Results of comparisons of the predictions of 17 dense gas dispersion models with observations from the Jack Rabbit II chlorine field experiment

T. Mazzola[1], S. Hanna[2], J. Chang[3], S. Bradley[1], R. Meris[4], S. Simpson[4], S. Miner[4], S. Gant[5], J. Weil[6], M. Harper[7], J. Nikmo[8], J. Kukkonen[8], J.-M. Lacome[9], M. Nibart[10], O. Björnham[11], S. Khajehnajafi[12], K. Habib[13], P. Armand[14], T. Bauer[15], L. Fabbri[16], T. Spicer[17], N. Ek[18]

[1]SAIC, Lorton, VA USA <Thomas.A.Mazzola@saic.com>, Scott.Bradley@saic.com
[2]Hanna Consultants, Kennebunkport ME USA <stevenrogershanna@gmail.com>
[3]RAND Corporation, Arlington, VA USA <jchang@rand.org>
[4]DTRA, Ft. Belvoir, VA USA <ronald.g.meris.civ@mail.mil>, <steven.m.simpson27.ctr@mail.mil>, <sean.j.miner.ctr@mail.mil>
[5]HSE, Buxton, Derbyshire, UK <Simon.Gant@hse.gov.uk>
[6]NCAR, Boulder, CO USA <weil@ucar.edu>
[7]DNVGL Digital Solutions, Stockport, UK <Michael.Harper@dnvgl.com>
[8]FMI, Helsinki, Finland <Juha.Nikmo@fmi.fi>, <Jaakko.Kukkonen@fmi.fi>
[9]INERIS, Verneuil-en-Halatte, France <jean-marc.lacome@ineris.fr>
[10]ARIA, Boulogne-Billancourt, France <mnibart@aria.fr>
[11]FOI, Umea, Sweden <oscar.bjornham@foi.se>
[12]SAFER Systems, Westlake Village, CA USA <shah@safersystem.com>
[13]BAM, Berlin, Germany <karim.habib@bam.de>
[14]CEA, DAM, DIF, F-91297 Arpajon, France <patrick.armand@cea.fr>
[15]NSWC, Dahlgren, VA USA <timothy.bauer@navy.mil>
[16] EC JRC, Ispra, Italy <Luciano.FABBRI@ec.europa.eu>
[17]Un. Arkansas, Fayetteville, AR USA <tos@uark.edu>
[18]Environment and Climate Change Canada, Dorval, Canada nils.ek@canada.ca

Abstract

The Jack Rabbit II (JR II) chlorine field trials in 2015 and 2016 involved nine 5 to 20 ton releases of pressurized liquefied chlorine from a tank mounted 1 m above a broad flat desert sand surface. A model comparison study was initiated, where 17 widely-used

dense-gas dispersion models were run by scientists in seven countries. Predictions were submitted following specified formats, using specified emissions and meteorology inputs. To compare with the model predictions, sets of observations were defined for the arc-maximum 1-3 s averaged concentrations (arc max C) and for cloud widths and heights (to 20 ppm and 200 ppm contours) at distances from 0.2 to 11.0 km from the release. The initial focus is on the three field trials (1, 6, and 7) that have the highest observed concentrations and that have detailed emissions information.

It is found that these models are able to satisfactorily simulate (generally within a factor of two) the observed arc max C's and their variation with downwind distance at this flat desert site. At each downwind distance, the scatter in the arc max C predictions covers about 1 ½ orders of magnitude, but the observed arc max C is within the range of the predictions. The median of the cloud width predictions is about 50 % larger than the observed value for the three trials. The median of the cloud height predictions is within about 10% of the observed value. For both cloud width and/or height, there are a few models with large (factor of 3 or higher) overpredictions.

Of the 17 models, when compared to observations, there is a core group of 5 or 6 with consistently (across all three trials and all distances) less mean error and scatter in their predictions of arc max C and cloud width and height. However, as a group, the 17 models are performing adequately (using the "factor of two" rule of thumb). An important caveat is that, at the JR II desert site, chlorine deposition is minimal. At a site with vegetation and/or organic-rich soil, the effects of removal of chlorine by deposition are expected to be significant.

Keywords: Jack Rabbit II chlorine field experiment, dense gas dispersion, model evaluation

1 Introduction

The Jack Rabbit I (in 2010) and II (in 2015 and 2016) field experiments, at a flat desert site within Dugway Proving Ground (DPG) in Utah, were initiated because there had been much concern about the possible effects of pressurized liquid chlorine and anhydrous ammonia released from storage tanks and transportation vessels (e.g., see Hanna et al., 2008, 2012 and 2016). A prime contributor to the concern was the finding that, for the Macdona and Graniteville chlorine railcar accidents, the available models predicted concentrations that, when combined with exposure and health effects models, were suggesting that there would be large numbers of casualties. Yet, at both accident sites, there were far fewer casualties than expected by the model systems. This discrepancy could be caused by several factors, such as overestimates of emissions to the atmosphere, biased dispersion models, limited knowledge of deposition to the surface, and limitations to the health effects modules. JR II addresses the first and second factors – the emissions and dispersion models. In the current paper, though, we make use of the emissions model testing and improvements resulting from detailed observations of the JR II tank and releases (Spicer and Tickle 2020) and provide all of the modelers with the optimized emissions estimates. We had hoped to also test the deposition modules at JR II, but discovered that it was not possible due to the chlorine-containing salt that was naturally present in the desert surface at the test site. In fact, the chlorine deposition is minimal for these releases, thus the JR II field tests are best for evaluating the dispersion models.

Observations from three of the JR II trials (1, 6, and 7) are used in the current paper to compare with the predictions of 17 dense gas model runs, from the U.S., Canada, five European countries, and the European Commission (see Table 1). The models are briefly described in the next section. To avoid confusion, note that two of the model runs were for the same model (SLAB) run by different persons at different groups (SLAB-I by INERIS in France and SLAB-R by RAND

Corp in the US). An overview of the JR II field experiment and a description of the model evaluation exercise are given by Fox et al. (2020) as an introduction to the special JR II issue of Atmospheric Environment.

Many of the modelers who are listed in Table 1 participated for three or four years in the JR II Modeling Working Group, which was advising the project managers. These modelers ran test cases to assist in guiding the placement of samplers in the field test experiments. In contrast, other modelers in Table 1 had seen only limited subsets of the JR II data prior to this exercise. All modelers were sent a set of recommended inputs for emissions (Spicer and Miller, 2017) and for meteorology (Hanna, 2018) and were asked to send model outputs of the arc-maximum concentrations (arc-max C) for averaging times of 1 s, and cloud widths and heights (to the 20 ppm and 200 ppm concentration contours) at distances of 0.2 0.5, 1, 2, 5, and 11.0 km from the release. This is summarized in the Jack Rabbit II Model Inter-Comparison Specification (Mazzola, et al. 2019). More details on the recommended emissions and meteorology are given by Spicer and Tickle (2020) and Hanna (2020), respectively, in this special issue. Further discussions on the concentration observations are given by Chang and Mazzola (2020) and on the cloud width and height observations by Mazzola (2020).

We briefly summarize the model inputs in Table 2 (meteorology) and in Table 3 (emissions). Hanna (2020) and Spicer and Tickle (2020) have full details supporting these tables. In addition, for models which can treat these, we specified a surface roughness of 0.5 mm, and a chlorine deposition velocity assumption of 0.04 cm/sec, for the Dugway desert playa.

An over-arching goal for this model inter-comparison exercise is to facilitate assessment and improvement of all models used for analysis of hazardous gas

releases. It is intended to generate collaborative studies involving experts in the field, and the test data from JR II have been widely distributed and hopefully will generate additional research. With continuing collaboration, it is hoped that scientific gaps related to modeling of hazardous gas releases will be identified, leading to future research, and that all models can become more accurate.

Table 1. List of models being compared and organizations running the models.

Model(s) run	Organization
Accident Damage Analysis Module (ADAM)	European Commission Joint Research Centre (JRC), Italy
ALOHA, SLAB-R	Rand, USA
Britter & McQuaid workbook (B&M)	Hanna Consultants, USA
Canadian Urban Dispersion Model (CUDM)	Environment and Climate Change, Canada
DRIFT	Health & Safety Executive (HSE), UK
ESCAPE	Finnish Meteorological Institute (FMI)
HPAC	Defense Threat Reduction Agency (DTRA), USA
Integral Dense-gas Dispersion Model (IDDM)	National Center for Atmospheric Research (NCAR), USA
PHAST	DNV GL, Ltd, UK
Parallel-Micro-SWIFT-SPRAY (PMSS)	Aria Technologies and Atomic and alternative Energies Commission (CEA), France
Puff model of atmospheric dispersion (PUMA)	Swedish Defence Research Agency (FOI)
RAILCAR-ALOHA, RAILCAR-QUIC	Naval Surface Warfare Center, USA
Safer Trace	Safer Systems, USA
SLAB-I	Ineris, France
VDI 3783 Parts I & II	BAM, Germany

Table 2. Basic recommended meteorological inputs for trials 1, 6 and 7 (mixing height, wind speed, 1/L, Pasquill stability class). L is determined by observations at 2 or 2.5 m.

