations involving sarin at moderate concentrations (0.5 LD50).[57] A similar study by Husain and Somani, also funded by DOD, on the delayed brain effects of low-dose sarin (0.05 LD50) in combination with pyridostigmine and exercise, confirmed these findings. In particular, it demonstrated that the neuronal damage from very low doses of sarin affected primarily the basal ganglia region of the brain (striatum).[58]

A study by Henderson and colleagues, with DOD funding, found that repeated inhalation exposure to low-level sarin at subsymptomatic doses

[55]K. Husain and others, "Delayed Neurotoxic Effect of Sarin in Mice after Repeated Inhalation Exposure," *Journal of Applied Toxicology* 13 (1993): 143–45, and "A Comparative Study of Delayed Neurotoxicity in Hens Following Repeated Administration of Organophosphorus Compounds," *Indian Journal of Physiology and Pharmacology* 39 (1995): 47–50.

[56]Mohamed B. Abou-Donia and others, "Neurotoxicity Resulting from Coexposure to Pyridostigmine Bromide, DEET, and Permethrin," *Journal of Toxicology and Environmental Health* 48 (1996): 35–56, and "Increased Neurotoxicity Following Concurrent Exposure to Pyridostigmine Bromide, DEET, and Chlorpyrifos," *Fundamentals of Applied Toxicology* 34 (1996): 201–22.

[57]Mohamed B. Abou-Donia and others, "Combined Exposure to Sarin and Pyridostigmine Bromide Increased Levels of Rat Urinary 3-Nitrotyrosine and 8-Hydroxy-2'-Deoxyguanosine, Biomarkers of Oxidative Stress," *Toxicology Letters* 123 (2001): 51–58; "Disruption of the Blood-Brain Barrier and Neuronal Cell Death in Cingulate Cortex, Dentate Gyrus, Thalamus, and Hypothalamus in a Rat Model of Gulf-War Syndrome," *Neurobiology of Disease* 10 (2002): 306–26; and "Sarin: Health Effects, Metabolism, and Methods of Analysis," *Food and Chemical Toxicology* 40 (2002): 1327–33.

[58]K. Husain and S. Somani, "Delayed Toxic Effects of Nerve Gas Sarin and Pyridostigmine under Physical Stress in Mice," *Journal of Burns and Surgical Wound Care* 2 (2003): 2–19.

(0.1 LCt50) for 5 or 10 days, with or without heat stress, produced no immediate effects.[59] But at 30 days after exposure to sarin, damage was produced to cholinergic receptors in several brain regions, including the basal ganglia. In the same study, Henderson and colleagues identified evidence of an autonomic nervous system injury affecting the function of T-cells in the immune system as well.[60] In addition, chronic abnormalities of neuronal metabolism in the basal ganglia have been implicated in ill Gulf War veterans by several investigators through the use of magnetic resonance spectroscopy.[61]

Two recent laboratory studies at the U.S. Army Medical Research Institute of Chemical Defense, Aberdeen Proving Ground, support the animal studies. Scremin and colleagues demonstrated that moderate doses of sarin (0.5 LD50) in combination with pyridostigmine bromide produced prolonged elevations in rats' cerebral blood flow but that neither agent alone had a prolonged effect on cerebral blood flow.[62] A companion study, by Roberson and colleagues, demonstrated that repeated administration of sarin to guinea pigs in doses of 0.2 or 0.4 LD50 produced no immediate ill effects on behavior, weight, body temperature, flinch threshold, or EEG brain wave activity. But at 100 days postdosing, abnormal brain function was found, indicating neurochemical or pathological brain cell changes that affect behavior.[63]

[59] R. F. Henderson and others, "Response of F344 Rats to Inhalation of Subclinical Levels of Sarin: Exploring Potential Causes of Gulf War Illness," *Journal of Toxicology and Industrial Health* 17 (2001): 294–97 and 18:1 (2002): 48.

[60] See Henderson and others, "Response of Rats to Low Levels of Sarin," and "Subclinical Doses of the Nerve Gas Sarin Impair T Cell Responses through the Autonomic Nervous System," *Journal of Toxicology and Applied Pharmacology* 184 (2002): 82–87.

[61] See R. W. Haley and others, "Brain Abnormalities in Gulf War Syndrome: Evaluation by ^{1}H Magnetic Resonance Spectroscopy," *Radiology* 215 (2000): 807–17, and "Effect of Basal Ganglia Injury on Central Dopamine Activity in Gulf War Syndrome: Correlation of Proton Magnetic Resonance Spectroscopy and Plasma Homovanillic Acid," *Archives of Neurology* 57 (2000): 1280–85, as well as D. J. Meyerhoff and others, "Reduced N-Acetylaspartate in the Right Basal Ganglia of Ill Gulf War Veterans by Magnetic Resonance Spectroscopy," *Proceedings of the International Society of Magnetic Resonance Medicine* 9 (2001): 994.

[62] O. U. Scremin and others, "Effects of Chronic Exposure to Low Levels of Cholinesterase Inhibitors on Cerebral Blood Flow," paper for the Society for Neuroscience Meeting, Orlando, Florida, 2002.

[63] Melinda Roberson and others, "Depression of Cholinesterase Activity by Low-Dose Sarin Exposure May Lead to Persistent Changes That Influence Behavior," Society for Neuroscience, Washington, D.C., Program no. 205.3 (Abstract, 2002).

Conclusions

In evaluating the weaknesses of the plume models, we conclude that the results from the CIA and DOD modeling can never be definitive. Plume models can allow only estimates of what happens when CW agents are released in the environment. Such estimates are based on mathematical equations, which are used to predict an actual event—in this case, the direction and extent of the plume. However, in order to predict precisely what happens, one needs to have accurate data on source term and meteorological conditions. DOD had neither of these.

Given the unreliability of the input data, the lack of troop location data, and the divergent results of modeling, DOD's analyses cannot adequately determine the extent of U.S. troops' exposure. In particular, the models selected were not fully developed for projecting long-range environmental fallout, and the assumptions used to provide the source term data were flawed. Even when models used the same source term data, their results diverged. In addition, the models did not include many potential exposure events and some key materials—for example, binary chemical weapons, mustard agent combustion by-products, and CW agent precursor materials. It is likely that if models were more fully developed, and if more credible data for source term and meteorological conditions were included in them, particularly with respect to plume height as well as level and duration of exposure, the hazard area would be much larger and most likely would cover most of the areas where U.S. troops and Coalition forces were deployed. However, given the lack of verifiable data for analysis, it is unlikely that any further modeling efforts would be more accurate or helpful.

The results of DOD's modeling efforts were, nonetheless, used in epidemiological studies to determine the troops' CW exposure classification—exposed versus nonexposed. As we noted in 1997, to ascertain the causes of veterans' illnesses, it is imperative that investigators have valid and reliable data on exposure, especially for low-level or intermittent exposures to CW agents.[64] To the extent that veterans are misclassified as to exposure, relationships will be obscured and conclusions misleading. In addition, DOD combined the results of individual models that showed smaller plume size and ignored the results

[64]U.S. General Accounting Office, *Gulf War Illnesses: Improved Monitoring of Clinical Progress and Reexamination of Research Emphasis Are Needed*, GAO/NSIAD-97-163 (Washington, D.C.: June 23, 1997).

of the LLNL model, which showed much larger plume size. Given the uncertainties in source term data and divergences in model results, DOD cannot determine—with any degree of certainty—the size and path of the plumes or who was or who was not exposed.

Recommendations for Executive Action

We recommend that the Secretary of Defense and the Secretary of Veterans Affairs not use the plume-modeling data for future epidemiological studies of the 1991 Gulf War, since VA and DOD cannot know from the flawed plume modeling who was and who was not exposed.

We recommend that the Secretary of Defense require no further plume modeling of Khamisiyah and the other sites bombed during the 1991 Gulf War in order to determine troops' exposure. Given the uncertainties in the source term and meteorological data, additional modeling of the various sites bombed would likely result in additional cost while still not providing DOD with any definitive data on who was or was not exposed.

Agency Comments and Our Evaluation

We obtained comments on a draft of this report from VA, DOD, and the CIA. VA concurred with our first recommendation (see appendix V). Nevertheless, VA stated that it has already completed three studies that incorporated the DOD plume model as part of the parameters for the research and has submitted these studies to scientific journals for publication. In addition, VA is currently collaborating with other research groups that may have used the DOD plume model. These studies are under way and will be completed as planned. Given our assessment, it is important that VA inform the researchers to include appropriate caveats, indicating the limitation of the conclusions based on flawed modeling data.

DOD did not concur with our first recommendation, indicating that the "GAO recommendation apparently represents a blanket prohibition against plume modeling in the future, where the limitations of the 1991 Gulf War may not apply" (see page 77). The intent of our recommendation was directed only at epidemiological studies involving the DOD and CIA plume modeling data from the 1991 Gulf War and was not a blanket prohibition of plume modeling in the future (see appendix VI). We have clarified the recommendation along these lines.

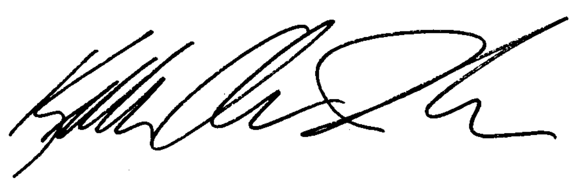
The CIA did not concur with the report, indicating that it could not complete a review in the time allotted. The CIA indicated that a comprehensive review would require 3 to 4 weeks. Nevertheless, from its preliminary review of our report, the CIA identified several statements that it considered inaccurate, such as those about source term data. The CIA contended that the agent source term is complete and accurate to a known certainty. Since our initiation of this inquiry in late 2002, we have met with the CIA on a number of occasions, most recently on April 7, 2004, to identify the methodologies pursued in establishing source term parameters used in the modeling. At the suggestion of the CIA, we met with UNMOVIC officials to determine what UNSCOM inspections disclosed and the methodologies used in determining source term data.

Our point-by-point evaluation of the detailed comments provided by DOD are presented in appendix VI.

As we agreed with your offices, unless you publicly announce the contents of this report earlier, we plan no further distribution until 30 days from its issue date. We will then send copies of the report to other interested congressional members and committees. In addition, the report will be available at no charge on GAO's Web site at http://www.gao.gov.

If you or your staff have any questions about this report or would like additional information, please contact me at (202) 512-6412 or Sushil Sharma, Ph.D., Dr.PH., at (202) 512-3460. We can also be reached by e-mail at rhodesk@gao.gov and sharmas@gao.gov.

Individuals who made key contributions to this report were Venkareddy Chennareddy, Susan Conlon, Neil Doherty, Jason Fong, Penny Pickett, Laurel Rabin, Katherine Raheb, and Joan Vogel. James J. Tuite III, a GAO consultant and recognized expert on Gulf War issues, provided technical expertise.



Keith Rhodes, Chief Technologist
Center for Technology and Engineering
Applied Research and Methods

Appendix I: DOD's Chronology of Khamisiyah Modeling Events

Date	Event
1995	
June	DOD formed the Persian Gulf Illnesses Investigation Team; by October, it had identified some of the U.S. forces that had occupied the area around Khamisiyah during the Gulf War, including the 37th Engineer Battalion
Aug.	President Bill Clinton created the Presidential Advisory Committee on Gulf War Veterans' Illnesses
1996	
Spring	The Presidential Advisory Committee on Gulf War Veterans' Illnesses directed the CIA to model chemical agent release from Bunker 73
May	UNSCOM inspected Khamisiyah
June	DOD confirmed publicly that "US soldiers from the 37th Engineer Battalion destroyed ammunition bunkers [at Khamisiyah] in early March 1991.... It now appears that one of these destroyed bunkers contained chemical weapons"
July	The CIA briefed the Presidential Advisory Committee on Bunker 73 modeling results
Oct.	The CIA requested LLNL to perform atmospheric dispersion calculations, using a hypothetical release scenario; the Deputy Secretary of Defense sent a memorandum to 21,000 veterans who had been identified as being within 50 km of Khamisiyah
Nov.	The Secretary of Defense established the Office of the Special Assistant for Gulf War Illnesses (OSAGWI) to focus ongoing DOD investigations and expand the investigation into Gulf War veterans' complaints of undiagnosed illnesses
Dec.	IDA released its interim report, critical of the model the CIA used; it recommended rocket demolition testing to determine how rockets behaved without high explosives and an ensemble approach with prognostic models
1997	
Jan.	The Deputy Secretary of Defense sent letters with a survey to veterans, saying that chemical weapons had been present at Khamisiyah when the demolitions occurred and urging them to call the Persian Gulf Incident Hotline with any additional information they had about Khamisiyah
Jan.–Feb.	The Special Assistant agreed to remodel Khamisiyah, as well as Al Muthanna, Muhammadiyat, and Ukhaydir, following entreaties by the Presidential Advisory Committee on Gulf War Veterans' Illnesses and IDA's recommendations
May	DOD and the CIA conducted open-field demolition tests on 122 mm rockets similar to those destroyed in the Khamisiyah pit at Dugway Proving Ground, Utah, to determine how CW agents in Iraq's rockets might have been released
May (to June 1998)	The Department of the Army and OSAGWI hosted G3/S3 conferences to elicit more correct information on unit locations during the war
June–July	CW agent release was modeled at the pit
July	IDA released its final report, "Report of the Panel Reviewing Analysis of the Khamisiyah Pit Release of Nerve Agent, March 1991"; DOD and the CIA jointly announced the results of Khamisiyah dispersion modeling. Given the unit locations available then, the modeling indicated a hazard area where troops may have been exposed to low levels of nerve agent. DOD sent written notices to 98,910 veterans in the potential hazard area and approximately 10,000 notices to those who had received the Deputy Secretary of Defense's letter and survey but were not in the potential hazard area
Dec.	An independent scientific panel (Anthes and others) reviewed and commented on the methodology used to complete the Khamisiyah modeling, making recommendations for improvements in future modeling[a]

Date	Event
1998: Jan.	President Clinton created the Presidential Special Oversight Board for the Department of Defense Investigations of Gulf War Chemical and Biological Incidents to provide recommendations, based on its review of DOD investigations into possible detection of, and exposures to, chemical or biological weapons agents, as well as environmental and other factors that might have contributed to Gulf War illnesses
2000	
Jan.	DOD completed efforts to remodel the Khamisiyah release using updated meteorological and dispersion models, with revised source terms from the CIA and with a better understanding of where U.S. forces had been. As a result, DOD's estimate of the number possibly exposed increased by about 2,000; almost 35,000 troops who had previously been notified of possible exposure were no longer in the possible hazard area, whereas about 37,000 newly identified troops probably were
Mar.	The scientific review committee completed its review of the revised methodology, commenting that "the methodologies are sound" and "the results ... very likely overestimate the dosages actually received by personnel"
2002: Apr.	DOD published its final Khamisiyah report, with complete technical report documenting methodologies

Source: Department of Defense, Office of the Assistant Secretary of Defense for Health Affairs, Deployment Health Support Directorate.

[a] Richard A. Anthes and others, "Comments by Review Panel on Khamisiyah Modeling Report and Presentations on November 4–5, 1997," report for the Directorate for Deployment Health Support of the Special Assistant to the Under Secretary of Defense (Personnel and Readiness) for Gulf War Illnesses, Medical Readiness, and Military Deployments, Fairfax, Virginia, December 11, 1997.

Appendix II: Power-Law Formula

A 1969 Sandia National Laboratories empirical study established a power-law formula for calculating plume heights attributable to high-explosive detonations. The power-law formula was derived from data on 23 test shots, ranging from 140 lbs. to 2,242 lbs. of high explosives at DOE's Nevada Test Site (the National Exercise, Test, and Training Center), and it provides a cloud-top height at 2 minutes after detonation. Most of the shots were detonated during near-neutral conditions, where the clouds continued to rise after 2 minutes; data for 5 minutes after detonation on some shots show tops rising to nearly double the 2-minute values. The 2-minute values better represent the final cloud-top heights during stable conditions.

This formula is represented as

$$h = 76\ (w^{1/4})$$

where

h = height of plume in meters and
w = weight of explosives in pounds

According to this formula, an MK-84 or GBU-24 bomb (942.6 lbs. of high explosives) would generate a plume of 421 meters:

$$\text{h} = 76\ (942.6 \text{ pounds of high explosives})^{1/4}$$

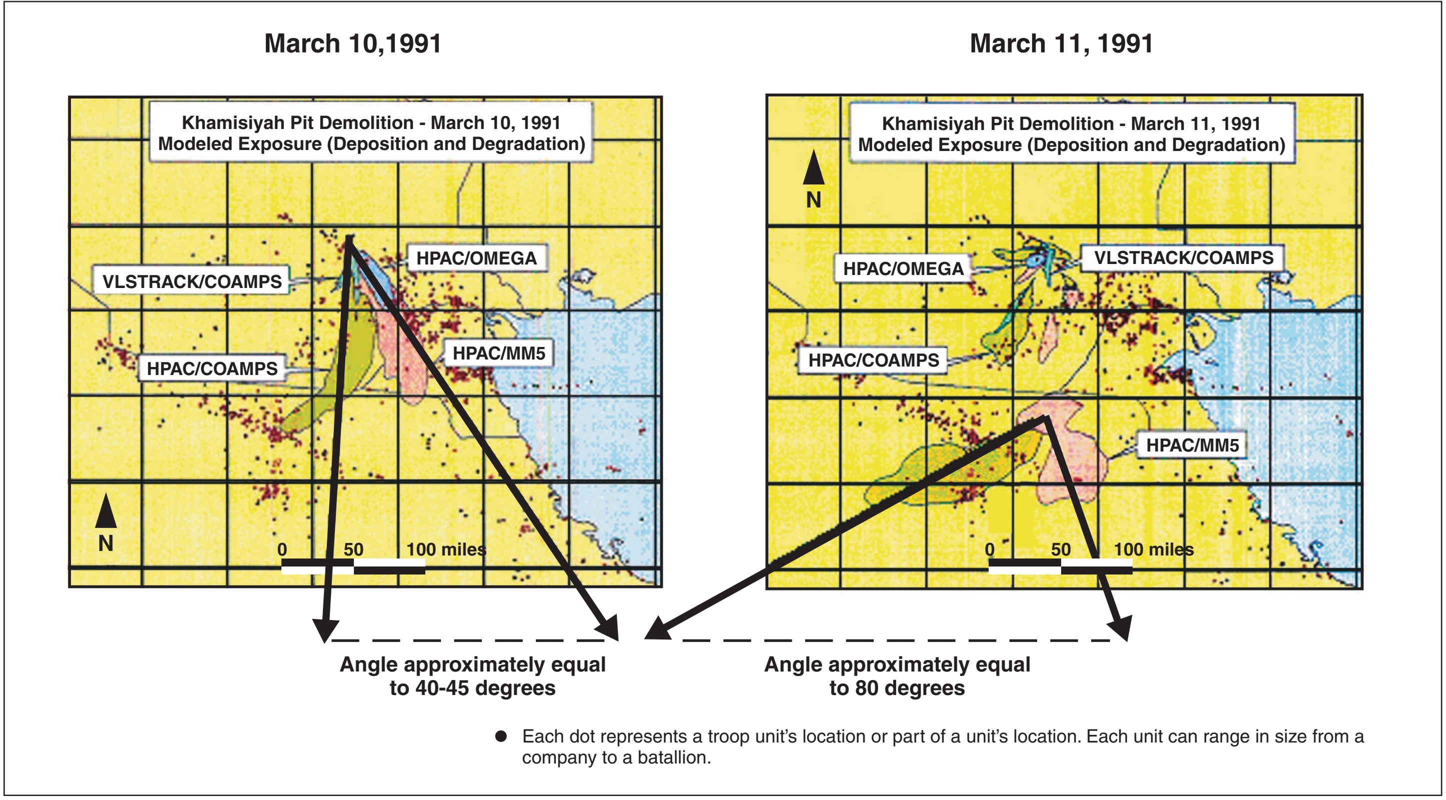
$$\text{h} = 76\ (5.541)$$

h • 421 meters

Appendix III: DOD's Model Divergences

Even among the models selected for use by the DOD panel, widely divergent directional outcomes were observed. For example, in figure 13, the differences among various models for hazard areas during the first 2 days of the modeling period for Khamisiyah can be seen.

Figure 13: Divergence in Models Used to Construct DOD and CIA Composite Analyses

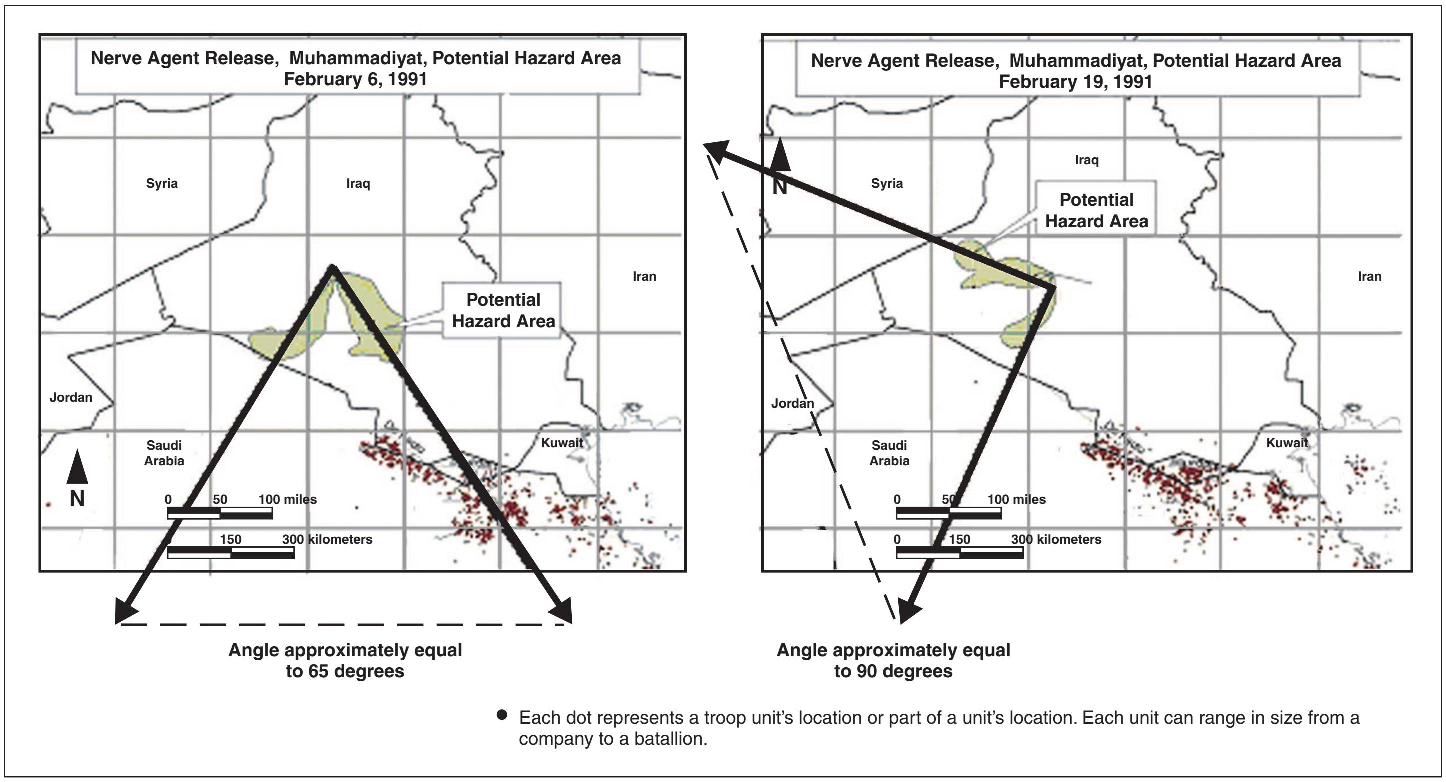


Source: Naval Surface Warfare Center.

In the March 10, 1991, section of the figure, an approximately 40 to 45 degree divergence between the HPAC/OMEGA and the HPAC/COAMPS models can be seen; in the March 11, 1991, section, an approximately 80 degree divergence can be seen.

The uncertainty attributed to this divergence is not limited to the Khamisiyah modeling. According to a modeling analyst involved with the modeling of Al Muthanna, the COAMPS and OMEGA weather models showed the plume going in different directions, at a difference of 110 to 120 degrees. According to the analyst, COAMPS showed the plume going north northwest, while OMEGA showed it going south. Similar divergence among model projections was also observed in the modeling of Muhammadiyat, as shown in figure 14.

Figure 14: Divergence in DOD Muhammadiyat Models

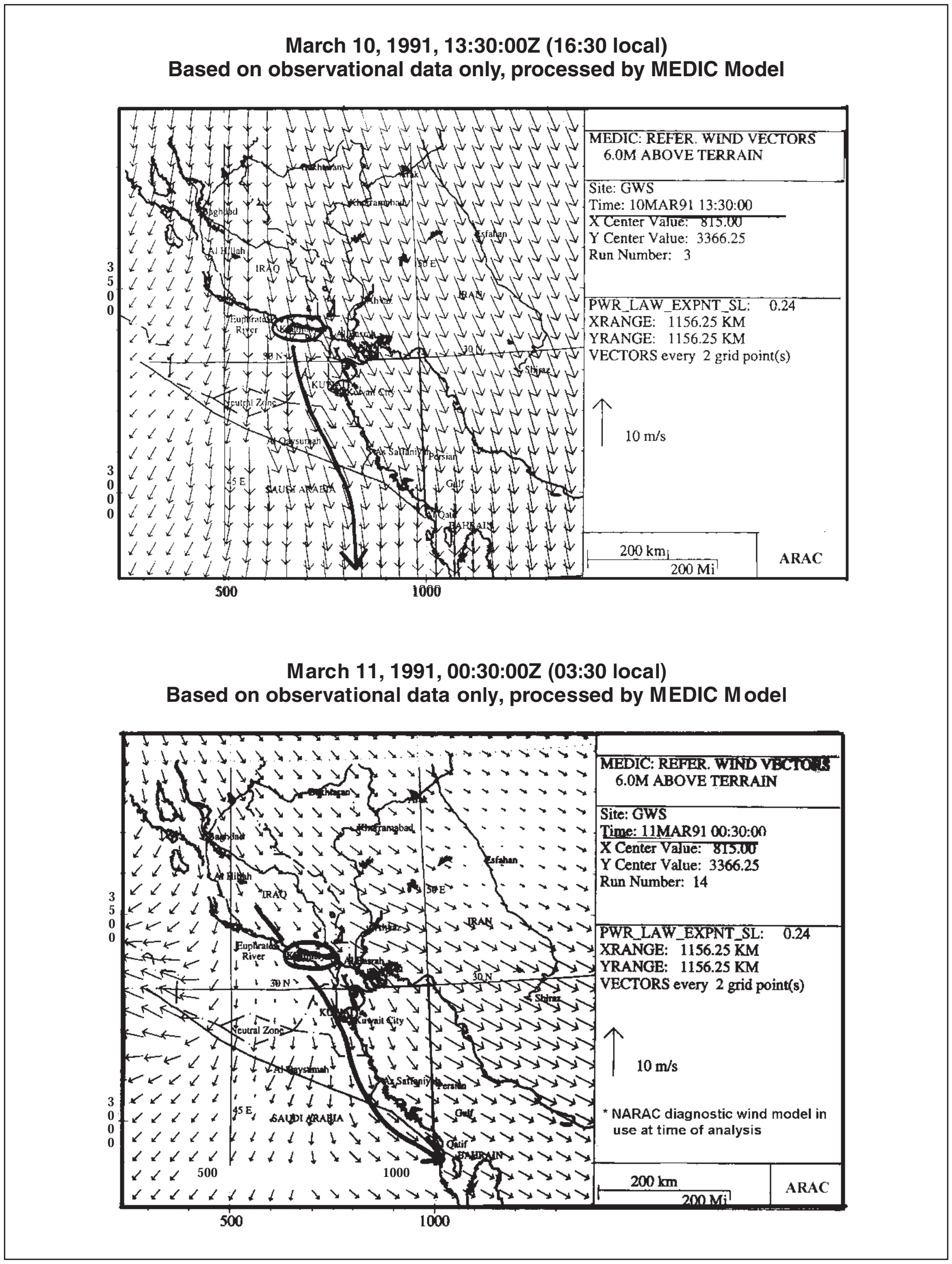


Source: Department of Defense.

Appendix IV: Divergence and Wind Field Models

In figure 15, LLNL projections for divergence of wind field vectors 6.0 m above terrain are based on observational data the Meteorological Data Interpolation Code (MEDIC) model processed.

Figure 15: Lawrence Livermore National Laboratory Diagnostic Wind Model, Based on Observational Data



Source: Lawrence Livermore National Laboratory.

Figure 16 shows the wind field vector model, based on European Centre for Medium-Range Weather Forecast (ECMWF) projections and processed by the MEDIC model.

Figure 16: Lawrence Livermore National Laboratory Diagnostic Wind Model, Based on ECMWF Projections

March 10, 1991, 12:00:00Z (15:00 local)
Based on ECMWF data, processed by MEDIC Model

MEDVEL __10MAR91__1200__00__000__0600__001.BIF

Barbs every 2

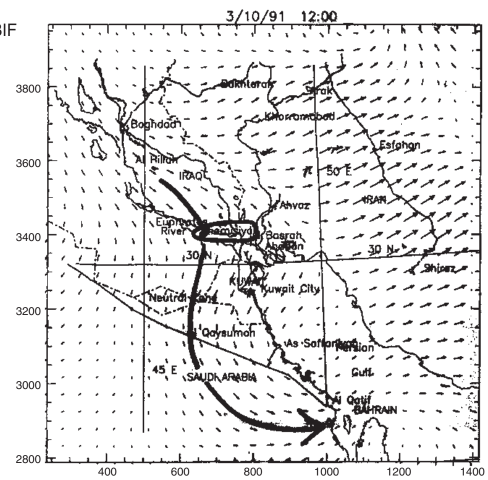
Level: 1

X Center value: 699.37

Y Center value: 3250.62

X Range: 925.00

Y Range: 925.00



Annotation line: approximate streamline

March 11, 1991, 00:00:00Z (03:00 local)
Based on ECMWF data, processed by MEDIC Model

MEDVEL __11MAR91__0000__00__000__0600__001.BIF

Barbs every 2

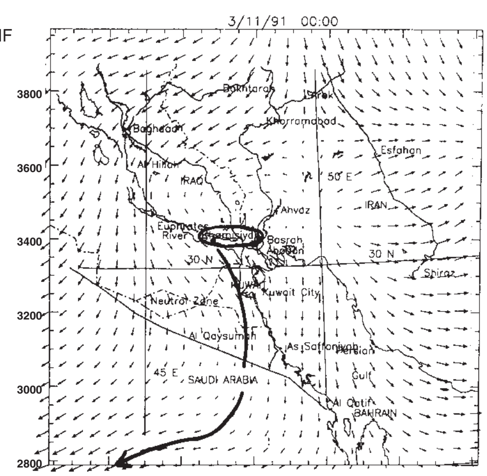
Level: 1

X Center value: 699.37

Y Center value: 3250.62

X Range: 925.00

Y Range: 925.00



Annotation line: approximate streamline

Source: Lawrence Livermore National Laboratory.

The wind field vector model in figure 17 is based on COAMPS simulations at the U.S. Naval Research Laboratories.