Trial	Date	Release time	Time of observation	Mixing height	Wind speed at z = 2 m	1/L	Pasquill stability class
		MDT = UTC -6	UTC	m	m/s	1/m	
1	8 24 15	1336	1335	1000	1.5	0.0678	E or F
6	8 31 16	1424	1430	1000	2.4	0.056	E
7	9 2 16	1356	1356	1000	4.0	0.0229	D or E

134 **Table 3**. Averaged simplified emission rates and parameters for Trials 1, 6 and 7 for both
135 single source and dual source assumptions.

	Trial 1	**Trial 6**	**Trial 7**
Single source, tank release into atmosphere			
Primary release			
Discharge rate (kg/s)	224.	260.	259
Discharge period (s)	20.3	32.2	33.3
Temperature (^{o}C)	-37.3	-37.4	-37.4
Vapor fraction	0.171	0.172	0.172
Density (kg/m^3)	18.32	18.15	18.12
Velocity (m/s)	50.8	44.2	44.2
Area (m^2)	0.241	0.324	0.323
Dual source, tank release with rainout and evaporation			
Primary release modified for rainout			
Discharge rate (kg/s)	146	168	162
Discharge period (s)	20.3	32.4	33.6
Temperature (^{o}C)	-37.3	-37.4	-37.4
Vapor fraction	0.262	0.266	0.274
Density (kg/m^3)	11.96	11.79	11.41
Velocity (m/s)	50.8	44.2	44.2
Area (m^2)	0.240	0.323	0.322
Evaporated rainout			
Discharge rate (kg/s)	43.2	34.0	34.0
Discharge period (s)	36.8	86.4	93.4
Temperature (^{o}C)	-37.3	-37.4	-37.4
Vapor fraction	1	1	1
Density (kg/m^3)	3.160	3.152	3.144
Area (m^2)	491	491	491

136

137

2 Overview of Models

Most of the 17 models have been routinely run by members of the JR II modeling working group. A few models were added in order to have broader coverage. The models can be divided into two groups – publicly-available and proprietary. Publicly-available models include those that are widely distributed around the world (such as SLAB) and those that are used only within the country or laboratory of development (e.g., PUMA). Proprietary models generally do not make their source code available and also charge a fee to users.

Some models include only a transport and dispersion module. Other models also include a meteorological processing module, an emissions module, and/or an exposure and health effects module. In the current exercise, we do not include health effects estimates, and we provide suggestions for input meteorology and emissions. However, modelers were given the option of using their own emissions and/or meteorological modules.

Below we provide brief summaries of each model and references where further details can be found. Some models (ADAM, B&M, DRIFT, HPAC, IDDM, PUMA, and PMSS) have papers in the current special issue.

2.1 Publicly available models

ADAM (Fabbri et al., 2017 and 2020, Fabbri and Wood, 2019) is a consequence assessment tool, recently developed by the Joint Research Centre of the European Commission. Its atmospheric dispersion module is an in-house modified version of SLAB, which introduces several modeling improvements such as the exact calculation of the average concentration for instantaneous releases, the calculation for time-varying releases, the inclusion of the contribution from pool evaporation in case of rainout, and the extension of the calculation for downward jets.

ALOHA (NOAA/EPA 1992) is a comprehensive dispersion model for hazardous gasses that is part of the CAMEO modeling software distributed by the US EPA and NOAA to emergency responders. Its passive (neutral) gas algorithm is based on the Gaussian plume model and its dense gas algorithm is based on DEGADIS (Havens and Spicer, 1990).

The B&M model is based on the Britter and McQuaid (1988) Workbook for Dense Gas Dispersion. It consists of two nomograms (one for instantaneous releases and one for continuous releases) where dimensionless concentration variations with downwind distance are presented as a function of a dimensionless cloud stability (depending on cloud excess density and size). Field and laboratory data are used to develop the curves. A crucial assumption for the current chlorine release trials regards specification of the initial cloud density and dimension scale (see the Hanna 2020 paper in this special issue).

The Canadian Urban Dispersion Modelling (CUDM) system was developed in a collaboration involving Environment Canada, the University of Alberta, Waterloo University, and the Canadian Department of Defence. The system consists of a wind model and a plume model component. The former is a high-resolution Reynolds-averaged Navier-Stokes (RANS) model with k-epsilon closure, that computes detailed winds around explicitly-resolved urban-scale streets and buildings provided from databases of 3D city building layouts. Inflow boundary conditions are supplied either from larger-scale 3D NWP analyzed or forecast fields or from single wind profiles that assume neutral vertical stability. An optional horizontally-uniform k-eddy wind configuration with no obstacles is also available. The plume model component of the CUDM is the well-mixed 3D Lagrangian stochastic particle trajectory model of Thomson (1987). The urban Lagrangian stochastic (urbanLS) model is based on Wilson et al. (2009) and uses passive scalar particles with dry deposition on surfaces parameterized by a deposition velocity. It does not account for plume density effects.

193 DRIFT (Gant et al 2018 and 2020) is a widely-used dense gas dispersion model
194 developed by the UK Health and Safety Executive (HSE). It is an integral model
195 similar in basic principle to DEGADIS, HEGADAS, and SLAB, and has been
196 widely validated with available dense gas field experiment observations.

197 The ESCAPE model (Expert System for Consequence Analysis and Preparing
198 for Emergencies) was developed at the Finnish Meteorological Institute and can
199 be used for evaluating the releases of toxic and flammable gases into the atmos-
200 phere, their source terms, atmospheric dispersion and the effects on humans and
201 the environment. Kukkonen et al. (2017) have presented the mathematical treat-
202 ments of this model and an evaluation of the model against selected experimental
203 field data. The treatments for atmospheric heavy gas and passive dispersion are
204 modified versions of the DRIFT model. The operational implementation of the
205 ESCAPE modelling system can be accessed anywhere using internet browsers, on
206 laptop computers, tablets and mobile phones. The model has been adapted to be
207 able to automatically use the real time predictions and forecasts of the numerical
208 weather prediction model HIRLAM, "HIgh Resolution Limited Area Model".

209 The HPAC/SCIPUFF (Hazard Protection Assessment Capability/Second Order
210 Closure Puff) model system was developed for US Department of Defense scenar-
211 ios (DTRA 2008, Sykes et al. 2014, Simpson et al. 2020). The dispersion model
212 can be described as a Lagrangian puff model. It can treat clouds with positive or
213 negative buoyancy, and has been validated with a broad array of hazardous gas
214 release field experiments.

215 IDDM (Integral Dense-gas Dispersion Model), developed at the US National
216 Center for Atmospheric Research (NCAR) accounts for source momentum, densi-
217 ty effects, and boundary layer processes in predicting the transport, growth, and
218 maximum concentrations due to dense gas releases. An early version called the

219 "NCAR" model was presented in Gant et al. (2018, Appendix A) with the more
220 recent IDDM form given by Weil (2019) and Weil and Alessandrini (2020).

221 PUMA is a real-time Gaussian puff model, developed by the Swedish Defence
222 Research Agency, with the purpose of providing support during ongoing emer-
223 gencies (Sigg et al., 2018). It operates on the local scale with meteorological
224 properties described by similarity theory and it has dense gas capabilities de-
225 scribed by Björnham and Burman (2016).

226 RAILCAR was originally developed by the US Navy to simulate emissions of
227 hazardous gases released from railcars. It uses basic physics and chemistry, along
228 with observations from several railcar accidents, to parameterize emissions condi-
229 tions for input to dispersion models. RAILCAR also includes a basic dispersion
230 model. See (Bauer 2013 and 2015, Bauer et al. 2018) for details. The US Navy
231 also funded the linking of RAILCAR with two dense gas dispersion models,
232 ALOHA and QUIC. The current model comparison exercise used RAILCAR-
233 ALOHA and RAILCAR-QUIC. However, the model outputs that we compare
234 here are from previous applications by the model developer to the JR II Trials 1,
235 6, and 7, using their own assumed emissions and meteorology, rather than the
236 inputs suggested by us.

237 SLAB (Ermak, 1990) is perhaps the most widely-used dense gas dispersion
238 model in the public domain. As its name infers, it is an integral slab model. Here
239 we include SLAB runs by two groups, INERIS (called SLAB-I) and RAND
240 (called SLAB-R). This exemplifies the typical differences encountered when the
241 same model is run by two expert users in different organizations.

242 VDI (1990) is a German regulatory model that is used to simulate dense gases.
243 The VDI Guideline 3783, Part 2 "Dispersion of Heavy Gas Emissions by Acci-
244 dental Releases" was released in 1990. The Guideline is based on results which
245 have been gained by systematic wind tunnel experiments. In these experiments the

246 release type was varied between an instantaneous release and a continuous gas
247 source. Overall 26 generic dispersion areas were investigated, ranging from a free
248 field without any buildings to different built up situations. By the application of
249 dimensional analysis relations, the results which have been gained in the small-
250 scale model can be transferred directly to fullscale released masses and to the en-
251 vironmental and ambient conditions, which are of practical interest. The guideline
252 is designed for the near field dispersion, where heavy gas effects are dominating,
253 down to a minimum concentration of 1 Vol.%. For estimations of the far field
254 effects, the VDI 3783 Part2 can be coupled with Part 1 of this guideline, enabling
255 us to estimate the gas dispersion down to ppm values.

256 2.2 Proprietary models

257 Phast is a commercial package developed by DNVGL containing a wide range
258 of models for consequence and risk assessment. The dispersion model (Witlox
259 and Holt, 1999) allows a range of source terms (evaporating pools, instantaneous
260 releases, finite duration and time-varying releases). It accounts for dense and
261 buoyant gases, 2-phase releases, droplet modelling and rainout. For cases that
262 rainout, an integrated pool spreading and evaporation model is included. Plumes
263 may evolve through one or more different stages: initial expansion, momentum
264 jet, heavy gas, and passive. In its most recent versions it includes enhancements to
265 model along-wind spreading for short duration and time-varying releases (Witlox
266 and Harper, 2014). Phast is widely-used by the chemical processing industry and
267 by many government agencies.

268 PMSS (Parallel Micro-SWIFT-SPRAY) combines algorithms developed by
269 public research agencies such as the French Atomic and alternative Energies
270 Commission (CEA) and by a commercial company, Aria Technologies. The
271 model was run for this exercise by Aria Technologies. PMSS aims to provide a
272 simplified but rigorous CFD solution of the flow and dispersion in built up envi-

ronments in a moderate amount of time. It is the efficiently parallelized version of MSS (Oldrini et al., 2017). MSS encompasses the local scale versions of SWIFT and SPRAY. Micro-SWIFT is a mass-consistent diagnostic model taking account of obstacles (such as the CONEX array). In its recent version, it also includes a RANS momentum solver. Micro-SPRAY is a Lagrangian Particle Dispersion Model able to account for obstacles (Tinarelli et al., 2012). It is also able to deal with the dispersion of a dense gas (Anfossi et al. 2010, Gomez et al. 2020). PMSS has been thoroughly validated against numerous wind tunnel and field experiments in complex environments, especially in the framework of the COST ES 10006 project (Trini Castelli et al., 2017).

SAFER/TRACE (SAFER Systems, 2014) is similar to Phast in that it is widely-used in the chemical processing industry. TRACE algorithms cover a ground-based or elevated release of dense or non-dense gasses. It can handle dispersion of gas or vapor/aerosol streams. It has comprehensive algorithms for release rate estimation from tank, pipe, and liquid pool (e.g., Meel and Khajehnajafi 2012 and Khajehnajafi and Pourdarvish 2011). Users have the ability to create gaseous or liquid mixtures. The liquid mixture model includes popular equation of state models (e.g., Peng Robinson, Soave, Redlich-Kwong) for hydrocarbon mixtures and Antoine, and Wilson for non-ideal solutions. Its chemical database covers most of the chemicals used in Industry. The TRACE fire model includes pool fire, jet fire, fireball, and flashfire. The explosion model includes vessel burst, BLEVE, and unconfined vapor cloud explosion. SAFER TRACE™ incorporates an intelligent wizard feature that allows the user to easily and rapidly build a scenario, create backs up and share scenarios with other users. Results can be viewed in tabular or graphical formats. Output information can be exported to other applications like word processors, spreadsheets and presentation managers, and Google Earth.

3 JR II brief description

As described by Fox et al. (2020), JR II took place over a broad flat desert at DPG. Releases were initiated between about 7 to 9 AM local time in late summer. The pressurized liquid chlorine tank was at the center of a 25 m diameter concrete pad, and the tank's bottom was 1 m above the surface. The main release was a two-phase momentum jet from a 15.2 cm diameter hole, with duration of about 20 to 40 seconds. The hole was at the tank's bottom for Trials 1 and 6, and pointing 45° downwards and downwind for Trial 7. During 2015, the source location was within a mock urban array, consisting of about 80 CONEX shipping containers set up on a packed gravel area 122 m square. For 2016, the CONEX obstacle array was removed. Concentrations were measured in the area containing the obstacle array, and also on 90° arcs at distances of 0.2, 0.5, 1, 2, 5, and 11 km. Continuous (with 1 to 3 s averaging time) samplers were used, as well as several UV lidars and still photography and videos. Table 4 lists the general characteristics of the three JR II trials used in the current paper. The details of the tank and the pressurized releases are described by Nicholson *et al.* (2017), Spicer and Miller (2017), and Spicer and Tickle (2020). Further discussion of the JR II field tests and model inter-comparisons is found in Fox *et al.* (2020).

Table 4. General characteristics of Trials 1, 6, and 7 in JR II.

Trial	Time UTC (MDT = UTC - 6)	Mass released kg	Duration of release s	Mass release rate kg/s	Wind speed at z = 2 m m/s	Wind direction z = 2 m	1/L 1/m	Pasquill stab class
1	8 24 15 1336	4547	20.3	224	1.5	147	0.068	E or F
6	8 31 16 1424	8372	32.2	260	2.4	147	0.056	E
7	9 2 16 1356	8625	33.3	259	4.0	150.0	0.0229	D or E

Trials 1, 6 and 7 were chosen as initial representative trials. They have downwards (Trials 1 and 6) or 45^{o} downwards (Trial 7) jet directions, relatively large observed concentrations, and relatively steady winds. Because the plume was not on the edge of the sampler network, full definition of peak concentration and plume width was possible. In the future, the six additional JR II trials may be evaluated by some modelers as further research funding allows. Although there were samplers at distances less than 200 m, these were not considered here as many models cannot simulate the initial dense momentum jet or the detailed effects of the obstacle array. The samplers used for arc max concentrations were located at a height of 0.3 m. To estimate cloud height, a few short towers were available with observations at 0.3, 3, and 6 m. See Chang and Mazzola (2020) for more information on the sampler arrays.

4 Results of model comparisons

After the 17 model runs were completed for the three JR II Trials in Table 4, two types of methods for comparing model predictions with observations were used. The first type is a simple visual inspection approach using tables and scatter plots. There are few enough numbers that they can be easily seen in a single table or scatter plot. In addition to the visual inspections, we generated some simple comparisons, such as the fractions of cases when a model would overpredict or underpredict by more than a factor of 2. The second type is application of the BOOT model evaluation software (Chang and Hanna 2004), where performance measures such as FB (Fractional mean Bias), MG (Geometric Mean bias), NMSE (Normalized Mean Square Error), VG (Geometric Variance, FAC2 (fraction of predictions within a factor of 2 of observations), and FAC5 and FAC10 are calculated, and statistical significance tests are applied. The first four performance measures are defined as follows:

$$FB = (\langle C_o \rangle - \langle C_p \rangle)/(0.5(\langle C_o \rangle + \langle C_p \rangle))$$

$$MG = \exp(\langle \ln C_o \rangle - \langle \ln C_p \rangle)$$

$$NMSE = \langle (C_o - C_p)^2 \rangle / (\langle C_o \rangle \langle C_p \rangle)$$

$$VG = \exp(\langle (\ln C_o - \ln C_p)^2 \rangle)$$

where C_o and C_p are observed and predicted variables (arc max concentrations in the below applications), respectively, and the < > brackets indicate an average. A perfect model, with C_o and C_p always equal to each other, has FB = NMSE = 0, MG and VG = 1, and FAC2 = 1.