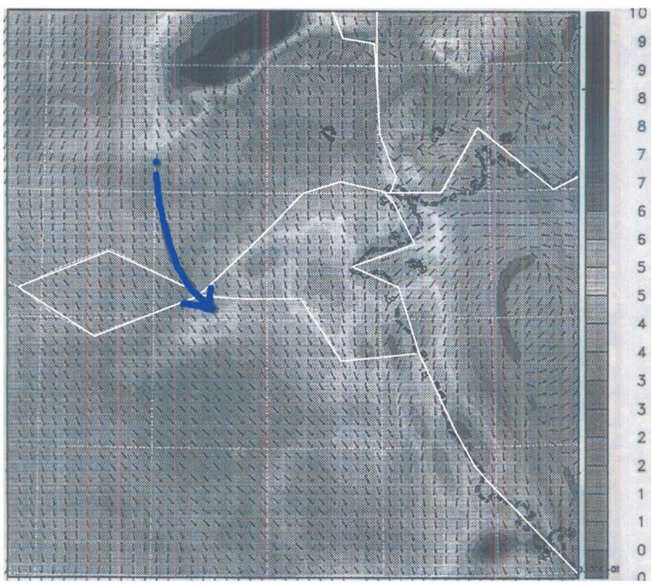
Figure 17: Wind Field Vector Model, Based on COAMPS

NRL COAMPS Simulation Wind analysis for 12z, 10 March 1991

COAMPS GRID 3, 100 X 100 X 30 5.00 km

10.0 m wind field analysis at 1991031012

surface land sea analysis at 1991031012



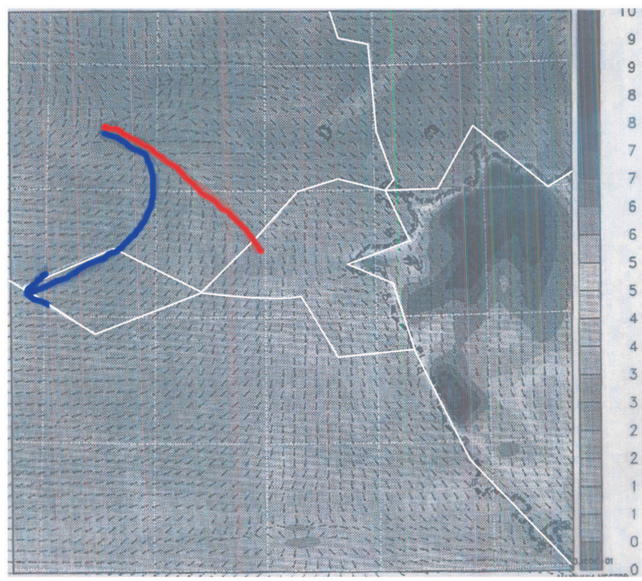
Blue Line: approximate streamline

NRL COAMPS Simulation Wind forecast for 00Z, 11 March 1991

COAMPS GRID 3, 100 X 100 X 30 5.00 km

10.0 m wind field
12 h 0 m 0 s forcast valid at 1991031100 from 1991031012

surface land sea
12 h 0 m 0 s forcast valid at 1991031100 from 1991031012



Blue Line: approximate streamline
Red line: line of diffluence (note very strong diffluence near Kamisiyah)

Source: Lawrence Livermore National Laboratory.

Appendix V: Comments from the Department of Veterans Affairs

THE SECRETARY OF VETERANS AFFAIRS
WASHINGTON

May 26, 2004

Mr. Keith Rhodes
Chief Technologist
Center for Technology and Engineering
Acquisition and Sourcing Management Team
U.S. General Accounting Office
441 G Street, NW
Washington, DC 20548

Dear Mr Rhodes:

The Department of Veterans Affairs (VA) has reviewed your draft report, ***GULF WAR ILLNESSES: DOD's Conclusions about U. S. Troops' Exposure Cannot Be Adequately Supported*** (GAO-04-159) and concurs with your recommendation to no longer use the nuclear, biological, chemical (NBC) plume model in future research studies on Gulf War veterans illnesses.

Nevertheless, it should be understood that VA has already completed three studies that incorporated the Department of Defense NBC plume model as part of the parameters for the research. In fact, VA has submitted these studies to scientific journals. Furthermore, VA is currently collaborating with other research groups that may have used the NBC plume model. These studies are underway and will be completed as planned.

The Department appreciates the opportunity to comment on your draft report.

Sincerely yours,

Anthony J. Principi

Appendix VI: Comments from the Department of Defense

Note: GAO comments supplementing those in the report text appear at the end of this appendix.

ASSISTANT TO THE SECRETARY OF DEFENSE
3050 DEFENSE PENTAGON
WASHINGTON, DC 20301-3050

NUCLEAR AND CHEMICAL AND BIOLOGICAL DEFENSE PROGRAMS

MAY 21 2004

Mr. Keith Rhodes
Chief Technologist
Center for Technology and Engineering,
Applied Research and Methods
U. S. General Accounting Office
441 G Street, N.W.
Washington, DC 20548

Dear Mr. Rhodes:

This is the Department of Defense (DoD) response to the General Accounting Office (GAO) draft report, "GULF WAR ILLNESSES: DoD's Conclusions About U.S. Troops' Exposure Cannot Be Adequately Supported," dated May 11, 2004 (GAO Code 460530/GAO-04-159). The Department non-concurs with the first recommendation and concurs with the second recommendation. A detailed analysis explaining DoD's non-concurrence and comments to the report for clarification purposes is enclosed.

We appreciate the opportunity to comment on the draft report. My point of contact for this report is Mr.Wayne Davis at (703) 697-5561, wayne.davis@osd.mil.

Sincerely,

for Debra Thedford
Klaus O. Schafer, MD, MPH
Deputy for Chemical/Biological Defense
Acting

Enclosure:
As stated

GAO DRAFT REPORT - DATED MAY 11, 2004
GAO CODE 460530/GAO-04-159

"GULF WAR ILLNESSES: DOD'S CONCLUSIONS ABOUT U.S. TROOPS' EXPOSURE CANNOT BE ADEQUATELY SUPPORTED"

DEPARTMENT OF DEFENSE COMMENTS TO THE REPORT

Meteorological and dispersion models are powerful tools to study various atmospheric phenomena such as weather forecast, global climate change, and fate of chemical and biological agents once released into the atmosphere. State-of-the-art meteorological models are highly sophisticated because they need to account for a myriad of inter-dependent atmospheric processes. Dispersion models also need to deal with inherently random, turbulent processes. Despite these challenges, meteorological and dispersion models continue to improve and make valuable contributions to many DoD missions. For example, during the 2000 Presidential Inauguration in Washington, D.C., and the 2002 Winter Olympics in Salt Lake City, a suite of meteorological and dispersion models were run continuously to provide critical information in the event of a possible terrorist attack. During Operation Iraqi Freedom, a suite of meteorological and dispersion models run continuously to support many military operations.

The outputs of DoD models, as with all models, are limited and inherently uncertain as a result of the random, turbulent, and complex processes that are modeled, as well as a result of the data that feeds the models, which may be incomplete, absent, or uncertain. It is the strength of the DoD modeling process in 2000 to study the events of the 1991 Gulf War that we use a disciplined, scientific process based on independent peer review to yield modeling results that are validated, and that limits bias in model design and defines limits of uncertainty. The Department accepts that uncertainty exists in models and source terms. Epidemiological and other studies based on DoD dispersion models recognize the inherent uncertainty. However, the Department does not accept the thesis posited in the GAO report that the existence of uncertainty is an indication of a "flawed" model nor that such models have limited or no value to support studies and assessments.

See comment 1.

RECOMMENDATION 1: The GAO recommended that the Secretary of Defense not use the plume-modeling data for future epidemiological studies, since DOD cannot know from the flawed plume modeling who was and who was not exposed.

DOD RESPONSE: We recognize that modeling the possible chemical warfare agent releases during the 1991 Gulf War was an extremely difficult task due to lack of measured source data and onsite meteorological data. Nevertheless, the use of state-of-the-art, validated computer modeling techniques is the most feasible option to determine what might have happened. Computer modeling is routinely used to address many important issues. One such example is global climate change, where obviously future source and meteorological data are unavailable, and yet important economic and environmental policies must be made based on the results of

ENCLOSURE 1

computer modeling. Another example is the collateral damage analysis before bombing missions or military operations, where again the actual source and meteorological data would be unavailable.

The Department believes that to implement this recommendation would display *negligence*, by ignoring potentially beneficial analytic tools, and demonstrate a *disregard* for the safety, health, and welfare of the members of the military Services and veterans. Additionally, the implementation of this recommendation would replace any rational basis for epidemiological assessment with an irrational basis for such studies, namely, assessments of those potentially exposed based on incomplete data without logical extrapolation to probable areas of exposure.

We take exception to the term "flawed" in reference to the models. The modeling was not flawed. While the models were not able to counter the lack of information on the source terms or eliminate the inherent uncertainty in calculating for atmospheric turbulence, the models have been extensively and independently validated. The presence of uncertainty is a condition of all models. It does not indicate a flaw.

The GAO recommendation apparently represents a blanket prohibition against plume modeling in the future, where the limitations of the 1991 Gulf War may not apply. Certainly, the narrow breadth of this audit does not support such a sweeping recommendation that extends well beyond incidents related to the 1991 Gulf War. The GAO cannot rule out that a future incident may present good source and meteorological data. The DoD may have even greater confidence in modeling and its ability to answer questions or predict the outcome or impact of events. Moreover, there is the risk of this recommendation spilling over into other areas beyond epidemiology – to other areas where modeling is used for planning purposes.

See comment 2.

RECOMMENDATION 2: The GAO recommended that the Secretary of Defense require no further plume modeling of Khamisiyah and the sites bombed during the 1991 Gulf War in order to determine troops' exposure. (p. 63/GAO Draft Report)

Now on p. 60.

DOD RESPONSE: The Department concurs with this GAO recommendation. Despite significant enhancements in meteorological modeling as well as transport and diffusion modeling since the events at Khamisiyah were studied, uncertainties in the data will remain.

General comments concerning the GAO report:

The DoD used state-of-the-art meteorological and dispersion models to study the potential impacts associated with possible releases of chemical warfare agents during the 1991 Gulf War. Determined to use the best available methodology for the analysis, yet recognizing the difficulty due to lack of on-site source, meteorological, and exposure data, the DOD sought guidance in 1996 and 1997 from an Institute for Defense Analyses (IDA) panel consisting of world renowned scientists in the field of meteorology and dispersion. This independent panel provided recommendations adopted by the DoD, including the use of the ensemble approach to address issues of uncertainty. The US National Centers for Environmental Prediction and the European Centre for Medium Range Forecast use the ensemble approach to forecast weather by considering a suite of meteorological models. A 2003 report from the National Academies' Board on Atmospheric Sciences and Climate (http://www.nap.edu/html/tracking/reportbrief.pdf)

See comment 3.

2

also suggests the use of ensemble modeling to account for uncertainty. These facts reinforce the ensemble approach adopted by the DoD as the best practice in the scientific community.

Used in the DoD/CIA modeling in 1997, the ensemble methodology was refined and improved in response to scientific peer review of the 1997 results. Following its application of improved modeling in 2000, the DoD again asked a peer review panel consisting of nationally recognized modeling experts to review the results and the panel conclusion was that "the methodologies are sound." Consequently, we believe the GAO report errs when it accuses the DoD of using "flawed" plume modeling. In summary,

See comment 4.

- This audit fails to objectively evaluate the modeling capabilities of the DoD.
- The audit contains errors of fact, confusion of facts, and conclusions without scientific evidence or support.
- The audit recounts events that occurred in 1996 and 1997, without regard for improvements made in modeling, analysis, and epidemiological research in the 1999 to 2002 time frame.
- The audit fails to document the inclusion of nationally recognized modeling experts who participated in or reviewed this work.
- The audit dismisses the conclusions of all epidemiological studies that used these data, including those done with the improved 2000 modeling that was scientifically peer-reviewed and published in a technical report in April 2002.

See comment 5.
See comment 6.
See comment 7.
See comment 8.
See comment 9.

The GAO's multi-year work on this audit shows some misunderstanding of the events of the Gulf War and little comprehension of modeling in general and the DoD modeling specifically. The report contains errors of fact, confusion of facts, and conclusions that evolve without scientific evidence or support. It is also apparent that this GAO team is stuck in 1996-7. The report reflects little analysis of the final modeling, choosing instead to point out weaknesses in 1996 efforts, which were improved by subsequent work. In addition, the report fails to address the conclusions of the 2000 DoD peer review panel, other than a single line of entry in Appendix I that concludes that the DoD's "methodologies are sound," which contradicts the GAO's position that the "DoD's modeling efforts were flawed."

See comment 10.

The analysis of the epidemiology is equally rooted in 1997, reflecting none of the work done in 2000. The GAO never mentions several completed epidemiology studies that used the results of the 2000 plume modeling. The GAO's review of the epidemiological literature is selective, incomplete, and outdated. As for the scientific studies on the effects of exposures to chemical warfare agents, the GAO ignored the *Department of Defense Low-Level Chemical Warfare Agents (CWAs) Research Master Plan*, June 2003, or the results of any of the research indicated in the plan. The GAO does not explain what it thinks would be a better method to assess exposure. It appears to favor some studies of 1991 Gulf War veterans that are mentioned on pages 52 to 61, which seems to imply that GAO thinks these studies used a superior method of exposure assessment, compared to the DoD modeling. However, some of the studies GAO mentioned did not evaluate possible chemical warfare exposure at all.

See comment 11.
Now on pp. 50–55.

3

Following are **specific comments** concerning the GAO report. (Page numbers (P) refer to the draft report and may not correspond to the page in the final report.)

Now on p. 1.
See comment 12.

P.1: "... 16 of the 21 sites that were bombed were destroyed. Many US and British troops were located near some of these sites..." **This statement is not correct.** The air campaign occurred from January 17 to February 24, 1991. US and British ground forces did not enter Iraq until February 25, 1991, well after any bombing pertinent to this study. In addition, many of the sites in question were well north of the fighting area, in parts of Iraq where US and British forces did not penetrate.

Now on p. 2.
See comment 13.

P.2, para 1: "...in October 1991, United Nations Special Commission (UNSCOM) inspectors had found evidence that US troops had destroyed munitions containing CW agents at Khamisiyah. **[Evidence found only in the Pit.]** Specifically, among the nearly 100 bunkers at Khamisiyah, remnants of 122 mm rockets were identified at Bunker 73. **This is not correct.** UNSCOM could not determine if Bunker 73 contained chemical warfare agents at this time because damaged munitions made it too dangerous to get close enough to sample or take Chemical Agent Monitor readings. However, on a return visit to the site in May 1996, UNSCOM conclusively determined that debris (e.g., burster tubes, fill plugs, and plastic inserts) in the rubble of Bunker 73 was characteristic of chemical munitions.

Now on p. 2.
See comment 14,

P.2, para 2: "In June 1996, DoD estimated that 300 to 400 US troops who participated in the demolition of Khamisiyah Bunker 73 had been exposed. **This statement is not correct.** The June 21, 1996, press conference, Mr. Bacon said, "300 to 400 troops were involved in the actual detonation of the bunkers." The DoD did not say that the demolition of Bunker 73 led to any possible exposures.

Now on p. 2.
See comment 15.

P.2, para 3: "within a 25 km radius of Khamisiyah and, therefore, had potentially been exposed." **The GAO has misinterpreted the DoD.** The Office of the Special Assistant for Gulf War Illnesses used the 25-kilometer and 50-kilometer radii to identify the area where servicemembers might have information about the demolitions at Khamisiyah. In letters to those servicemembers, the DoD surveyed them for additional information, but did not identify them as "potentially exposed" as claimed by the GAO.

Now on p. 3.
See comment 16.

P.3, Scope and Methodology, para 1: "US demolitions and Coalition bombings of Khamisiyah..." **Chemical warfare agent release at Khamisiyah was not due to Coalition bombings.**

Now on p. 6.
See comment 17.

P.6, para 1: "Troops under the path of the plume were classified as exposed, those not under the path as nonexposed." **This statement is not correct.** The Office of the Special Assistant for Gulf War Illnesses said that servicemembers who were with their units within the hazard areas might have been exposed to levels of chemical warfare agents that exceeded the general population limit, but did not classify them as "exposed" or "nonexposed."

Now on p. 5.
See comment 18.