4.1 Arc max C

First, we will consider the maximum peak concentration (arc max C) on each of the sensor arcs. The sensor data have undergone extensive quality assurance and control. Some sensors saturated (exceeded their calibration limits), but we still include such peaks with qualification. Although the sensors used to determine the arc max C observation were all located at a height of 0.3 m, some models predicted concentrations at a height of 0 m (ground level). At the downwind distances of these sensors (x greater than or equal to 200 m), the differences in predicted concentration between heights of 0 and 0.3 m are expected to be less than a few per cent. Chang and Mazzola (2020) further describe these sensor data.

Figure 1 is a plot containing model predictions and observations of arc max C as a function of distance for Trial 7. Of the three trials, Trial 7 has the largest observed concentrations, and there is more of a tendency for slight model underprediction. In this figure, eight of the models are within a factor of 2 of the observations at more than half of the downwind distances. The other models tend to underpredict by more than a factor of two at most downwind distances. At most distances, the range of the 17 model predictions covers about 1 ½ orders of magnitude. It is also seen that a -5/3 power law reasonably fits the data from unsaturated sensors. Dense gas modelers have recognized for almost 50 years that the arc max C from a ground-level dense gas source will tend to follow an approximate -5/3 power law (e.g., see Hoot and Meroney 1974, whose wind tunnel data yielded a -1.68 power law). In order to see individual model points more clearly, a zoomed-in plot for just the outer four arcs is presented in Figure 2. Here the tendency to underpredict by most models is clearer. For example, we can see that all but one of the models underpredict at 5 km distance, and most of the models underpredict by more than a factor of 2.

It is worth pointing out that, of the three trials being studied here, Trial 7, with a 45^{o} downward jet, was likely the most challenging for the models. For instance, the ADAM model can simulate either downward or horizontal jet directions, but not an angle in-between. For this comparison we simply averaged results for these two geometries in ADAM.

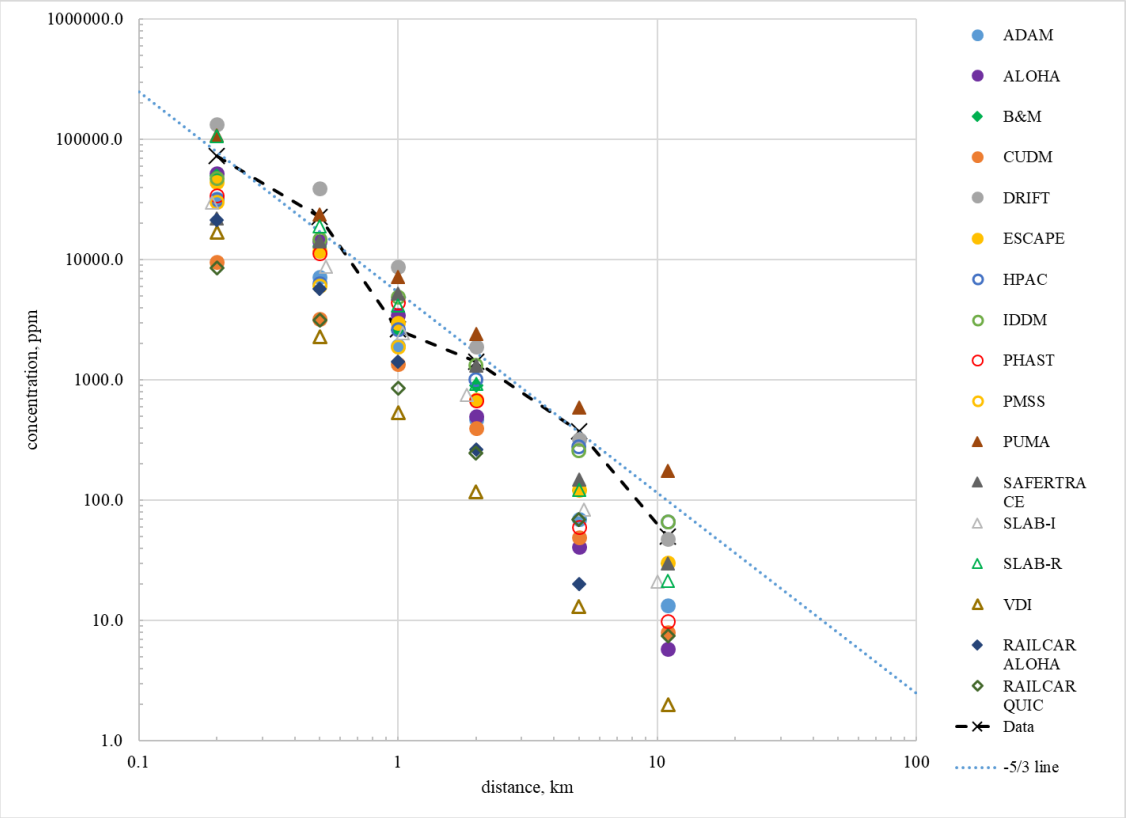


Figure 1. Plot of predicted and observed arc max C for all models for Trial 7. The observations (averaging times of 1s-3s) are the X's connected by a dashed line. Saturated observations (at 1 and 11 km) have a light gray box around the X. The straight blue-dotted line represents a -5/3 slope.

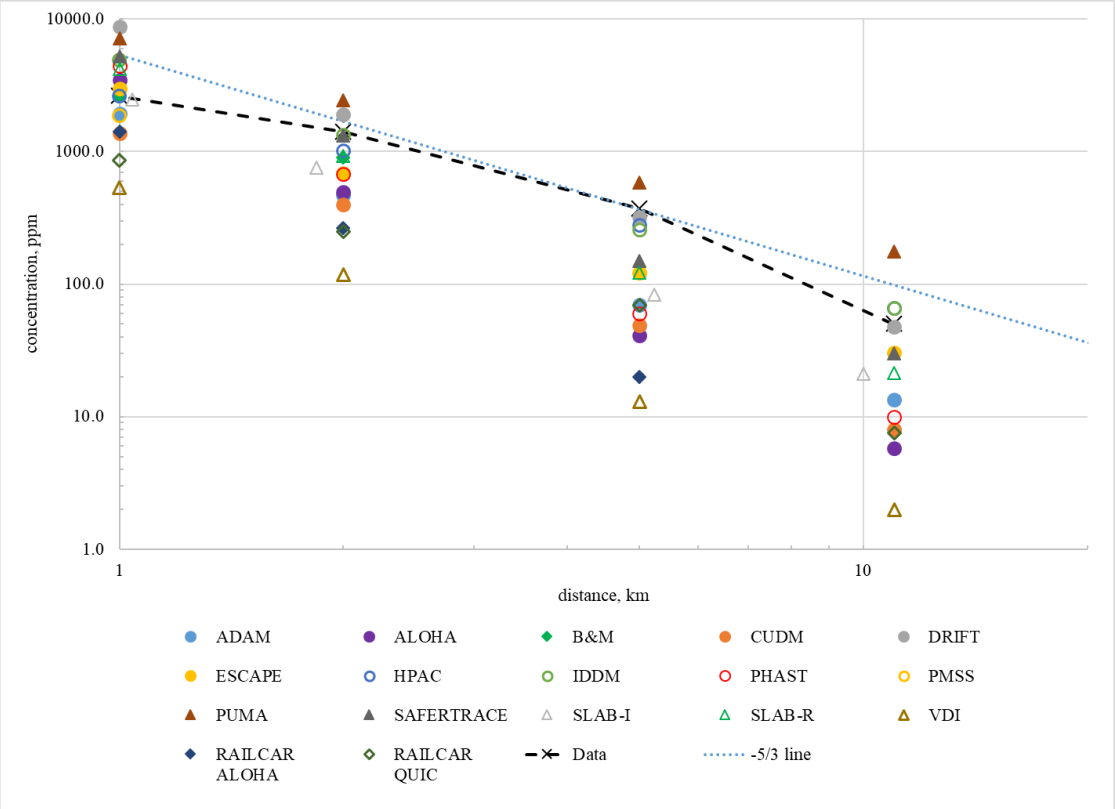


Figure 2. Zoomed plot of predicted and observed arc max C as a function of distance for all models for Trial 7 for the most distant four arcs at 1, 2, 5 and 11 km. Otherwise, as for Figure 1.

For Trial 7, tallies (counts) have been made of the numbers of times that the modeled arc max C is larger by more than a factor of two, smaller by more than a factor of two, and within a factor of two of the observed arc max C. The counts are displayed in Table 5. Of the 101 model entries, 49 are within a factor of two (black), 49 underpredict by greater than a factor of two, and only 3 overpredict.