P.5, para 3: "DoD's and VA's conclusions—that there was no association between exposure to CW agents at Khamisiyah and U.S. troops' rates of hospitalization and mortality—also cannot be adequately supported." This implies that there is an association between exposure and hospitalization and mortality rates, yet **the GAO offers no evidence to support this claim**. The GAO's conclusion is inconsistent with its first finding that plume modeling cannot be used to determine who was exposed. If modeling cannot be used to determine which populations were

4

more likely to have been exposed and which were less likely to have been exposed, then it must be assumed that all populations were exposed. If this were true, then hospitalization and mortality may be associated with exposure, yet good health (or lack of illness) would also be associated with exposure.

See comment 19.

P.6, para 3: **The DoD does not concur with this recommendation.** It would be a reckless approach, without scientific basis, and could endanger the health of our servicemembers. The data the DoD used was and is the best data available and any research that desires to use it would know the limitations of the data. We have to make operational and health decisions based on the best information available. The purpose of modeling is to estimate conditions based on incomplete, but best available data. If all data were available, no modeling would be necessary. To postpone or deny decisions until all data are available is not putting the needs of the veterans first.

Now on p. 10.
See comment 20.

P.10, last bullet: **This is not part of the modeling methodology**. Information on potentially exposed populations, etc., is not necessary to complete the modeling, though it is necessary to complete the subsequent analysis to match the results of the modeling to determine who may have been exposed.

Now on p. 11.
See comment 21.

P.11: Figure 2 is incorrect. There should be an arrow connecting from global weather models to regional weather models, and only regional weather models should be connected to transport and diffusion models. Outputs from global weather models are mainly used as initial and boundary conditions for regional weather models, and should not be directly fed into transport and dispersion models. The first paragraph on P.11 should also mention this hierarchical relationship between global and regional weather models.

Now on p. 11.
See comment 22.

P.11, para 2: **The relationship of HPAC and SCIPUFF is incorrect.** The SCIPUFF is a component of the HPAC, rather than the other way around. Also, the DoD and CIA modeling efforts did not include NUSSE and ADPIC. If the paragraph intends to include all dispersion models that have ever been used by any research groups to study the 1991 Gulf War, then this list should be much longer.

Now on p. 12.
See comment 23.

P.12: The DoD and CIA modeling efforts were never intended to provide "definitive conclusions about the size or path (that is, the direction) of the plumes." This is clearly indicated by the ensemble approach adopted by DoD and CIA. Because turbulent diffusion is inherently a random process, there can never be any "definitive" conclusions.

Now on p. 12.
See comment 24.

P.12: Table 1 is incorrect. The DoD did not use MATHEW or ADPIC in its ensemble modeling of Khamisiyah. Lawrence Livermore Labs used these models in response to other requests, but the assumptions used by Lawrence Livermore were not related to the data collected by the DoD to document the demolition of Khamisiyah.

Now on pp. 13–14.
See comment 25.

P.13-14: "DoD's Models Were Not Fully Developed for Analyzing Long-Range Environmental Hazards" **The GAO omits MM5 from its discussion here.** MM5 is the most commonly used mesoscale meteorological model used in the scientific community with hundreds of peer-reviewed journal articles published on the model. The DoD used it in both the 1997 and 2000 modeling. See http://www.mmm.ucar.edu/mm5/Publications/mm5-papers.html for a complete list of journal articles published on MM5. The GAO's claim that the models used by DoD were not fully developed for analyzing long-range environmental hazards is unfounded. First previous

5

URL reference describes the long history of development and application for MM5. The Navy also has a long history of developing and running COAMPS. Refer to http://www.nrlmry.navy.mil/projects/coamps/frame_pub.htm for a comprehensive list of publications on the model. Although OMEGA is a relatively new model, partly developed with funding support from the Defense Threat Reduction Agency, there are many publications on OMEGA that describe its evaluation. See http://vortex.atgteam.com/papers/ for a comprehensive list. In addition, the developers for the three mesoscale meteorological models all published their modeling results for Khamisiyah in peer-reviewed journals. Section 4.3.2 of *Verification and Validation report of HPAC 3.0*[1] describes the validation of HPAC for long-range diffusion applications, including ANATEX (Across North America Tracer Experiment) and ETEX (European Tracer Experiment). VLSTRACK has been validated with chemical agent releases up to 220 km[2].

Now on p. 13.
See comment 26.

P.13, para 2: "…OMEGA consistently underpredicted surface wind speeds by a factor of 2 to 3…" **OMEGA did not under predict surface wind speeds.** In fact, a diffcrcnt schcmc, donc <u>outside</u> OMEGA, extrapolated model predictions to a common comparison height. Because the three mesoscale meteorological models have different vertical grid structures, it was necessary to choose a common height for the purpose of comparing surface wind speeds. Once the extrapolation scheme was made consistent, the "under prediction" no longer existed. In short, DoD *did not* use these "factor of 2 to 3" under predicted winds for plume modeling.

Now on p. 14.
See comment 27.

P.14, para 3: "…comparison of the hazard-prediction models HPAC and VLSTRACK documented substantial differences—by factors between 5 and 1,000-between the models…" **The GAO report does not provide sufficient details on what is meant.** Are these differences in terms of concentration, dosage, or hazard area? Furthermore, it is possible that the VLSTRACK results (which represent ensemble means) were compared to the 99% probabilistic output generated by HPAC.

Now on p. 14.
See comment 28.

P.14, para 3: "The most significant errors in the coding and the potential for misuse were found in HPAC and its subcomponent models. Given these problems with the analyses conducted up to 1998, HPAC could not be considered reliable." **This statement is contradicted by numerous validation and verification studies by different research groups**; all suggest the model's satisfactory performance[3].

Now on p. 14.
See comment 29.

P.14, para 4: "models were available but not used by DOD, such as MM5." **This statement is not correct.** The DoD used MM5 in both its 1997 and 2000 modeling.

Now on p. 15.
See comment 30.

P.15: We agreed with the 1996-1997 IDA panel recommendations regarding the uncertainty in the source term. Subsequent Dugway testing was conducted to reduce uncertainty. It would be ideal to duplicate all the conditions associated with the Khamisiyah event at Dugway. However,

---

[1] Bradley, S., T. Mazzola, R. Ross, D. Srinivasa, R. Fry, and D. Bacon, "Verification and Validation of HPAC 3.0," Prepared for Defense Threat Reduction Agency, 45045 Aviation Drive, Dulles, VA 20166-7517, DSWA-TR-98-66, 2000.
[2] Bauer, T. and M. Wolski, "Initial Validation of VLSTRACK Version 1.2," NSWCDD/TR-92/647, Naval Surface Warfare Center, Dahlgren, VA, 22448, 1993.
[3] Bradley, S., T. Mazzola, R. Ross, D. Srinivasa, R. Fry, and D. Bacon, "Verification and Validation of HPAC 3.0," Prepared for Defense Threat Reduction Agency, 45045 Aviation Drive, Dulles, VA 20166-7517, DSWA-TR-98-66, 2000.

6

practical constraints prevented this from happening. For example, lack of onsite meteorological data is a *fact*, rather than an *optional deficiency* that DoD could avoid.

Now on p. 16.
See comment 31.

P. 16: The GAO report seems to suggest that the agent purity should not differ widely for Khamisiyah, Al Muthanna, Muhammadiyat, and Ukhaydir. However, there is no basis to expect that agent purity would be the same for these locations because different manufacturing dates and different fill dates were involved. In addition, testing by UNSCOM and Iraq provided empirical evidence of the difference in purity.

Now on p. 18.
See comment 32.

P.18, para 3: "…shock blast effects of the munitions or the higher altitude plumes generated from the transfer of mass associated with the shock waves." **It is incorrect to assume shock (blast) waves will lead to plume mass transfer.** As mentioned in the GAO report, the typical time scale associated with shock waves is about 10^{-6} seconds. On the other hand, the typical time scale for dispersion is a few minutes to a few hours. Therefore, there is at least eight orders of magnitude difference between the two time scales, and shock waves will not have enough time to influence dispersion. In addition, the main purpose of the National Academy of Sciences cited in the GAO report is on damage to buildings due to shock waves, rather than "mass transfer" due to shock waves.

Now on p. 18.
See comment 33.

P.19, para 1: Concerning Al Muthanna, "UNMOVIC, however, informed us that UNSCOM had not physically inspected this bunker for safety reasons relating to structural instability." **This information is misleading.** UNSCOM, predecessor of UNMOVIC, inspected Al Muthanna and clearly visited Bunker 2. Pictures of the inside of Bunker 2, provided by UNSCOM, appear in the Persian Gulf War Illnesses Task Force's final report on chemical warfare issues in the 1991 Gulf War.[4]

Now on p. 19.
See comment 34.

P.19, para 2: "… at Muhammadiyat, the munitions were targeted with multiple high-explosive bombs." **Actually, any targeting was aimed at the munition storage bunkers, but at Muhammadiyat, Iraq had moved its chemical weapons out to open areas.**

Now on p. 19.
See comment 35.

P.19, para 2: "However, the type and quantity of explosives used in the Dugway testing—and, therefore, the resulting effects—are not comparable with the type and quantity of munitions that were actually used at Muhammadiyat." **The GAO misstates the relationship of the Dugway testing to Muhammadiyat modeling.** The Dugway testing was conducted to better understand the Khamisiyah Pit event, not Muhammadiyat. Nevertheless, because the sarin releases at Muhammadiyat occurred due to leakage, the evaporation and degradation rates for the chemical warfare agent were appropriate for Muhammadiyat modeling.

Now on p.19.
See comment 36.

P.19, para 3: "The major unresolved issues concerning DOD's modeling include assumptions about (1) Khamisiyah Bunker 73 …". **The DoD did not model a release from Khamisiyah Bunker 73.**

Now on p. 20
See comment 37.

P.20, para 7: "…given DoD's conclusion that all but 2.5 percent of the agent was degraded…" **The DoD did not model a release at Bunker 73 and the CIA developed the source term, not the DoD.**

Now on p. 20.
See comment 38.

P.20, last para: "on only one occasion did the CIA and DOD express any concern about agent release." **This is not correct.** The DoD has on many occasions expressed concern about agent

[4] Persian Gulf War Illnesses Task Force, "Chemical Warfare Agent Issues During the Persian Gulf War, April 2002.

7

release. It modeled possible releases at multiple locations to determine if concerns were justified.

Now on p. 20.
See comment 39.

P.20, last para: Discussions of UNSCOM inspections of Al Muthanna seem to contradict the previous paragraph where UNMOVIC said UNSCOM didn't inspect Al Muthanna.

Now on p. 21.
See comment 40.

P.21, para 3: "Therefore, for [the Muhammadiyat] model, DOD chose a duration of 24 hours. However, if agent concentration would dissipate to a miniscule level after 24 hours, it is not clear why DOD would choose a 72-hour duration far Khamisiyah models." **The explanation is clear in the DoD reports on these events.** The DoD modeled Muhammadiyat for 24 hours for each air strike because the source duration was assumed to be one hour. Therefore, towards the end of the 24-hour simulation period, the agent cloud would have been sufficiently diluted and concentrations essentially zero. Since potential hazard areas are determined by time-integrated concentrations (or dosages), we could have simulated Muhammadiyat for 72 hours and potential hazard areas would have remained the same. This is because the additional simulation time, with nearly zero concentrations, would contribute little to time-integrated concentrations. Khamisiyah was modeled for 72 hours because the source term lasted a much longer time due to evaporation of agents from soil and wood, with about 90% of the total agent released within the first 30 hours after the release. Different simulation times were chosen for Muhammadiyat and Khamisiyah is because of the fundamental difference in source durations for the two events.

For Muhammadiyat, we were unable to pinpoint the date of the release, so we assumed all agent was released at one time, which maximized the amount of chemical warfare agent for modeling. Total release at one time led to a higher estimate of concentration of dosage at any location than would be assumed if the agent release occurred over several days.

Now on p. 23.
See comment 41.

P23, para 2 and 3: "Credible scientific evidence suggests that the reported detections of CW agents were reliable." **It is unclear to what "reported detections" this section refers.** If the discussion is about the Czech and French reports of detection, the DoD did not "assert that the various detections are not valid because the source of the agents cannot be detected." In 1996, the DoD indicated that some detections appeared valid; the CIA called two of them credible. In fact, the Czech government acknowledged only two of the reports of detection, but the French have never acknowledged any detections. Based on the lack of confirmatory tests, no obvious source for the chemical warfare agents detected, and the fact that the governments never indicated these detections occurred nor provided information about them, these report were assessed as "Indeterminate."

Now on p. 25.
See comment 42.

P.26: Figure 3 should include the time of day for each satellite image. Also, improve the image quality because the axis labels and scales are not legible.

Now on p. 27.
See comment 43.

P.28, para 2: "This period of reporting on chemical agent detection … coincides directly with the initial Coalition bombings of confirmed and suspected Iraqi CW research, production, and storage facilities January 17-24,1991." While this is an interesting observation, the GAO provides no evidence to support a chemical warfare agent release of the magnitude required to generate even low-level detections hundreds of kilometers from the event.

Now on p. 27.
See comment 44.

P.28, para 2: "…figure 4 shows pressure gradient data …" **This is not correct.** Figure 4 *does not* show pressure gradient data. Figure 4 probably shows an infrared image that mainly reflects cloud-top (not ground) temperatures.

8

Now on pp. 27 and 29.
See comment 45.

P.28, 30: The inversion argument is overly emphasized in the GAO report. As indicated in Figure 6, capping inversion is almost a constant feature in the atmosphere, as it separates the atmospheric boundary layer from the free atmosphere. The height of this elevated inversion can change during frontal passage. Therefore, the inversion argument cannot be selectively used to suggest agent detections.

Now on p. 28.
See comment 46.

P.29, Figure 4: **The figure is incorrect.** The x-axis labels should indicate "east" longitude. The figure should also include the time of day.

Now on p. 29.
See comment 47.

P.30, para 2: "The assessments by the CIA and DOD that the detections of the Czech Chemical Detection Units are not credible … are unsound…" **This is not correct.** The DoD did not say the reports were not credible. In 1996, the DoD indicated that some detections appeared valid; the CIA called two of them credible. The Czech government acknowledged only two of the reports of detection, and the French have never acknowledged any detections. Based on the lack of confirmatory tests, no obvious source for the chemical warfare agents detected, and the fact that the governments never indicated these detections occurred nor provided information about them, the DoD assessed the detections as "Indeterminate."

Now on p. 30.
See comment 48.

P.31, para 4: "Modeling experts from LLNL [Lawrence Livermore National Laboratory] who participated in only the initial modeling of the Khamisiyah pit site …" It **is unclear what the GAO means by "initial modeling."** In fact, the LLNL never participated in DoD's efforts to model the Khamisiyah demolition. The LLNL produced a capability modeling of Khamisiyah in response to the IDA panel's request but, as the GAO points out later, LLNL models were not part of the DoD ensemble.

Now on pp. 30-34.
See comment 49.

P.31-35: The GAO report suggests that the actual plume height might have been significantly higher for the cases considered by DoD and CIA. The report then cites a power law formula based on a Sandia empirical study to claim that plume height can be in excess of 400 m. The report also cites an LLNL prewar analysis that the bombing of Iraqi storage facilities for CW agents would be a surface-based plume with a height of 493 m. These arguments are flawed or cannot be adequately supported based on at least three reasons.

1) The Khamisiyah pit demolition, the most significant event investigated by DoD and CIA, did not involve aerial bombings (*i.e.*, the conditions assumed in the Sandia and LLNL studies). Instead, chemical agent-filled rockets were destroyed using explosives, and many of these rockets were cracked, rather than completely demolished by the explosions.

2) It appears that the GAO report refers to plume height as the *top boundary* of the plume (see Figure 8), rather than the *centroid (center of mass)* of the plume. All dispersion models track the plume according to its centroid height. Therefore, a plume top boundary might be as high as 400 m, but its centroid height might be roughly 200 m.