Table 5. 17 model predictions and observations of arc max C for Trial 7 for the six downwind distances. **Black** indicates prediction and observation within a factor of 2; **Blue** indicates underprediction larger than a factor of two; **Red** indicates overprediction

larger than a factor of two. For the observed concentrations, **Orange** indicates observed value is saturated.

x (km)	0.2	0.5	1	2	5	11
ADAM	32368	7192	1923	471	69	13
ALOHA	52300	14100	3470	497	41	6
B&M	50000	13000	2600	900	120	30
CUDM	9560	3230	1365	396	49	8
DRIFT	132908	38929	8694	1892	321	48
ESCAPE	44044	11833	3001	685	122	30
HPAC	31372	6314	2633	1005	278	66
IDDM	47472	14738	4886	1336	257	66
PHAST	34153	11304	4418	681	60	10
PMSS	30216	6044	1878	743	128	29
PUMA	108480	23835	7163	2428	586	176
SLAB-I	29800	8750	2460	750	83	21
SLAB-R	107047	18793	4185	925	121	21
TRACE	22154	14110	5198	1312	149	30
RAILCAR ALOHA	21290	5729	1417	266	20	
RAILCAR QUIC	8598	3138	861	248	69	8
VDI	16823	2277	537	117	13	2
Observed	**72580**	**22686**	**2621**	**1417**	**373**	**50**

Arc max C summary plots, similar to that for Trial 7 in Figure 1, are given in Figures 3 and 4 for Trials 1 and 6, respectively. For the three JR II trials being analyzed (1, 6, and 7), some differences in mean bias and scatter in arc max C model predictions and observations are expected. For example, the observed arc max C values are seen to generally fall below the median of the range of model

predictions for Trial 1, near the median for Trial 6, and above the median for Trial 7, There are many rational scientific explanations, such as variations in the mass released, the wind speed, the ambient stability at the time of release, the presence of the CONEX array, and the location of the hole. A few of the differences can be explained by basic science principles, such as the fact that both modeled and observed concentrations approximately scale with mass of chlorine released.

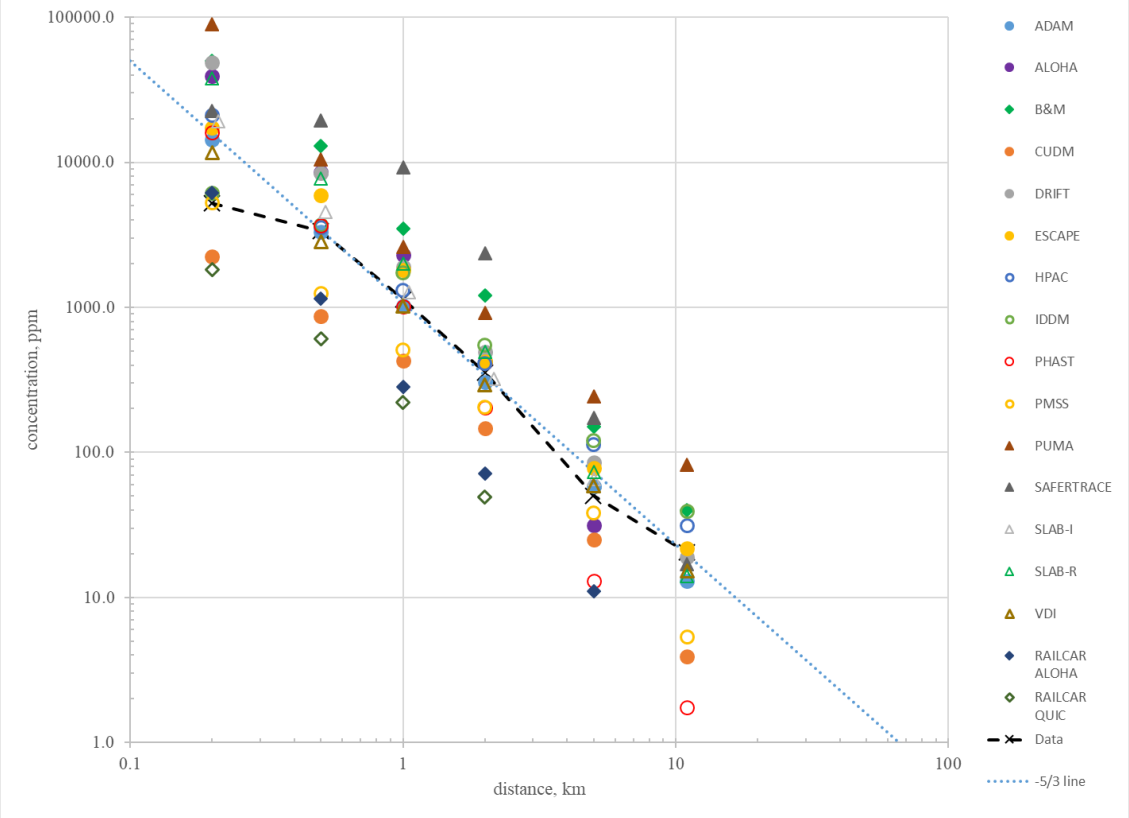


Figure 3. Plot of predicted (for all models) and observed arc max C for Trial 1. The observations (averaging times of 1 – 3 s) are indicated by the X's connected by a dashed line. Saturated observations are at 0.2 and 5 km. The straight blue-dotted line represents a -5/3 slope.

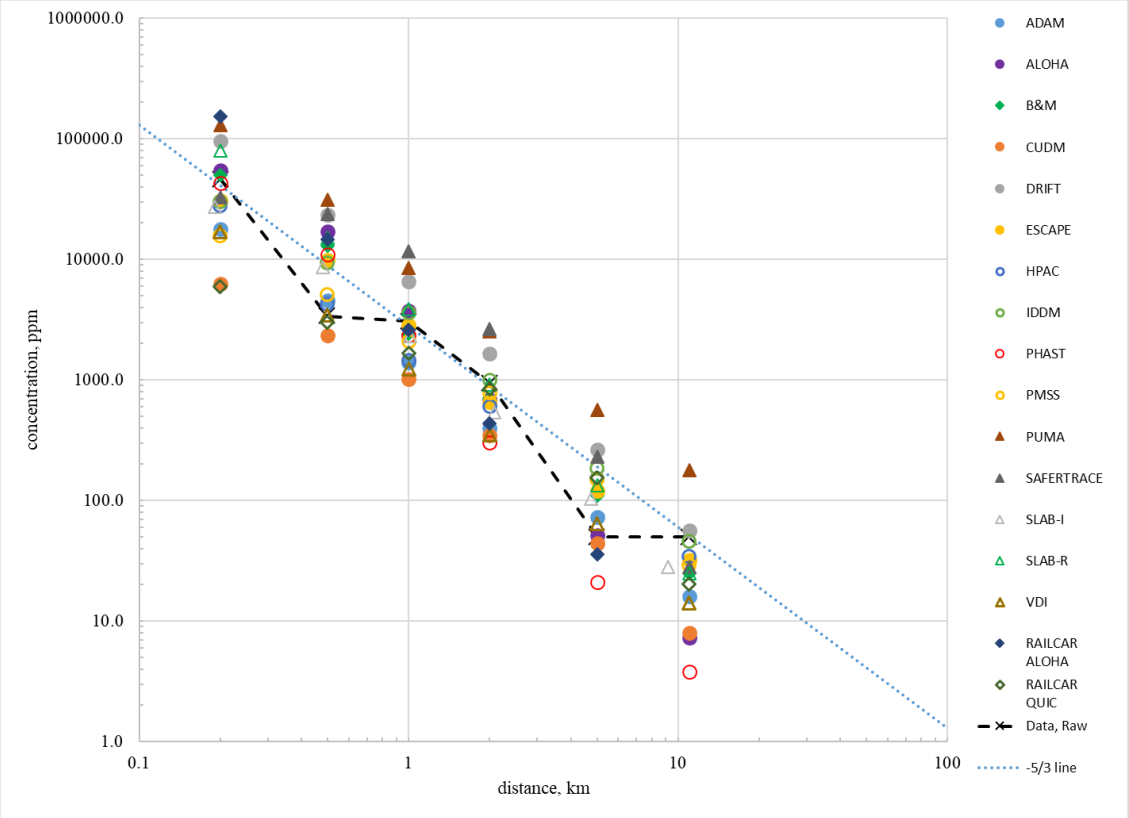


Figure 4. Plot of predicted (for all models) and observed arc max C for Trial 6. The observations (averaging times of 1 – 3 s) are the X's connected by a dashed line. Saturated observations are at 0.5 and 5 km. The straight blue-dotted line represents a -5/3 slope.

Considering all 17 model predictions of arc max C for all 3 trials and at 6 downwind distances, the range of model predictions for any trial or distance is consistently about 1 to 1 ½ orders of magnitude in Figures 1 through 4. The observed arc max C values are always inside the range of the 17 model predictions. A comparison of the magnitudes of the arc max C's along the best fit $x^{-5/3}$ lines for the three trials suggest some empirical conclusions. At any x, the ratio of concentrations on the best fit line during Trial 1 to those during Trial 6 is 0.39, and the ratio during Trial 6 compared to Trial 7 is 0.52. Thus, the ratio for Trial 1 to Trail 7 is 0.20. Trial 1 likely has the smallest concentrations because of two major

466 reasons mentioned earlier: (1) its mass released was about half of that for Trials 6
467 and 7, and (2) the CONEX array would increase turbulence and therefore increase
468 rates of dispersion in the first 100 m. Trial 7 likely has the largest concentrations
469 because of two major reasons: 1) its 45° downward and along-wind jet orienta-
470 tion, thus decreasing the width of the initial jet, and 2) the wind speed was more
471 than twice as large as that for Trials 1 and 6, causing the cloud to move downwind
472 faster with less total dispersion at any downwind distance. Britter and McQuaid
473 (1988) provided further details of the rationale for why the short-term arc max C
474 can be larger (a worst case) for large wind speeds. This is the opposite of what
475 happens for continuous releases.
476
477 Table 6 extends the factor of two analyses for Trial 7 in Table 5 to Trials 1 and
478 6. Counts of the over, under and within factor of two predictions by all models
479 for these three trials and for all distances are listed. It is seen that about half of
480 the model predictions are within a factor of 2 of the observations and there are a
481 few less overpredictions than underpredictions by greater than a factor of 2.
482 Models with more arc max C comparisons within a factor of two of observations
483 include ESCAPE, HPAC, and IDDM. It was found that there are 6 models that
484 have less mean bias and scatter for all three trials. Some models have larger mean
485 biases for one trial, and smaller mean biases for the other two trials. A few models
486 show consistent under or over predictions by a factor of 2 for all three trials.
487
488

Table 6. Summary of numbers of times the observed arc max C is overpredicted (**red**) or underpredicted (**blue**) by more than a factor of two, or predictions were within a factor of two (**black**) of observations for the three trials and the six downwind distances. Percentages for each model's predictions are in parentheses.

	number red (overpredicted)	**number black (within factor of 2)**	**number blue (underpredicted)**
ADAM	1 (5.6%)	8 (44.4%)	9 (50%)
ALOHA	3 (17.6%)	10 (58.8%)	4 (23.5%)
B&M	7 (38.9%)	10 (55.6%)	1 (5.6%)
CUDM	0 (0%)	4 (22.2%)	14 (77.8%)
DRIFT	7 (38.9%)	11 (61.1%)	0 (0%)
ESCAPE	3 (16.7%)	13 (72.2%)	2 (11.1%)
HPAC	3 (16.7%)	12 (66.7%)	3 (16.7%)
IDDM	3 (16.7%)	15 (83.3%)	0 (0%)
PHAST	2 (11.1%)	7 (38.9%)	9 (50%)
PMSS	1 (5.6%)	10 (55.6%)	7 (38.9%)
PUMA	14 (77.8%)	4 (22.2%)	0 (0%)
SLAB-I	3 (17.6%)	10 (58.8%)	4 (23.5%)
SLAB-R	4 (22.2%)	12 (66.7%)	2 (11.1%)
TRACE	9 (50%)	7 (38.9%)	2 (11.1%)
RAILCAR ALOHA	2 (13.3%)	4 (26.7%)	9 (60.0%)
RAILCAR QUIC	1 (5.9%)	3 (17.6%)	13 (76.5%)
VDI	1 (5.6%)	7 (38.9%)	10 (55.6%)
Sum	**64 (21.3%)**	**147 (49.0%)**	**89 (29.7%)**

Despite the biases and scatter discussed above, it is clear that these widely-used dense gas models are all able to simulate the basic behavior of the observa-

tions of arc max C. There is not a consistent mean bias one way or another. The scatter also is consistent over the three trials and six downwind distances. There is no general trend in bias or scatter with distance.

In order to assess the significance of model performance differences in their predictions of arc max C, the BOOT statistical analysis methods were applied (Chang and Hanna 2004). Table 7 presents the performance measures VG, MG, FAC2, and FAC5 for the 17 models for Trials 1, 6, and 7, and for all arcs except the 11 km arc. Not all models produced predictions on the 11 km arc. The medians across all 17 models can be determined: 2.69 for VG, 1.04 for MG, 0.47 for FAC2 and 0.93 for FAC5. Recall that for a perfect model, all of these performance measures will equal 1.0. MG < 1.0 indicates an overprediction mean bias and MG > 1.0 indicates an underprediction mean bias. The ranges across all 17 models are 1.4 to 11.9 for VG, 0.31 to 3.55 for MG, 0.2 to 0.8 for FAC2, and 0.4 to 1.0 for FAC5.

Chang and Hanna (2004) and Hanna and Chang (2012) suggested some "acceptance criteria" for the model performance measures for rural and for urban scenarios, respectively. For example, if FAC2 > 0.6 a model is "acceptable" for rural areas. The criterion is FAC2 > 0.3 for urban areas where modeling has more uncertainties. These criteria were developed based on results of many model evaluation exercises with many research-grade field experiments. The Jack Rabbit II field study took place over a flat desert playa for downwind distances > 100 m. For the 2015 trials, with the mock urban array (CONEXs) in place at x < 100 m, there might be some effect at distances beyond the end of the array. Also, the above acceptance criteria were developed without consideration of dense gas experiments. Nevertheless, in Table 7, FAC2 > 0.3 for 12 of the 17 models, and FAC2 > 0.6 for 3 of the models.

Table 7. BOOT calculations of the model performance measures Geometric Variance VG, Geometric Mean MG, FAC2 and FAC5 for the arc max concentration for the three trials and for arcs at 0.2, 0.5, 1, 2, and 5 km.

	VG	MG	FAC2	FAC5
ADAM	1.91	1.50	0.47	0.93
ALOHA	2.69	0.92	0.60	0.87
B&M	2.80	0.66	0.47	0.93
CUDM	5.34	3.16	0.27	0.73
DRIFT	3.04	0.43	0.53	0.80
ESCAPE	1.58	0.93	0.67	1.00
HPAC	1.65	1.04	0.60	1.00
IDDM	1.40	0.79	0.80	1.00
PHAST	2.24	1.44	0.40	0.93
PMSS	1.81	1.59	0.53	1.00
PUMA	6.89	0.31	0.27	0.80
RC/ALOHA	6.24	2.20	0.27	0.80
RC/QUIC	11.87	3.55	0.20	0.40
SLAB-I	1.74	1.13	0.60	1.00
SLAB-R	2.03	0.65	0.67	0.93
TRACE	5.97	0.45	0.27	0.73
VDI	7.92	2.43	0.40	0.80

The reader can see that, for each performance measure in Table 7, there are several models with values of, say, MG or FAC2 that are close to each other (say, within plus or minus 10 or 20 %). The question therefore arises whether the models with similar MG or FAC2 are significantly different, using the statistical definition. This question can be answered by BOOT, since one of the outputs is an indication about whether or not it can be concluded, with 95% confidence, that there is a significant difference in a performance measure for two models. The results for the Geometric Mean MG are recorded in Table 8, where ‘X’ means

that the difference in the natural logarithm (ln) of the geometric mean MG for two models is significantly different from 0, at the 95 % confidence level. This table was developed using the 17 models whose performance measures are in Table 7, and shows that there is no significant difference in ln(MG) for the five models with MG closest to 1; ALOHA, ESCAPE, HPAC, IDDM, and SLAB-I. However, the same holds true for the four models with the largest underpredictions and MG greater than 2; CUDM, RAILCAR/ALOHA, RAILCAR/QUIC, and VDI; and for the three models with the largest overpredictions and MG less than 0.5; DRIFT, PUMA and TRACE. Considering all 17 models compared with each other, only about 20 % of the pairs in Table 8 show no significant difference.

Table 8. Summary of significance of model performance differences based on differences in ln of geometric mean bias in arc max C for the three trials and the five closest downwind distances. An X indicates a model pair where there is a significant difference at the 95% confidence level. Beige indicates those models with MG closest to 1; yellow those with MG > 2; and blue with MG < 1/2. Note that the table is symmetric but the bottom half is not shown for visual simplicity.