3) For some events modeled by DoD, chemical agents were not instantaneously released as a buoyant puff, as suggested by the Sandia and LLNL studies. Rather, a significant portion of the release occurred from evaporation or leakage of chemical agents, in which case no buoyancy was involved. For example, based on field and laboratory testing, DoD found that for the Khamisiyah pit demolition, less than 10% of the total source term was instantaneously released into the atmosphere, and the remaining 90% of the source term was released due to slow evaporation from wood and soil. Furthermore, since the pit demolition occurred around

9

4:15 PM local time when the atmospheric boundary layer was convective and well mixed, the initial plume would have been quickly well mixed vertically throughout the boundary layer, thus rendering the model results insensitive to the initial plume height and the argument of nocturnal low level jet irrelevant.

Now on p. 31-32.
See comment 50.

Page 32, last sentence: "…video images that showed some of the plume data, such as (1) those taken from ground level at Khamisiyah …" **This is misleading.** There was no video of the demolition of the munitions in the Pit at Khamisiyah. Even if there had been a video, it would not show "data."

Now on p. 32.
See comment 51.

P.33: Figure 6 does *not* show "winds in the surface layer in the stable boundary layer often accelerate to higher speeds, in a phenomenon referred to as the low-level or nocturnal jet." In fact, the figure does not say anything regarding wind speed.

Now on p. 34.
See comment 52.

P.35, para 1: "Empirical studies and observed events tend to refute the assumptions with which the CIA and DOD supported the alternative assumption that the plume was transported by low-level jets." **This is not correct.** The DoD never assumed low-level jets transported the plume.

Now on p. 34.
See comment 53.

P.35: The GAO testimony many times mentions the nocturnal low-level jet (LLJ) as if it is a ubiquitous phenomenon in the atmosphere, i.e., occurring everywhere and every night. However, although LLJs are not rare, they are not ubiquitous either. For example, based on the analysis of two years of wind data from 47 rawinsonde stations in the United States, Bonner[5] observes that LLJs most frequently occur in central plains, particularly in the Kansas and Oklahoma region, where 30% of all rawinsonde soundings have LLJs. In addition, Brook[6] reports LLJs on 19% of the winter nights in parts of Australia. An LLJ generally forms at nighttime overland under clear sky conditions. Favorable conditions for the wind maximum of a LLJ include such factors as sloping terrain, radiative cooling in the air, surface cooling rates, conditions at sunset, and frictional decoupling of the air aloft from the surface. **The GAO report does not present any evidence to support the assumption of the presence of LLJs** on the days in question. In fact, without on-site meteorological measurements at the time, it is doubtful that such evidence would ever exist.

Now on p. 34.
See comment 54.

P.35, para 2: "nighttime detections of low levels of CW agents, associated with turbulence-resulting from aircraft-related sonic booms and incoming missiles and artillery-mixing the upper-level and lower-level atmospheric layers." **This statement lacks any technical merit.** The GAO report seems to suggest that a plume can first be bodily transported by the LLJ hundreds of miles, and then mixed to the ground due to turbulence. This doesn't happen because transport and dispersion always happen *simultaneously*. A plume cannot travel while undiluted. Furthermore, the presence of the LLJ tends to *generate* more turbulence because of shear instability. In addition, all mesoscale meteorological models (COAMPS, MM5, and OMEGA) contain the relevant physics (e.g., pressure gradient, friction, radiative transfer, and the rotation of the earth) to simulate the evolution of an LLJ. However, the diagnostic MATHEW wind field model favored by the GAO cannot simulate the LLJ, because the model does not account for atmospheric dynamics. This is especially a deficiency when the observational data are sparse.

[5] Bonner, W.D., "Climatology of the Low Level Jet," *Monthly Weather Review*, 1968, Volume 96, p.833-850.
[6] Brook, R.R., "Koorin Nocturnal Low-level Jet," *Boundary-Layer Meteorology*, 1985, Volume 32, p.133 154.

10

Now on pp. 35–37.

P.36-38: "DoD's Field Testing Did Not Realistically Simulate Actual Conditions" **The report shows a lack of understanding of the purpose, goals, and results of the Dugway testing**. The Dugway testing was not intended to recreate the demolition at Khamisiyah, to simulate the environmental conditions at Khamisiyah (except as related to the evaporation testing) or to estimate the plume height. Testing was intended to simulate portions of the Khamisiyah pit demolition to derive release characterization. The tests were to determine the mass balance of the release, that is, the amount vaporized, cast into the atmosphere as droplets, burned, or spilled. Subsequent work developed the evaporation and degradation rates.

Since explosion is a phenomenon involving a very short time scale, meteorology usually does not play a significant role. This is the reason why, for example, the methodology recommended by the EPA Risk Management Program to estimate the distance to one pound-per-square-inch overpressure for vapor cloud explosions does not involve meteorology. In fact, the power law formula listed in Appendix II of the report for plume height due to high explosive detonations also does not involve meteorology. Nevertheless, the subsequent evaporation of liquid agents spilled unto soil and wood does depend on meteorology because of longer time scales, and the Dugway laboratory testing of agent evaporation accounted for appropriate temperatures.

See comment 55.

Moreover, the Dugway study used a reduced scale of the rocket stacks in the pit because of financial and other practical constraints. It is a well-accepted practice in all disciplines of research to study a scaled-down model to estimate the effects of a full-scale phenomenon. For example, water tanks and scaled model ships are used to design warships, and the results from limited-scale field and laboratory experiments are used to develop dispersion models that are applicable to the atmospheric boundary layer.

Now on p. 37.

P.38, para 4: "The models … showed great divergence, even with the same source term data." Because the atmosphere is turbulent in nature, even the same release under the same meteorological conditions will lead to different observed concentrations at the same sampler location. In addition, because of different physics and methods of data processing, different models almost always yield different results even with the same input data. This is the reason why model uncertainty traditionally has been an important research topic in dispersion, meteorology, and many other disciplines. The ensemble approach, used by the DoD, is a state-of-the-art methodology to account for uncertainty.

Hanna and Yang[7] report that the root-mean-square-error of predicted surface wind directions is about 50°, and is 3 to 4 m/s for wind speeds. They suggest that these uncertainties in wind speeds and directions are primarily due to random turbulent processes that cannot be simulated by the models, and to sub-grid variations in terrain and land use. Therefore it is unlikely that the errors can be reduced much further.

The ensemble approach[8] is routinely used in weather forecast to account for uncertainty due to the fact that different weather models give different forecasts, even when the same observational data are used. For example, the US National Centers for Environmental Prediction (NCEP) and

[7] Hanna, S.R. and R. Yang, "Evaluations of Mesoscale Model Predictions of Near-Surface Winds, Temperature Gradients, and Mixing Depths," *Journal of Applied Meteorology*, 40, 1095-1104, 2001.
[8] Hamill, T.M., S.L. Mullen, C. Snyder, Z. Toth, and D.P. Baumhefner, "Ensemble Forecasting in the Short to Medium Range: Report From a Workshop," *Bulletin of the American Meteorology Society,* 2000, Volume 81, p. 2653-2664.

11

the European Centre for Medium Range Forecast (ECMWF) all use the ensemble approach to forecast weather by considering a suite of meteorological models, and no attempt is made to resolve the difference among these models. Moreover, a 2003 report from the National Academies' Board on Atmospheric Sciences and Climate (http://www.nap.edu/html/tracking/reportbrief.pdf) also suggests the use of ensemble modeling to account for uncertainty. The DoD used the ensemble approach in its modeling because of the uncertainty associated with individual model simulations. There is, however, one important principle for the ensemble approach. For the sake of consistency, all model simulations included in an ensemble should be comparable (e.g., in terms of model capability and input data requirements).

See comment 56.

It is important to point out that model uncertainty should not be treated as a "deficiency" in dispersion modeling because we will always be confronted with the same challenge regardless of model sophistication and data availability. In fact, Appendix IV of the GAO report presents two sets of wind vector plots generated by LLNL-ARAC (Figures 15 and 16), where there is also a great divergence between the two.

Now on p. 38.
See comment 57.

P.39: "In addition, DOD chose not to include in the composite model the results of the LLNL model, created at the IDA expert panel's request, which showed a different and larger plume size and path than DOD's models showed." **This is not correct.** The LLNL did not model the quantities and conditions developed by the CIA and the DoD; the LLNL was never part of the ensemble approach; and the LLNL results were for the benefit of the IDA panel, which regarded the LLNL models as less capable. In fact, since that time, the LLNL has replaced its meteorological models with one the DoD used.

Now on pp. 41 and 43.

See comment 58.

P.42 and 44: Figures 10 and 11: **It is misleading to overlay the DoD composite model results with the LLNL model** because there are numerous differences in how hazard areas were calculated. For example:

- The source term is different. The source term used by DoD assumed contents (both GB and GF) of 225 rockets released over several days, whereas the source term used by LLNL assumed contents of 550 rockets released instantaneously.
- The area-defining thresholds are different. The potential hazard area calculated by DoD is based on a threshold dosage of 0.0432 mg-min/m^3 for GB, and 0.0144 mg-min/m^3 for GF, over a 24-hour exposure period. We believe the "general population limit" area calculated by LLNL is based on a threshold dosage of 0.01296 mg-min/m^3 (for GB only) over a 72-hour exposure period.
- The dosage accumulation times are different. The DoD calculations involve the estimation of potential hazard areas for *each day*, thus the 24-hour exposure period mentioned above. This is because the troop location database has a 24-hour resolution. The composite "potential hazard area" is a superposition of three potential hazard areas for 10 through 12 March 1991. On the other hand, the LLNL's "general population limit" area is based on a single total exposure period of 72 hours.
- The meteorological fields are different. The DoD calculations are based on COAMPS, MM5, and OMEGA "prognostic" meteorological models that incorporate a wide range of atmospheric physics. The LLNL calculations are based on the MATHEW "diagnostic"

12

meteorological model that mainly interpolates observational data and is not based on first principles. In a data-rich environment, the difference between diagnostic and prognostic models may not be great. However, in a data-sparse region, such as Iraq in 1991, the difference between the two types of models can be significant. Because prognostic models are based on fundamental conservation laws of mass, momentum, and energy, they can predict the evolution of weather systems such as fronts, land and sea breezes, and low-level jets. Diagnostic models, on the other hand, do not have such capability. In other words, if the observational data used by MATHEW do not already include these weather phenomena, due to either sparse data or absence of such phenomena, then the gridded wind fields generated by the model will also not include such phenomena.

Now on p. 44.
See comment 59.

P.45, para 1: "…divergence in the modeling outcomes could be explained by a directional split—a line of diffluence …" This **is not entirely correct.** Diffluence in the predicted wind field is only one of the possible reasons for different model results. Another possible explanation is a time shift of flow patterns. For example, two meteorological models might predict exactly the same evolution of flow fields, but there is a slight time shift between the two, (*e.g.*, one model predicted a frontal passage at 0200Z, but the other model predicted a frontal passage at 0300Z). Then the cloud trajectory driven by the two flow fields can differ as a result.

Now on p. 44.
See comment 60.

P.45: The inter-model bias shown in Figure 12 is consistent with the root-mean-square-errors in wind direction as suggested by Hanna and Yang,[9] which further strengthens the need for ensemble modeling to account for model uncertainty.

Now on p. 45.
See comment 61.

P.46, para 1: "the VLSTRACK and SCIPUFF/HPAC results were complicated by the use of significantly different source term inputs." **This statement is misleading.** The GAO report fails to mention that the 2000 modeling used a consistent source term for SCIPUFF and VLSTRACK.

Now on p. 45.
See comment 62.

P.46, para 2: "…a plume of plumes, rather than a plume based upon data.'" **This statement is misleading.** The "plume of plumes" appears to refer to the DoD's ensemble approach. However, each "plume" in the ensemble is based on data, so the ensemble "plume of plumes" is based on data.

Now on pp. 49–60.
See comment 63.

P.50-61: "DOD's And VA's Epidemiological Conclusions On CW Exposure And Hospitalization And Mortality Rates Cannot Be Adequately Supported." **This is not correct.** The GAO focuses on only two epidemiological studies related to Khamisiyah, one on hospitalization and one on mortality. Several major, relevant studies have been published related to Khamisiyah that GAO did not mention. Both the DoD hospitalization study and the VA mortality study have been updated, using 2000 plume modeling data.[10]

[9] Hanna, S.R. and R. Yang, "Evaluations of Mesoscale Model Predictions of Near-Surface Winds, Temperature Gradients, and Mixing Depths," *Journal of Applied Meteorology,* 2001, Volume 40, p. 1095-1104.
[10] Smith, T.C., G.C. Gray, J.C. Weir, J.M. Heller, and M.A. Ryan, "Gulf War Veterans and Iraqi Nerve Agents at Khamisiyah: Postwar Hospitalization Data Revisited," *American Journal of Epidemiology,* September 1, 2003, Volume 158(5), p. 457-467; and Kang, H.K. and T.A. Bullman, "Mortality among US Gulf War Veterans who were Potentially Exposed to Nerve Gas at Khamisiyah, Iraq," Washington, DC: Department of Veterans Affairs, May 2002.

13

First, the report only mentions the original DoD epidemiologic study based on the 1997 plume estimates[11] (1). These estimates were revised. Additionally, although more than 695,000 were deployed to the first Gulf War, 351,299 were identified as Army with complete covariate data and those were used in this initial report. Only Army was used because those determined to be possibly exposed were also Army so this restriction was used in order to compare within a more homogeneous population.

The hazard areas changed with the remodeling in 2000 and the possibly exposed population changes accordingly. The revised population was used in the second epidemiologic paper[12]. In both reports, the DoD used sophisticated techniques to allow those to be included in the follow-up period until they left active duty service. Follow-up time was nearly equal among the possibly exposed and non-exposed.

Now on pp. 51–60.

See comment 64.

P.52-61: "Some Studies Suggest an Association between CW Exposure and Gulf War Illnesses." Studies of 1991 Gulf War Veterans that used self-reported data on chemical warfare agent exposure – Some of the studies mentioned on these pages attempted to evaluate possible chemical exposure while others did not. The studies that attempted to evaluate exposure used self-reported exposures, based on questionnaires several years after the war. Some of these studies did not evaluate chemical warfare agents as a possible etiology of illnesses,[13] so it is unclear why the GAO even mentions them, so no conclusions can be drawn on whether chemical warfare agents have an association with the symptoms reported by the study subjects in those particular studies. None of these articles focused on potential chemical exposure due to the demolitions at Khamisiyah.

Recall bias is a serious concern in any questionnaire study, when asking questions about possible exposures that may have taken place several years before the study is performed. Some studies have systematically evaluated the potential problem of recall bias in questionnaire studies of Gulf War veterans, and concluded that self-reported exposure information needed to be validated.[14] For example, in one Gulf War study, the authors concluded: "A veteran may report exposures that are highly improbable given the dates of deployment and, in many instances be

[11] Gray, G.C., T.C. Smith, J.D. Knoke, J.M. Heller, "The Postwar Hospitalization Experience Among Gulf War Veterans Exposed To Chemical Munitions Destruction At Khamisiyah, Iraq," *American Journal of Epidemiology*, September 1, 1999, Volume 150, No 5.

[12] Smith, T.C., G.C. Gray, J.C. Weir, J.M. Heller, M.A. Ryan, "Gulf War Veterans and Iraqi Nerve Agents at Khamisiyah. Postwar Hospitalization Data Revisited," *American Journal of Epidemiology*, September 1, 2003, Volume 158(5), p. 457-467.

[13] B.Mackness, et al., "Low Paraoxonase in Persian Gulf War Veterans Self-Reporting Gulf War Syndrome," *Biochemical and Biophysical Research Communications*, 2000, Volume 276, p. 729-733; M. Hotopf et al., "Paraoxonase in Persian Gulf War Veterans," Journal of Occupational and Environmental Medicine, 2003, Volume 45, p. 668-75; and D.J. Meyerhoff et al., "Reduced N-Acetylaspartate in the Right Basal Ganglia of Ill Gulf War Veterans by Magnetic Resonance Spectroscopy," *Proceedings of the International Society of Magnetic Resonance Medicine*, 2001, Volume 9, p. 994.