	ADAM	ALOHA	B&M	CUDM	DRIFT	ESCAPE	HPAC	IDDM	PHAST	PMSS	PUMA	RC/ALOHA	RC/QUIC	SLAB-I	SLAB-R	TRACE	VDI
ADAM		X	X	X	X	X	X	X			X		X	X	X	X	X
ALOHA			X	X	X				X	X	X	X	X		X	X	X
B&M				X	X	X	X		X	X	X	X	X	X		X	X
CUDM					X	X	X	X	X	X	X			X	X	X	
DRIFT						X	X	X	X	X	X	X	X	X	X		X
ESCAPE									X	X	X	X	X	X	X	X	X
HPAC										X	X	X	X		X	X	X
IDDM									X	X	X	X	X	X		X	X
PHAST											X	X	X		X	X	X
PMSS											X		X	X	X	X	
PUMA												X	X	X	X		X
RC/ALOHA														X	X	X	
RC/QUIC														X	X	X	
SLAB-I															X	X	X
SLAB-R																	X
TRACE																	X
VDI																	

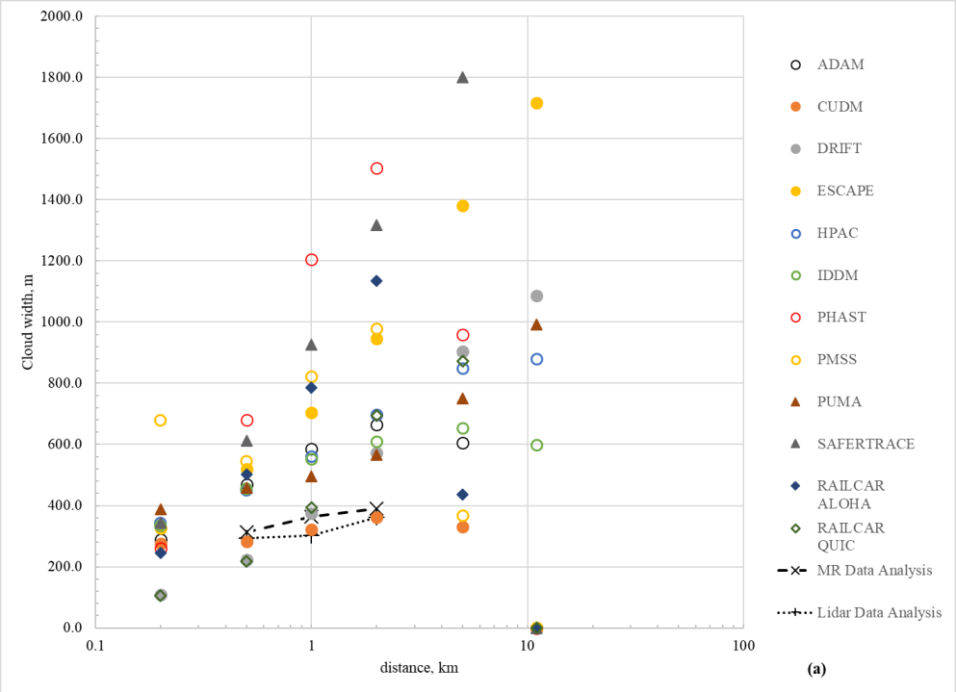
4.2 Cloud width and height

As described by Mazzola (2020), cloud width and height were estimated using a combination of samplers and lidars at distances of 0.5, 1, and 2 km. The observed cloud width and height were estimated at the time when the arc max C occurred at each arc. The cloud edges were defined at two concentration levels: 20 ppm and 200 ppm. Videos from a drone showed that the instantaneous visible cloud at distances of 0.5 to 2 km had a “mushroom” shaped footprint, with a broad leading “head” due to the initial momentum jet release and associated dense gas horizontal spreading, and a narrower trailing cloud due to evaporation from liquid rainout from the concrete pad around the source (see Fox et al. 2020). Since

575 the “head” had the largest concentrations, the cloud widths and heights correspond
576 to that part of the visible cloud. For example, Figure 5 shows the predicted versus
577 observed cloud widths and heights for Trial 7. Note that observations are plotted
578 for samplers (dashed line with X symbol) and for lidar (dotted line). As explained
579 by Mazzola (2020), observed cloud height was estimated from lidars where avail-
580 able plus a few instances where there was a short tower with samplers at heights
581 of 0.3, 3, and 6 m.
582
583 Predictions of cloud width and height are available from only about half to 2/3
584 of the models. It is seen in Figure 5 that most of the models overpredict the cloud
585 width, sometimes by as much as a factor of 3 or 4. The median of the biases cal-
586 culated for all the models indicates about 70 % overprediction. However, two of
587 the models have a bias less than about 25 % at distances of 0.5 and 1 km. At dis-
588 tances of 5 and 11 km, where there were no observations, the expected decrease in
589 cloud width at a fixed edge concentration is evident among about half of the mod-
590 el predictions.
591
592 Regarding cloud height during Trial 7, at a distance of 1 km, the range of the
593 model predictions encompass the observations (uncertainty limits are shown).
594 Over all models, there is about a 30 % median bias towards overprediction. The
595 median bias across all models is near 0 at 1 km, is about 53 % at 0.5 km and about
596 33 % at 2 km.
597
598 Similar to the format in Table 5, Table 9 provides counts of factor of 2 agree-
599 ment between model predictions and observations of cloud width and height to 20
600 ppm by distance. No models underpredict cloud width by more than a factor of 2.
601 58% of models predict within a factor of 2, while 42 % overpredict cloud width

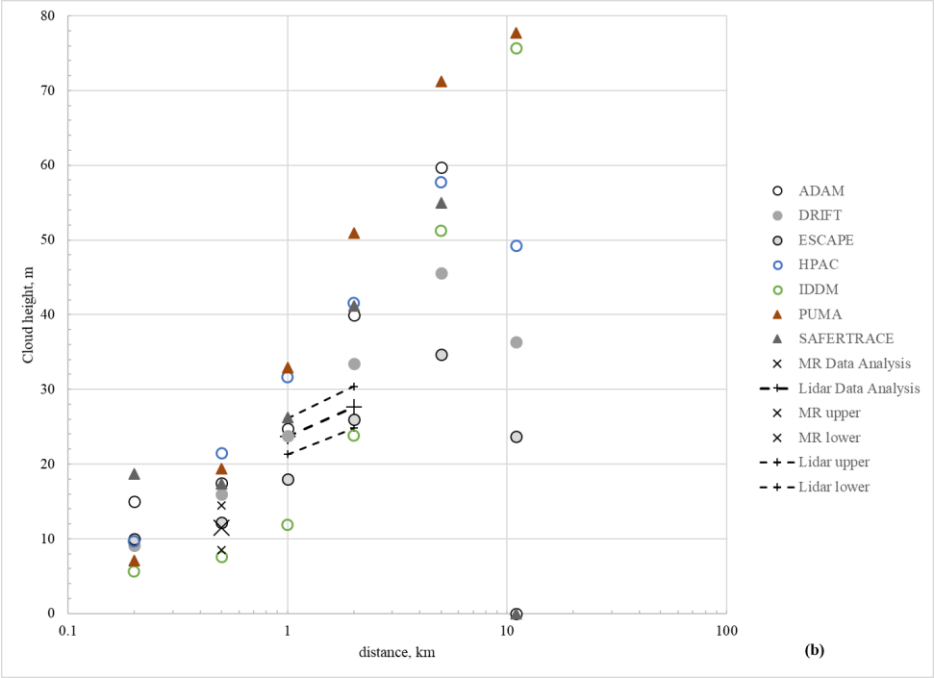
602 by greater than a factor of 2. For cloud height, all of the models' predictions are
603 within a factor of 2.
604
605 Although the plots for width and height at a concentration edge of 200 ppm are
606 not shown here, Table 10 provides a summary of performance. It is seen that
607 most models predict cloud dimensions to 200 ppm within a factor of 2, while
608 about 25 % of the models overpredict both width and height by greater than a fac-
609 tor of 2. The overpredictions by a model are listed for either cloud width or
610 height, but not for both.

611



612

613
614



615

616 **Figure 5.** Comparison for Trial 7 of model-predicted and observed cloud widths **(a)** and
617 heights **(b)** (to 20 ppm) as a function of downwind distance. The X points are from sam-
618 plers and the dashed line is from lidars, for which uncertainty estimates are shown.

619

620

621 **Table 9.** Model predictions and observations for 20 ppm cloud width (in m) (12 models)
622 and height (in m) (7 models) at time of arc max C for Trial 7 for three downwind
623 distances. **Black** is prediction and observation within a factor of 2; **Blue** is underpredic-
624 tion larger than a factor of two (none shown); **Red** is overprediction larger than a factor
625 of two.