[14] McCauley, L.A., S.K. Joos, P.S. Spencer, M. Lasarev, and T. Shuell, "Strategies to Assess Validity of Self-Reported Exposures During the Persian Gulf War," *Environmental Research*, October 1999, Volume 81(3): p. 195-205; Greenberg, N., A. Iversen, L. Hull, C. Unwin, M. Destrange, and S. Wessely, "Vaccination Records in Gulf War Veterans," *Journal of Occupational and Environmental Medicine,* March 2003, Volume 45(3), p. 219; and Wessely, S., C. Unwin, M. Hotopf, L. Hull, K. Ismail, V. Nicolaou, and A. David, "Stability of Recall of Military Hazards Over Time: Evidence from the Persian Gulf War of 1991," *British Journal of Psychiatry,* October 1, 2003, Volume 183; p. 314-322.

14

unreliable when measured at two different time points in time . . . Findings which make sole use of self-reported findings, for both outcome and exposure, must be viewed with caution or suspicion."[15]

Now on pp. 59–60.
See comment 65.

P.50, para2; P.52, para3; and P.59, para 3: "Animal Studies" Most of the animal studies outlined on pages 59-61 cannot be directly extrapolated to the possible health effects of low-level sarin exposure in 1991 Gulf War veterans. The 1991 Gulf War veterans did not report any symptoms at the time of the Khamisiyah demolitions. Many of these animal studies used very high sarin exposures that led to immediate, severe symptoms. In addition, many studies administered sarin through injection, which cannot be extrapolated to inhalation exposure, which was the potential route of exposure during the Gulf War. Some 1991Gulf War veterans have developed chronic symptoms years after the war. Animal studies would have to include long-term follow-up of health effects to be relevant to illnesses that developed years after the war. In most of these studies, the animals were sacrificed within hours to weeks of exposure, and there was no evaluation of long-term health effects. In addition, some of the studies mentioned by the GAO do not evaluate the health effects of sarin or other organophosphate nerve agents, so it is unclear why GAO even mentioned them, e.g., Abou-Donia, 1996a; Abou-Donia, 1996b; Abdel-Rahman and Abou-Donia, et al., 2002.

Now on pp. 50–51.
See comment 66.

P.50, para 2; P. 51, para 4; and P. 52, para 1: "DoD and VA Used an Insensitive Outcome Measure for Determining Hospitalization Rate" **The GAO's conclusion that hospitalization studies do not provide valid data to evaluate chronic illnesses is illogical and erroneous**. GAO does not provide any examples of chronic illnesses in 1991 Gulf War veterans that would not lead to hospitalization. Many studies have demonstrated that 1991 Gulf War veterans have been diagnosed with chronic illnesses, related to nearly all organ systems. Similarly, in many studies, Gulf War veterans have been hospitalized with chronic illnesses, related to nearly all organ systems. **The GAO did not mention several studies that evaluated various health outcomes, other than hospitalization, in relation to chemical warfare agent exposure**.

The use of hospitalizations as a measure of health effects has limitations and strengths. This analysis would be limited to morbidity severe enough to require admission to a DoD hospital for inpatient care diminishing the ability for examination of the full spectrum of health effects. However, hospitalization data are very complete for active-duty military personnel and they reflect a much more objective measure of illness than self-reported symptoms or illnesses.

Further, in a 2000 report the DoD investigated those who left the service soon after the war[16]. There were no indications that 1991 Gulf War veterans were suffering increased probability morbidity ratios for infectious diseases; neoplasms; endocrine diseases; blood diseases; skin conditions; or diseases of the nervous system; circulatory system, or musculoskeletal system. However, 1991 Gulf War veterans did experience proportionally more hospitalizations for various specific diagnoses, namely, fractures and bone and soft tissue injuries (from the DoD and

[15] McCauley, L.A., S.K. Joos, P.S. Spencer, M. Lasarev, and T. Shuell, "Strategies to Assess Validity of Self-Reported Exposures During the Persian Gulf War," *Environmental Research*, October 1999, Volume 81(3): p. 195-205.

[16] Gray, G.C., T.C. Smith, H.K. Kang, J.D. Knoke, "Are Gulf War Veterans Suffering War-Related Illnesses? Federal And Civilian Hospitalizations Examined June 1991 To December 1994," *American Journal of Epidemiology*, January 1, 2000, Volume 151, No 1.

15

California Office of statewide Health Planning and Development), various diseases of the respiratory and digestive systems (from the VA, and diverse symptom diagnoses (from the VA).

Now on pp. 49–50.
See comment 67.

P.50 to 52, and 62, GAO's Review of the Epidemiological literature**, The GAO does not define its criteria for including the studies they did or excluding many other relevant studies.** Some major, relevant studies have been completed, which GAO did not mention. Several studies have evaluated the possible health effects of the release of chemical warfare agents due to Khamisiyah, including:

- One hospitalization study, performed by the Naval Health Research Center;[17]
- One mortality study, performed by the VA Office of Public Health and Environmental Agents;[18]
- Two studies of several symptoms and medical diagnoses, performed by the Oregon Health Sciences University[19];
- Two studies about veterans who had chronic illnesses that were evaluated in the VA and DoD Gulf War registries, by the Naval Health Research Center;[20] and
- Three studies of mortality, hospitalizations, and symptoms, performed the Institute of Medicine (completed, but not published).

GAO did not mention these highly relevant studies, which have the following advantages over the studies that GAO favored:

- Large study populations that were randomly selected and, therefore, representative of the overall population of 1991 Gulf War veterans;
- Focus specifically on possible exposures due to the Khamisiyah demolitions, rather than on non-specific, unverified events;
- Exposure assessments of exposure that is methodologically superior to self-reported data;
- Various objective health outcomes, in addition to self-reported symptoms; and
- Follow-up of health status of servicemembers after discharge from active-duty.

Now on pp. 70–72.
See comment 68.

P.70-71: There is no indication of the "line of diffluence" as referenced on p. 45. "line of diffluence … Appendix IV illustrates this diffluence with three different data sets."

---

[17] Smith, T.C., G.C. Gray, J.C. Weir, J.M. Heller, and M.A. Ryan, "Gulf War Veterans and Iraqi Nerve Agents at Khamisiyah: Postwar Hospitalization Data Revisited," *American Journal of Epidemiology,* September 1, 2003, Volume 158(5), p. 457-467.

[18] Kang, H.K. and T.A. Bullman, "Mortality among US Gulf War Veterans who were Potentially Exposed to Nerve Gas at Khamisiyah, Iraq," Washington, DC: Department of Veterans Affairs, May 2002.

[19] McCauley, L.A., M. Lasarev, D. Sticker, D.G. Rischitelli, and P.S. Spencer, "Illness Experience of Gulf War Veterans Possibly Exposed to Chemical Warfare Agents," *American Journal of Preventive Medicine,* October 2002, Volume 23(3), p. 200-206; and Shapiro, S.E., M.R Lasarev, and L McCauley, "Factor Analysis of Gulf War Illness: What Does it Add to Our Understanding of Possible Health Effects of Deployment?" *American Journal of Epidemiology,* September 15, 2002, Volume 156(6), p. 578-585.

[20] Smith, T.C., B. Smith, M.A. Ryan, G.C. Gray, T.I. Hooper, J.M. Heller, N.A. Dalager, H.K. Kang, and G.D. Gackstetter, "Ten Years And 100,000 Participants Later: Occupational And Other Factors Influencing Participation In US Gulf War Health Registries," *Journal of Occupational and Environmental Medicine* August 2002, Volume 44(8), p.758-768; and Smith, T.C., D.L. Jimenez, B. Smith, G.C. Gray, T.I. Hooper, G.D. Gackstetter, J.M. Heller, N.A. Dalager, H.K. Kang, K.C. Hyams, and M.A. Ryan, "The Postwar Hospitalization Experience of Gulf War Veterans Participating in US Health Registries," *Journal of Occupational and Environmental Medicine*, April 2004, Volume 46(4), p.386-397.

16

Now on p. 72.
See comment 68.

P.72: The quality of Figure 17 should be improved. It is not legible, does not have a scale for the wind barbs, and the legend to the right of each plot is not explained. Furthermore, it appears that Figure 17 is not based on the latest (2000) COAMPS results.

17

GAO Comments

1. The use of the phrase "flawed plume modeling" in the report refers to the use of DOD models that were not fully developed for analyzing long-range dispersion of CW agents as an environmental hazard. In addition, the uncertain source term data used resulted in a flawed modeling outcome. Meteorological and dispersion modeling, as predictive and diagnostic tools, can have significant value in cases where detailed meteorological data are unavailable and in providing warning for potential environmental hazards, assuming that the necessary input data supplied to the model are accurate. DOD now asserts that it has made significant improvements in its models; however, we have not evaluated DOD's assertion, since it was beyond the scope of this study.

2. We revised the recommendation to state: "We recommend that the Secretary of Defense and the Secretary of VA not use the plume modeling data as a basis for future epidemiological studies of Gulf War Illnesses in Iraq, since DOD and VA cannot know who was or who was not exposed."

   DOD correctly states that the necessary input data (i.e., source term and meteorological data) were not available. However, the models that DOD used were not fully developed for long-range dispersion of CW agents as an environmental hazard. Consequently, the modeling results were not fully reliable for determining which troops were exposed or were not exposed.

   The report did not intend to suggest that modeling, in general, is a flawed approach for predicting the hazard potential resulting from the release of toxic materials; rather, it intends to suggest that for a retrospective event, such as events at Khamisiyah, the use of models that were not fully developed for deriving long-range environmental hazards, in conjunction with the uncertain source term data used, resulted in a flawed modeling outcome. As we mentioned before, DOD now asserts that it has made significant improvements in its models; however, we have not evaluated DOD's assertion, since this was beyond the scope of this study.

3. An ensemble approach can be a useful tool in addressing issues of uncertainty. However, the process DOD used discounted at least two simulations using models that resulted in plume footprints that either were much larger or traveled in different directions or both. To properly account for uncertainty, and the untested ability of the several DOD models to estimate long-range environmental fallouts, other more mature models such as the MATHEW/ADPIC models from LLNL should have been included in the model. In the absence of their inclusion, evidence that the plumes did not travel in a divergent direction needs to be produced.

4. Again, our conclusion was not intended to suggest that modeling, or even ensemble modeling, is a flawed approach to predicting the hazard potential resulting from the release of toxic materials. Rather, it was intended to suggest that using models that were not fully developed for deriving environmental fallout estimates in conjunction with uncertain source term data resulted in a flawed modeling outcome and unsupported results. In addition, the selection of data and sites for modeling has a profound impact on model predictions. It is this area of the use of incomplete and improbable data and uncertainty in the selection criteria for sites and times to model that has resulted in flawed modeling outcomes.

5. The objectives of this report were not to evaluate DOD's modeling capabilities in general but, rather, to (1) determine the validity of DOD and British Ministry of Defense (MOD) conclusions—based on CIA and DOD plume-modeling results—regarding U.S. and British troops' exposure to CW agents during the 1991 Gulf War; (2) determine the total costs for the CIA's and DOD's various plume-modeling efforts related to these exposures; and (3) determine DOD and VA conclusions from epidemiological studies, based on DOD's plume-modeling results, that there was no association between CW exposure at Khamisiyah and the troops' hospitalization and mortality rates. In other words, we examined how DOD's capabilities were applied to this specific case.

6. This comment is dealt with in detail in addressing specific DOD comments below.

7. Our observation regarding inaccurate, inappropriate, and incomplete source term data and assumptions would apply equally to simulations conducted throughout the period from 1996 to the present. Because flawed data were fed into those models, the fact remains that the modeling results are unsupported. As we mentioned before, DOD now asserts that it has made significant improvements in its models; however, we have not evaluated DOD's assertion, since it was beyond the scope of this study.

8. The report refers to panels of experts who reviewed DOD reports and made comments and recommendations regarding the DOD modeling efforts. In at least one case, we have documented where an expert made a recommendation regarding the potential presence of meteorological phenomena not addressed in the DOD modeling studies.

9. We reviewed all published studies as well as technical reports DOD and VA prepared. We agree with DOD that "scientifically peer-reviewed and published" is considered a high standard of validity when it implies anonymous review by scientists coordinated by the editor of a reputable scientific journal leading to publication of findings in that scientific journal. The peer review and 2000 publication to which DOD refers here had not gone through that process. The 2002 Kang and Bullman study has not been published in a peer-reviewed journal

and therefore should not have been included in a review of the scientific epidemiologic literature. This subject is discussed in further detail in our response to DOD's comments on the section "DOD's and VA's Epidemiological Conclusions on CW Exposure and Hospitalization and Mortality Rates Cannot Be Adequately Supported."

10. As mentioned in response to comment 7 above, the inaccurate, inappropriate, and incomplete source term data and assumptions that we describe in the report would apply equally to simulations conducted throughout the period from 1996 to the present. The statement in appendix I reflects DOD's chronology of modeling events, not our assessment or conclusion.

11. One of the central conclusions of our report is that DOD's plume modeling was flawed, and this conclusion applied to the 2000 plume modeling as well as to the 1997 plume modeling. As for the comment that "GAO never mentions several completed epidemiology studies that used the results of the 2000 plume modeling," these studies were not mentioned for varying reasons. Smith and others, showed that hospitalization rates for several ICD-9 diagnoses were higher in veterans categorized in the Khamisiyah 2000 plume than in those not in the plume, and the association for cardiac arrhythmias was statistically significant. However, that study suffered from the same deficiencies as the earlier study that we cited—namely, inappropriate use of hospitalization outcome measures rather than appropriate use of measures of Gulf War illness, which usually do not result in hospitalization, and use of plume modeling based on flawed data. The 2002 Kang and Bullman study has not been published in a peer-reviewed journal and therefore should not have been included in a review of the scientific epidemiologic literature. This subject is discussed in further detail in our response to DOD's comments on the section "DOD's and VA's Epidemiological Conclusions on CW Exposure and Hospitalization and Mortality Rates Cannot Be Adequately Supported."

    In response to the comment that "GAO's review of the literature is selective [and] incomplete," we do not agree with this characterization of the literature review. We reviewed all published literature. The review of the literature was focused on assessing the validity of DOD and VA conclusions from the epidemiological studies, based on DOD's plume-modeling results that there was no association with CW exposures at Khamisiyah and troops' hospitalization and mortality rates. We address specific reasons for excluding reports DOD identified in our response to DOD's comments in the section "DOD's and VA's Epidemiological Conclusions on CW Exposure and Hospitalization and Mortality Rates Cannot Be Adequately Supported."

    With respect to the comment that we ignored the Department of Defense Low-Level Chemical Warfare Agents (CWAs) Research Master Plan, June 2003, or the results of any of the research indicated in the plan, we repeat that the review of the literature was focused on assessing the validity of DOD and VA conclusions from the epidemiological studies, based on DOD's plume-modeling results that there was no association with CW exposures at Khamisiyah and troops' hospitalization and mortality rates.

With regard to the comment that "GAO does not explain what it thinks would be a better method to assess exposure," we believe that some of the methodologies we cite in this report provide valuable insight into the identification of both the effects of exposure and those who were likely to have been affected by exposure. The methodologies include the identification of biomarkers or genetic polymorphisms, animal model studies that attempt to recreate the suspected event and evaluate the appearance of a similar outcome, epidemiological studies in which the cohort classifications can be safely made (such as deployed, nondeployed, or deployed outside the period of potential exposure).

Regarding the comment that we "appear to favor some studies of 1991 Gulf War veterans that are mentioned on pages 52 to 61 [now 50 to 55], which seems to imply that GAO thinks these studies used a superior method of exposure assessment, compared to the DOD modeling," we believe that in the absence of more reliable meteorological and source term data relating to the 1991 Gulf War, the DOD plume modeling must be considered inferior to other methodologies that avoid the selection bias likely to be introduced as a result of using the DOD Khamisiyah ensemble plume modeling, for the reasons cited in this report.

Concerning the comment "some of the studies GAO mentioned did not evaluate possible chemical warfare exposure at all," we evaluated only studies that examined possible CW agent exposure, genetic polymorphisms believed to be associated with CW agent exposure, animal model studies associated with CW agent exposure, and studies that used the DOD ensemble plume models (1997, 2000).

12. The initiation of the bombing of Iraq's CW research, production, and storage sites began on January 17, 1991. The release of environmental hazards associated with the open-air destruction of these facilities would have commenced on that date rather than on the ground invasion that began on February 25, 1991. U.S .and British forces did not have to penetrate these sites to be at risk from the potential health consequences of the fallout of material released as a result of these bombings. In addressing fallout, we were referring to hazardous materials thrust into the air and potentially exposing troops to CW agents at subacute levels at significant distances downwind. "Located near" is a relative term intended only to reflect that the troops were close enough to be at risk for exposure. This is not only our observation; it was also a concern of the war planners before the onset of Operation Desert Storm, as reflected by requests for hazard assessments modeling the U.S. Air Force made to Lawrence Livermore National Laboratories.

13. We concur and the report has been clarified accordingly.

14. We concur and the report has been clarified accordingly.

15. We concur with this comment and have clarified the report accordingly.

16. We concur with this comment and have clarified the report accordingly.

17. Our statement addresses not how DOD classified the troops during the modeling process but, rather, how researchers later used these data to classify troops into exposed and not exposed study cohorts in conducting health-related studies.

18. Our determination that conclusive classification criteria are unsupported does not mean or assume that the entire population was exposed. But the classification by researchers is not supported by reliable scientific, or any other, evidence since the determination of who was and who was not exposed is based on flawed and inaccurate data and the exclusion from the ensemble modeling estimates of modeling simulations that projected larger and directionally divergent potential exposure areas. Given the uncertainty associated with determining who was and who was not exposed and with determining whether or not the demolition of the Khamisiyah pit represents a single exposure event, researchers will have to assess, independent of the modeling performed and the ensemble projections, who was and who was not exposed.

    The observation that not all those who were exposed are ill applies equally, whether the DOD models and the ensemble estimates are viewed as accurate or flawed. This phenomenon is much more likely to be attributable to the genetic susceptibility of certain individuals to be physically affected by exposure at varying exposure levels.

19. The recommendation is intended to apply only to the 1991 Gulf War, and we have amended the recommendation to clarify its intent.

20. While it is not part of modeling methodology, it was part of the process DOD used in examining this issue. The report has been clarified.

21. We concur with this comment and have clarified the report accordingly.

22. DOD is correct in characterizing SCIPUFF as a component of HPAC and we have changed the language accordingly. DOD comments regarding NUSSE and ADPIC, however, are contradicted by the record. On September 4, 1997, the CIA and DOD issued a report entitled “Modeling the Chemical Warfare Agent Release at the Khamisiyah Pit (U).” In that report, the CIA identified transport and diffusion models used in this effort to include SCIPUFF, VLSTRACK, and NUSSE4. The MATHEW/ADPIC suite of models were, based on documentation supplied to us by

officials at the National Atmospheric Release and Advisory Capability, Lawrence Livermore National Laboratory, used in simulations of the Khamisiyah pit performed at the request of IDA in late 1996 and early 1997.

23. The section DOD cites does not imply that DOD or the CIA suggested that there can be “definitive conclusions” from the modeling process. Rather, it states that “The models are neither sufficiently certain nor precise to draw definitive conclusions about the size or path (that is, the direction) of the plumes.” It was the use of this information to define study cohorts in follow-on government funded health-related research that resulted in our comment.

24. We concur with this comment and have clarified the report accordingly. However, the MATHEW/ADPIC models were used in the simulations of the Khamisiyah pit, performed at IDA’s request in late 1996 and early 1997.

25. DOD is correct that MM5 and COAMPS are commonly used mesocale meteorological models. However, when they were used with dispersion models (such as VLSTRACK), which were not fully developed for long-range environmental hazards and in conjunction with uncertain meteorological and source term data, meteorological models such as MM5 could not overcome the limitations of the dispersion models. The validation of VLSTRACK at 220 km still falls far short of the distances contemplated in this report.

26. This statement is based on an internal DOD memorandum, dated December 18, 1998, with the subject “Discounting the Results of the Omega Version 3.5 for the Khamisiyah Reanalysis and Al Muthanna Analysis.” The memorandum states that the Omega consistently underpredicts surface wind speeds by a factor of 2 to 3 from actual observation collected at the five world meteorological stations in the area.

27. This statement is based on a 1999 IDA report evaluating the variance of VLSTRACK and HPAC predictions for the dispersion of chemical and biological warfare agents. The IDA report noted that for chemical releases, the HPAC and VLSTRACK predictions of areas of hazard differed substantially and that for biological releases they differed by factors of 5 and 1,000.

28. The statement DOD quotes is based on a 1998 memorandum from DOD’s Deputy for Counterproliferation and Chemical and Biological Defense. This memo states: “VLSTRACK and HPAC generate hazard predictions that are significantly different from an operational perspective.” Correspondence to GAO from the Deputy’s Modeling and Simulation Advisor noted that “the 1998 project found significant errors in the coding of one of the models such that for analyses conducted prior to that date, I would not consider that model reliable for use.”

29. We concur with this comment and have clarified the report accordingly.

30. It is, in part, this observation that causes us to view as uncertain the plume data DOD subsequently presented. Even the most elegant and precise model will provide inaccurate results if it uses inaccurate data.

31. We reported that "Assumptions about the purity of the CW agents sarin and cyclosarin established for Khamisiyah, Al Muthanna, Muhammadiyat, and Ukhaydir differed widely. In each case, agent purity was a key factor in the CIA and DOD methodology for determining the amount of agent released. For example, for modeling purposes, 10 tons of agents with a purity of 18 percent would be represented as 1.8 tons of agent. The CIA relied on UNSCOM reporting on the amount of CW agents Iraq produced. But to establish these rates, UNSCOM relied primarily on Iraqi declarations and Iraqi production records, as well as assumptions about the extent of agent degradation." This section was included to demonstrate that varying rather than consistent methodologies of differing levels of credibility were used in deriving the estimated agent purity. It was also used to explain the methodology used by the CIA and DOD in determining the maximum amount of agent available for dispersion. The report was clarified to address DOD's comment.

32. We were not suggesting that complete transfer occurred. Rather, we were providing other science-based evidence that is contrary to the earlier DOD observation that little or no transfer occurred. The overpressures generated by high explosives have to go somewhere. In a building, the structure would be destroyed. In a bunker, if the structure were not destroyed, this overpressure release would occur through the openings in the structure.

33. The photographs DOD referred to show the view from the opening in the top of Bunker 2, as well as several aerial views. According to UNMOVIC, however, UNSCOM did not physically inspect this bunker for safety reasons relating to structural instability. This observation seems to be confirmed by the photos referred to in the report DOD cited.

34. Given the images available, while the munitions may not have been targeted directly, they certainly seem to have been hit.

35. We are aware of the design of the Dugway testing and its purpose to simulate the demolitions at the Khamisiyah pit. Contrary to the DOD's assertion, the use of Dugway testing data for evaporation and degradation at Muhammadiyat is inappropriate for either leakage or destruction by high-explosive bombs, neither of which were approximated by the conditions at Dugway or at Khamisiyah. For example, the type and quantity of explosives used in the Dugway testing and, therefore, the resulting effects were not comparable to the type and quantity of munitions that were actually used at Muhammadiyat. At Dugway Proving Ground, small explosive charges were placed on boxed rockets; at Muhammadiyat, the munitions were targeted with multiple high-explosive bombs.

36. The CIA Report on Intelligence Related to Gulf War Illnesses, dated August 2, 1996, included modeling of Bunker 73 and identified the effort as being done by the CIA in parallel with DOD's Persian Gulf Investigative Team, to determine whether U.S. troops were exposed to chemical and biological warfare agents during the Gulf War. The CIA's effort did not seek to duplicate DOD's; however, CIA analysts drew on and examined DOD information to clarify intelligence, obtain leads, and ensure a thorough and comprehensive intelligence assessment. We have clarified the report to reflect that this was a CIA modeling effort.

37. According to Iraqi declarations and UNSCOM, the stocks at Bunker 73 were part of the same lot as that discovered at the Khamisiyah pit, and the munitions in the pit were estimated to have a purity of up to 50 percent when demolitions occurred. Therefore, a 2.5 percent purity rate is not supported. We concur that the earlier modeling and source term were performed by the CIA; however, later, after the IDA panel judged the models used in the early modeling efforts to be inappropriate, no effort was made to reexamine the releases from this site.

38. Our statement was taken out of context. The entire statement reads "DOD reported that the Al Muthanna research, production, and storage facility for CW agents was repeatedly attacked. Despite its repeated bombing, however, on only one occasion did the CIA and DOD express any concern about agent release. According to DOD analysis of the destruction of Bunker 2 at Al Muthanna on February 8, 1991 . . . ." This is to suggest not that the CIA and DOD are unconcerned about this issue but only that they were only sufficiently concerned to publish the results of their modeling in connection with this singular event at Al Muthanna, despite the repeated bombings of this principal Iraqi CW agent research, production, and storage facility.

39. We do not agree with DOD. According to UNMOVIC, UNSCOM never inspected this bunker for safety reasons.

40. We explained the methodology used by DOD and have clarified the report by deleting the comparison with Khamisiyah.

41. This section refers to the detections identified in table 4, of the report, now on p. 26. The comment that no confirmatory testing was conducted is not accurate. The Czech chemical detection unit reported the detections to U.S. command officials immediately, as is reflected in both the Czech and CENTCOM NBC logs, but the responding units were unable to confirm their findings when they arrived hours after the initial detections on January 19, 1991, and were unable to confirm these reports. However, in addition to the field detections, the Czech chemical detection units conducted reagent based wet chemistry confirmation tests that supported the findings of the initial detections. The French never officially acknowledged the detections attributed to their units; however, the CENTCOM NBC logs again noted that the French reported detecting chemical nerve agents on January 19,

1991, and that the Czechs confirmed the French detections. In addition, according to an Agence France Presse report, on February 4, 1991, General Raymond Germanos, a spokesperson for the French Ministry of Defense, was attributed as having confirmed that chemical fallout, “probably neurotoxins” had been detected in small quantities, “a little bit everywhere,” from allied air attacks of Iraqi CW facilities and the depots that stored them.

42. Since this NOAA-11 AVHRR-1B imagery was being used to demonstrate meteorological activity between January 18 and January 23, 1991, and not at some specific time, the exact time each image was captured will not add to or detract from its evaluation.

43. The coincidence between these and other events involving the destruction of Iraq’s CW agent research, production, and storage infrastructure using aerial bombs and cruise missiles, and the reported airborne detections of CW agents by Czech, French, U.S., and British forces (see table 4) suggest that DOD and the CIA should have reassessed their positions regarding the potential for additional exposure events.

44. We concur and the figure has been clarified accordingly.

45. The discussion in this section of the report deals with temperature inversions. Unlike a capping inversion, which is almost a constant feature in the atmosphere, a temperature inversion is not. When a temperature inversion occurs, air pollutants can be trapped near the surface of the Earth.

46. Figure 4 has been corrected to axis labels and time of day 0008Z.

47. In the April 2002 Gulf War Illness Task Force Report, “Intelligence Update: Chemical Warfare Agent Issues during the Persian Gulf War,” the CIA assessed that “the Czech detections were unlikely to be from a chemical agent.” The Czech chemical detection units conducted wet chemistry confirmation tests supporting the initial detections. We have clarified the report to remove the reference to DOD in the relevant sentence.

48. We agree that the initial modeling referred to was that performed by LLNL for IDA and not included in the DOD ensemble model. This modeling effort actually assumed fewer rockets destroyed than were later assessed to be present at Khamisiyah and that produced a plume path considerably larger and more divergent from those selected for use in the DOD ensemble.

49. The use of the power-law formula was intended to illustrate the unreasonable heights assumed during the DOD modeling efforts. Clearly not all the agent would be released simultaneously into the atmosphere and the agent released would be distributed throughout the plume geometry, but the power-law formula also

projects the plume height at time = 2 minutes. After that time, the plume continues to grow in height. In some cases, at time = 5 minutes the plume heights nearly double. Further, the distribution used in the Khamisiyah pit is based on field and laboratory testing conducted at Dugway Proving Ground that inadequately simulated the conditions at the pit and did not simulate conditions at the other sites modeled at all. We concur that the Khamisiyah pit demolition, which occurred around 4:15 pm local time, when the atmospheric boundary layer was convective and well mixed, would have been insensitive to the argument of nocturnal low-level jet. The other sites, however, were bombed during the nighttime, and therefore this argument remains valid. The power-law formula demonstrates the relationship between the amount of explosives detonated and the resultant plume height. Whether this occurs by demolition or aerial bombing is irrelevant. The issue of top boundary versus centroid of the plume is addressed in our report (see figure 8). Regarding buoyant puff at Khamisiyah, DOD based its comment on Dugway field testing, which did not realistically simulate actual conditions at the Khamisiyah site.

50. Videos of the demolition operations at Khamisiyah have been widely released. While we concur that videos do not show data, they can certainly demonstrate that DOD data assumptions, such as plume height estimates, are inaccurate, because the plume height was higher than DOD assumptions.

51. Figure 6 shows the layers in which these activities occur. We have changed the report to clarify this reference.

52. This statement should read “Empirical studies and observed events tend to refute the assumptions with which the CIA and DOD discounted the alternative assumption that the plume was transported by low-level jets.” The empirical studies are those involving the likely plume heights reached in high-explosive explosions. The observed events are the reported detections of CW agents associated with temperature inversion activity and atmospheric turbulence. We changed the report accordingly.

53. We did not make this assumption; rather, it is a possibility we were obliged to consider when evaluating the potential for the long-range transport of CW agents. It is precisely the absence of on-site measurements leading to this additional element of uncertainty for a phenomenon that is far from rare that has resulted in our questioning DOD’s and CIA’s discounting this phenomenon, despite having been cautioned to consider the possibility by a DOD expert consultant.

54. Nothing in this section of the report suggests that either long-range advection (transport of pollution) or turbulence events occurs independent of dispersion or dilution. In fact, exposures occurring at these distances would almost by necessity be at low or subacute levels. The report also does not suggest that the low-level jets function independent of turbulence. But aircraft and artillery would

produce a directional shift in turbulence, possibly resulting in mixing to the surface. The characterization of the MATHEW wind field model as one we favored is not accurate. The MATHEW/ADPIC suite simulations are simply demonstrative of the uncertainty associated with the modeling process.

55. We understand the financial and practical limitations in conducting this sort of testing. Still, differences in experimental conditions can result in profound differences in outcome. For example, there may have been more agent dispersed immediately, leaving less to evaporate over time had the simulations been conducted under different conditions.

56. We agree with this comment. Predictive modeling is a crucial asset and should be so considered. Retrospective modeling, however, in the absence of robust data is far more easily criticized as "deficient," not necessarily because of deficiencies in the models or the approach, but because of the lack of validated input data and a selection process that is subject to limited available data, an inexact intelligence assessment process, and the potential for individual bias.

57. We do not agree with DOD's characterization. Despite not having modeled the same quantities, LLNL modeled a variety of release scenarios and used the meteorological data available for the region. This modeling effort actually assumed a similar number of rockets destroyed as were later assessed to have been destroyed at Khamisiyah by the CIA in 2002, yet it produced a plume path considerably larger and quite divergent from the models selected for use in the DOD ensemble. We do not understand why IDA characterized the model as less capable. The LLNL models had an established history of modeling the release of hazardous materials, including the 1991 Kuwaiti oil fires.

    In February 1991, during the last few days of the Gulf War, the Iraqis ignited about 605 oil wells, causing an unprecedented environmental disaster in the region. During spring and summer 1991, two working groups, one sponsored by the World Meteorological Organization in conjunction with the World Health Organization and the other consisting of the U.S. government's scientific community, conducted airborne sampling programs to evaluate the local and global consequences of these fires. The Atmospheric Release Advisory Capability (ARAC), incorporating the MATHEW/ADPIC modeling suite, provided daily forecasts of the location and density of the smoke plumes in support of these aircraft missions, and concurrently to all the countries affected.

The modeling was performed in the same region and during the same general time as the Khamisiyah event and the 1986 nuclear reactor accident at Chernobyl.[1] In conjunction with the Chernobyl event, LLNL's long-range particle-in-cell model accurately simulated the spread of the radioactive cloud over the entire northern hemisphere, as verified later by radiological measurements. The NARAC emergency response central modeling system LLNL currently uses consists of a coupled suite of meteorological and dispersion models. The data assimilation model ADAPT constructs fields of such variables as the mean winds, pressure, precipitation, temperature, and turbulence, using a variety of interpolation methods and atmospheric parameterizations. Nondivergent wind fields are produced by an adjustment procedure based on the variational principle and a finite-element discretization. The dispersion model LODI solves the 3D advection-diffusion equation using a Lagrangian stochastic; LODI includes methods for simulating the processes of mean wind advection, turbulent diffusion, radioactive decay and production, bioagent degradation, first-order chemical reactions, wet deposition, gravitational settling, dry deposition, and buoyant/momentum plume rise. The models are coupled to NARAC databases providing topography, geographical data, chemical-biological-nuclear agent properties and health risk levels, real-time meteorological observational data, and global and mesoscale forecast model predictions. The NARAC modeling system also includes an in-house version of the Naval Research Laboratory's mesoscale weather forecast model COAMPS. This is a mesoscale meteorological model that LLNL has incorporated into its modeling suite.

58. Regarding the comment, "the source term is different," LLNL modeled a variety of release scenarios and used the meteorological data available for the region. This modeling effort actually assumed a similar number of rockets destroyed as were later assessed to have been destroyed at Khamisiyah by DOD in 2002, yet it produced a plume path considerably larger and quite divergent from the models selected for use in the DOD ensemble. According to DOD's technical report, "Modeling and Risk Characterization of U.S. Demolition Operations at the Khamisiyah Pit" (April 16, 2002), released by William Winkenwerder, Jr., Assistant Secretary of Defense (Health Affairs) and Special Assistant to the Under Secretary of Defense (Personnel and Readiness) for Gulf War Illnesses, Medical Readiness, and Military Deployments, the input source parameters used in the 1997 Khamisiyah pit modeling included a best estimate of 500 rockets damaged in

[1]F. W. Whicker and others, "PATHWAY: A Dynamic Food-Chain Model to Predict Radionuclide Ingestion after Fallout Deposition," *Health Physics* 52 (1987): 717–37; L. R. Anspaugh and others, "The Global Impact of the Chernobyl Reactor Accident," *Science* 242 (1988): 1513–19; T. Straume and others, "The Feasibility of Using I129 to Reconstruct I131 Deposition from the Chernobyl Reactor Accident," *Health Physics* 71:5 (1996): 733–40; K. T. Bogen and others, *Uncertainty and Variability in Updated Estimates of Potential Dose and Risk at a U.S. Nuclear Test Site Bikini Atoll*, UCRL-JC-122616 (Livermore, Calif.: Lawrence Livermore National Laboratory, 1995).

demolition, based on UNSCOM reporting and intelligence information. According to the report, this number could be as high as 650 or as low as 170, based on number of rockets minus rockets found by UNSCOM that were undamaged. The same report estimated that the total number of rockets in the pit was 1,250. Again, according to the report, this number could be as high as 1,400 or as low as 1,100, based on size of crates and stacks of rockets.

Regarding the comment "area-defining thresholds are different," we concur. However, the legend in figure 10 acknowledges this difference. Regarding the comment "dosage accumulation times are different," we concur. Again, however, this issue is addressed in the footnote defining these factors. Regarding the comment "meteorological fields are different," again, we concur with this observation. However, this comment is self-evident in that the plume is moving in a different direction. It is important to note that the corrections above do not have any effect on our conclusions. The use of the composite image is intended to illustrate that the use of different models, based on different underlying principles and assumptions, can result in different outcomes. In that regard, it is not misleading.

The fact that Iraq is a data-sparse region only serves to strengthen our observation that the uncertainty associated with attempting to model the fallout from the Khamisiyah pit, and to an equal or greater degree the other sites modeled, is too great to provide meaningful results. If an ensemble approach is to be attempted, then a range of methodologies needs to be incorporated.

59. That our language "could be explained" suggests that it is only one possible reason. When we interviewed researchers at LLNL, they noted that this diffluence may account for the different outcomes.

60. Figure 12 not only illustrates intermodel bias; it also illustrates model, and potential ensemble, bias. In the case illustrated, an ensemble of the three models would still not incorporate the area containing the actual hazard. We concur that ensemble modeling is essential in minimizing uncertainty and providing hazard warning. But as illustrated by figure 12, even in using an ensemble approach, significant uncertainty remains.

61. Our report has been clarified to reflect the use of consistent source term for SCIPUFF and VLSTRACK in the 2000 modeling. This reconciliation in source term does not, however, change any of our other observations or recommendations regarding the uncertainty associated with the source term, including the observation that, even with harmonized source, model projections still differ.

62. We do not agree that this is a misleading statement. While each of DOD's modeled plumes was based on data, the composite or ensemble plume was based on a

simple graphic overlay of the projection of the three component plumes. The relationship between these plumes, therefore, was not based on data.

63. One of the central conclusions of our report is that DOD's plume modeling was flawed, and this conclusion applied to the 2000 plume modeling as well as to the 1997 plume modeling, because both suffered from the same weaknesses. That is, the models were not fully developed for long-range environmental hazards, and source term and input data were incomplete. Regarding the comment that "Several relevant studies have been published . . . that GAO did not mention," we have now incorporated these studies in our report.

64. As we explained in our report, the PON group of enzymes is a potentially important predisposing factor in Gulf War illnesses because one of its major functions in normal body physiology is to protect the nervous system from organophosphate chemical toxins, such as pesticides and nerve agents. This finding was remarkable because the only function of Q type of the PON enzyme group is to protect the nervous system from nerve agents sarin, soman, tabun, and VX. The R-type isoenzyme has as its main function protection from organophosphate pesticides, such as diazinon, malathion, and parathion. Thus, an association between Gulf War illnesses and blood levels of only the Q-type isoenzyme of PON points specifically to nerve agent exposure. Therefore, they were being cited to illustrate studies that have examined paraoxonase deficiencies in veterans reporting Gulf War Syndrome.

    DOD's second paragraph is inaccurate. We did review the three studies cited in DOD, footnote 14. First, we rejected the Greenberg and others reference (second citation in DOD footnote 14). It is a letter to the editor merely commenting on possible errors in the recall of vaccine receipt in British Gulf War veterans; it did not deal with self-reports of CW agents and was not peer reviewed.

    The two other studies DOD cited, McCauley and others (the first citation in DOD's footnote 14) and Wessely and others (the third citation in DOD's footnote 14) both reported studies that measured the level of agreement on self-reported endorsement of various wartime environmental exposures in ill and well Gulf War veteran populations. These studies do not demonstrate recall bias. In fact, they contradict that claim. DOD has overlooked that recall bias results from nondifferential misclassification of exposure measurements in the case and control groups. Simply finding that some veterans misreported their exposures does not establish recall bias; other criteria are required to establish that. Both studies showed moderate to good levels of test-retest agreement (kappa >0.4) on self-reports of CW agent exposure, but, more importantly, the level of test-retest agreement (kappa statistic) did not differ between the ill and well groups. This means that whatever errors in recall occurred were nondifferential (occurred at the same rate in both groups), and therefore they did not bias the estimates of the relative risks for CW agent exposure. This means also that the errors in recall did not result in recall bias.

65. In suggesting that "Most" of the animal studies "cannot be directly extrapolated to the possible health effects of low-level sarin exposure in 1991 Gulf War veterans," DOD has missed the important developments in recent studies cited in our report that can be directly extrapolated to the health effects of low-level sarin exposure in the 1991 Gulf War. The best examples are the DOD-funded experiments by Henderson and others from the Lovelace Respiratory Research Institute and from the U.S. Army Medical Institute of Chemical Defense (*Toxicol Appl Toxicol* 184 (2002): 67–76). This study modeled the low-level inhalation exposure in rodents to sarin, with and without heat stress. No immediate health effects were measured by clinical indicators and brain pathological examination. However, 30 days after cessation of exposure, the rodents were found to have developed physical evidence of brain cell damage, demonstrated using sophisticated microscopic, neurochemical brain examination techniques. Once structural changes are found, the capability of sarin to cause chronic brain illness is established, and following the animals in the longer term is unnecessary. Other studies we cited add depth to the significant findings of Henderson and others and establish a body of evidence demonstrating the potential of subacute sarin exposures to cause brain cell damage and related chronic symptoms.

66. DOD's response fails to distinguish between veterans having only symptoms relating to different organ systems and veterans having classic diseases of those organ systems to which physicians assign ICD-9 codes and for which they often hospitalize patients. The public health problem of 1991 Gulf War veterans is characterized by a collection of symptoms of various organ systems in which physicians typically do not recognize classic, diagnosable diseases for which ICD-9 codes exist. And since veterans with this problem are generally not critically ill enough to require hospitalization, their physicians do not hospitalize them more commonly than they do veterans without the condition. This is why DOD and VA studies with hospitalization as the outcome have not been productive.

    The Gulf War illness manifested by serious symptoms but no physical signs is the example of chronic illness in 1991 Gulf War veterans that would not lead to hospitalization. Recent studies by several investigators at different institutions have shown that this set of conditions stems from physical damage to brain cells in deep parts of the brain.

    By stressing that DOD "hospitalization data are very complete for active-duty personnel," DOD overlooks the significant selection bias that results from studies that rely on DOD hospitalization data alone as an outcome measure. Since most of the more severely ill Gulf War veterans left the military soon after the 1991 Gulf War, they were no longer eligible for hospitalization in DOD hospitals and, thus, their further hospitalizations were no longer counted. This "attrition" of the most ill veterans creates a strong selection bias in these studies toward falsely negative findings.

Gray and others (DOD's footnote 16) fails to address the observation that Gulf War illness does not satisfy the diagnostic criteria for ICD-9 codes for the illness categories Gray and others studied—for example, infectious diseases, neoplasms, endocrine diseases, and blood diseases. Gray's finding that these classic diagnoses, defined by ICD-9 codes, were no more common in those who left the military soon after the war does not address the issue of selection bias from using DOD hospital data. To the contrary, other data Gray and others provided in their 1996 paper, reviewed by Haley (*Am J Epidemiol* 148 (1998): 325–23), demonstrated conclusively that military personnel discharged soon after the war did have largely different reasons for being discharged from the service, which points to a selection bias.

The references to the studies involving the California Office of Statewide Health Planning and Development and VA are not relevant in assessing the health effects of CW agent exposure, because these studies did not relate illness to CW agent exposure measures.

67. We did review all the studies cited but found them unsuitable for consideration for the following reasons. Smith and others (DOD's footnote 17) used inappropriate outcome measures (ICD-9 diagnoses made in hospitalized patients) and unsupported measures of CW agent exposure—that is, the 2000 Khamisiyah plume model addressed in the report. Consequently, no useful conclusions can be derived from it. Kang and Bullman (DOD's footnote 18) is an internal VA technical report that has not been peer reviewed or published in a scientific journal and therefore is not appropriate for inclusion in a review of the scientific epidemiologic literature. McCauley and others (the first citation in DOD's footnote 19) found that 1991 Gulf War veterans who witnessed the Khamisiyah demolition (a more valid measure of Khamisiyah exposure than the flawed plume models) had more chronic symptoms on average than those who did not witness it. This study suggests a causal role of Khamisiyah-associated sarin exposure in chronic Gulf War illness. We did not include the paper by Shapiro, Lasarev, and McCauley (second citation in DOD's footnote 19) because an anomalous problem with its factor analysis method appeared to disqualify its findings on a methodological basis. The investigators performed factor analysis of symptoms in a random sample of 1991 Gulf War veterans and found that those who actually witnessed the Khamisiyah demolition were significantly more likely to have their syndrome factor 2 ("dysesthesia syndrome"). While this appears to support a causal role of Khamisiyah-associated sarin exposure in chronic Gulf War illness, the authors also reported that their factor analysis method was unreliable when applied to randomly generated variables. If this study were included, however, it would suggest a causal link with CW agent exposure. As for the two studies by Smith and

others (DOD's footnote 20), we did not include these because both used participation in the DOD and VA registries as a proxy case-definition for Gulf War illness. Since military personnel and veterans were free to participate in the registry, regardless of whether or not they were ill or, if ill, the nature of their illness, registry participation is an entirely nonspecific measure and not suitable for scientific research on the problem. The two studies showed that registry participation and hospitalization were more common in veterans who were present at the Khamisiyah risk area than those who were not. Although the findings would support a causal link of CW agent exposure and illness, their methods do merit inclusion in a scientific literature review.

68. We believe that the images in figure 17 are adequate to demonstrate the modeled diffluence in wind field data. While we cannot improve the quality of the figure, we have added an arrow to show the diffluence.

Appendix VII: Comments from the Central Intelligence Agency

25 May 2004

MEMORANDUM FOR: Mr. Jason Fong, General Accounting Office (GAO)

SUBJECT: Central Intelligence Agency (CIA) Interim Response to GAO Draft Report "*DOD's Conclusions about U.S. Troops' Exposure Cannot Be Adequately Supported*" (GAO-04-159)

1. We appreciate your providing a copy of the draft report, *Gulf War Illnesses: DOD Conclusions about U.S. Troops' Exposure Cannot Be Adequately Supported* (GAO-04-159), for CIA's review and comments. We regret, however, that we cannot complete our review in the ten days allotted and therefore must nonconcur with the report pending completion of a comprehensive review, which we estimate will require three to four weeks.

2. The CIA's Biological and Chemical Group conducted a preliminary review of the report and identified numerous statements that they consider inaccurate, such as "source term data quantity and purity of the agent were inaccurate because they were uncertain, incomplete, and nonvalidated." They contend that the agent source term is complete and accurate to a known certainty. The agent source term is based on testing by an Iraqi quality control laboratory as well as later by UNSCOM-sponsored testing in world-class laboratories.

3. The CIA will provide a formal response outlining all of our concerns in the coming weeks.

Sincerely,

Hannah P. Marter
Liaison Group
Office of Congressional Affairs
Central Intelligence Agency

c.c. Mr. John Davies
Chief, Iraq Team
Biological and Chemical Group

GAO's Mission

The General Accounting Office, the audit, evaluation and investigative arm of Congress, exists to support Congress in meeting its constitutional responsibilities and to help improve the performance and accountability of the federal government for the American people. GAO examines the use of public funds; evaluates federal programs and policies; and provides analyses, recommendations, and other assistance to help Congress make informed oversight, policy, and funding decisions. GAO's commitment to good government is reflected in its core values of accountability, integrity, and reliability.

Obtaining Copies of GAO Reports and Testimony

The fastest and easiest way to obtain copies of GAO documents at no cost is through the Internet. GAO's Web site (www.gao.gov) contains abstracts and full-text files of current reports and testimony and an expanding archive of older products. The Web site features a search engine to help you locate documents using key words and phrases. You can print these documents in their entirety, including charts and other graphics.

Each day, GAO issues a list of newly released reports, testimony, and correspondence. GAO posts this list, known as "Today's Reports," on its Web site daily. The list contains links to the full-text document files. To have GAO e-mail this list to you every afternoon, go to www.gao.gov and select "Subscribe to e-mail alerts" under the "Order GAO Products" heading.

Order by Mail or Phone

The first copy of each printed report is free. Additional copies are $2 each. A check or money order should be made out to the Superintendent of Documents. GAO also accepts VISA and Mastercard. Orders for 100 or more copies mailed to a single address are discounted 25 percent. Orders should be sent to:

U.S. General Accounting Office
441 G Street NW, Room LM
Washington, D.C. 20548

To order by Phone:	Voice:	(202) 512-6000
	TDD:	(202) 512-2537
	Fax:	(202) 512-6061

To Report Fraud, Waste, and Abuse in Federal Programs

Contact:

Web site: www.gao.gov/fraudnet/fraudnet.htm
E-mail: fraudnet@gao.gov
Automated answering system: (800) 424-5454 or (202) 512-7470

Public Affairs

Jeff Nelligan, Managing Director, NelliganJ@gao.gov (202) 512-4800
U.S. General Accounting Office, 441 G Street NW, Room 7149
Washington, D.C. 20548