	Width (m)			Height (m)		
x (km)	0.5	1	2	0.5	1	2
ADAM	469	586	665	17.5	24.8	39.9
CUDM	282	322	361			
DRIFT	223	374	573	15.9	23.7	33.4
ESCAPE	519	704	947	12.2	18.0	26.0
HPAC	451	561	697	21	32	42
IDDM	455	551	610	7.6	11.9	23.8
PHAST	680	1205	1503			
PMSS	545	822	977			
PUMA	459	495	566	19.4	32.9	50.9
TRACE	611	926	1317	17.4	26.2	41.2
RAILCAR ALOHA*	502	786	1134			
RAILCAR QUIC*	218	393	695			
Median for all models	**464**	**568**	**680**	**17.6**	**24.2**	**36.6**
Observed MiniRae	313	364	391	11.5		
Observed Lidar	293	303	363		23.7	27.6
Observed Avg	**303**	**334**	**377**	**11.5**	**23.7**	**27.6**

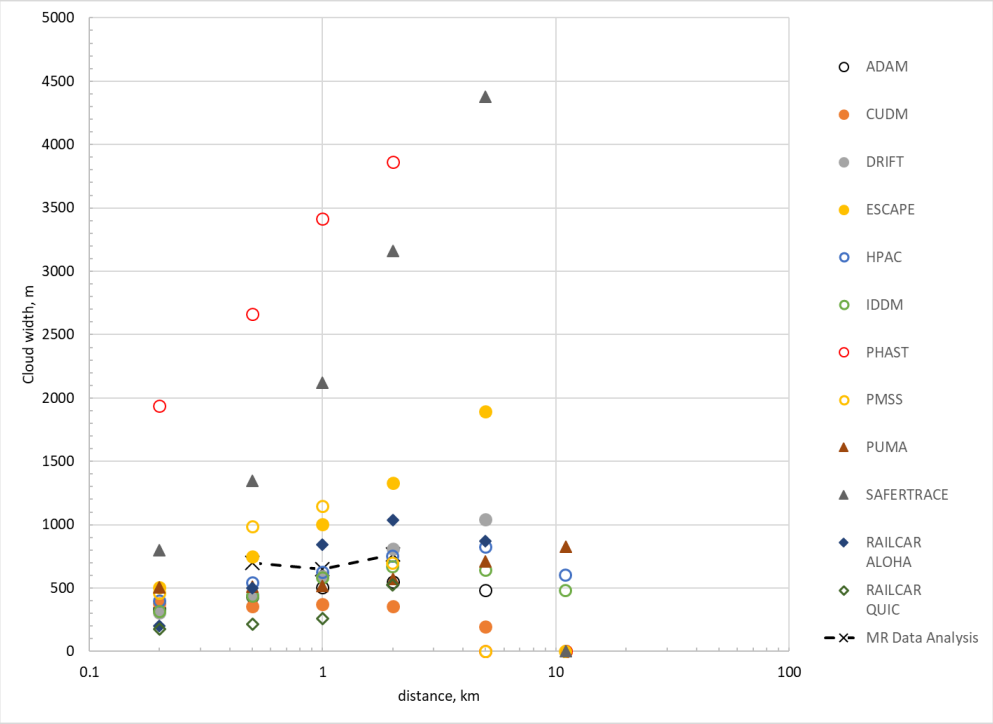
626 * - Results reported at 10 ppm, not 20 ppm used by all other models and observations
627
628

Table 10. Model predictions and observations for 200 ppm cloud width (in m) (10 models) and height (in m) (7 models) at time of arc max C for Trial 7 for three downwind distances. **Black** is prediction and observation within a factor of 2; **Blue** is underprediction larger than a factor of two (none shown); **Red** is overprediction larger than a factor of two.

	Width (m)			Height (m)		
x (km)	0.5	1	2	0.5	1	2
ADAM	372	415	335	12.5	16.7	19.5
CUDM	218	219	186			
DRIFT	218	351	497	10.8	15.4	19.8
ESCAPE	517	697	917	7.9	10.3	10.9
HPAC	364	431	456	17	23	27
IDDM	367	420	410	6.2	9.1	16
PHAST	536	905	870			
PMSS	504	717	170			
PUMA	381	377	395	15.4	25.1	36.9
TRACE	541	790	1027	11.7	15.7	18.3
Median for all models	**377**	**425**	**433**	**10.3**	**15.6**	**18.9**
Observed MiniRae	**292**	**339**	**301**	**6.21**	**15.8**	**10.0**

Figure 6 shows the predicted vs observed cloud widths to 20 ppm for Trial 1. Note that observations are plotted for samplers (dashed line with X symbol) and there are no lidar observations. The bias of the models' predictions of the cloud width is less for Trial 1 than for Trial 7, since the predicted widths are scattered more evenly about the observed width. Two of the models show the largest over-predictions (by about a factor of 4 or 5). Observed cloud height was estimated only at the 0.2 km distance, where there were two short towers with samplers at heights of 0.3, 3, and 6 m. The predicted model heights and the observed height

643 (the larger of the values from the two samplers) to 20 ppm are given in Table 11.
644 The median of the predicted heights at 0.2 km is only 6 % smaller than the ob-
645 served value. The range of the predictions is from 0.38 to 2.30 times the observed
646 height. The model differences may be partly caused by the different treatments by
647 each model of the effects of the CONEX array just upwind of this arc.
648
649



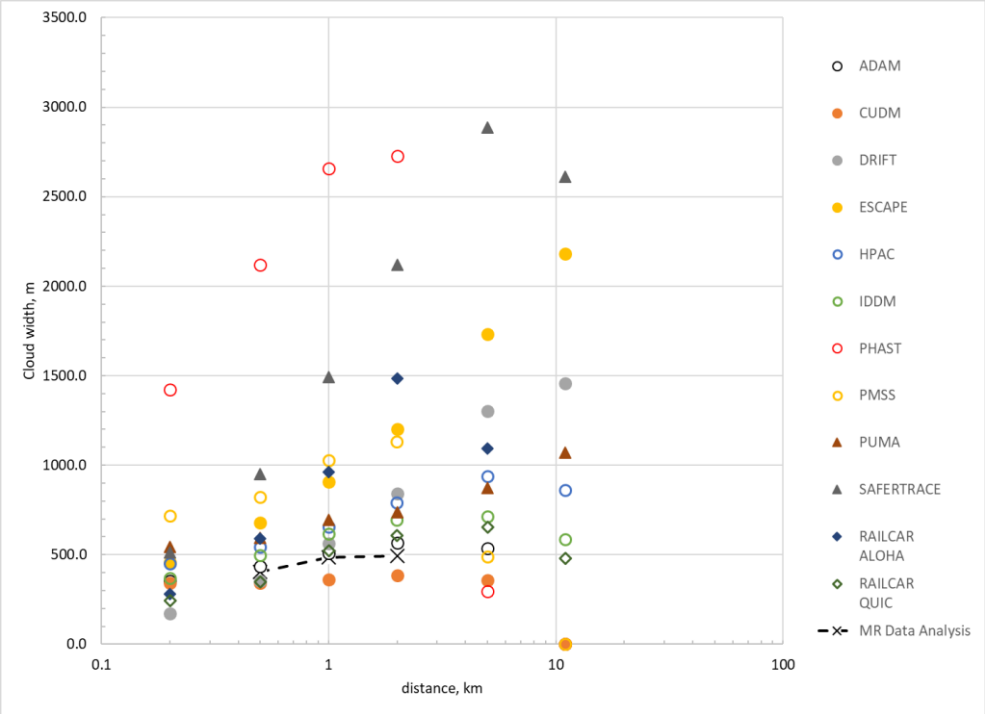
650

651 **Figure 6.** Comparison for Trial 1 of model-predicted and observed cloud widths (to 20
652 ppm) as a function of downwind distance. The X points are from samplers.
653
654

Table 11. Model predictions and observation of 20 ppm cloud height (in m) (7 models) at time of arc max C for Trial 1 for 0.2 km downwind distance. **Black** is prediction and observation within a factor of 2; **Blue** is underprediction larger than a factor of two; **Red** is overprediction larger than a factor of two.

x (km)	0.2
ADAM	3.9
DRIFT	5.1
ESCAPE	4.5
HPAC	10
IDDM	17.1
PUMA	8.5
TRACE	23.4
Median for all models	**8.5**
Observed MiniRae	**9.2**

Figure 7 shows the predicted vs observed cloud widths to 20 ppm for Trial 6. Note that observations are based only on sampler observations (dashed line with X symbol); there were no lidar observations. All but two of the models overpredict the cloud widths, with two models overpredicting by more than a factor of 5. The median predicted width over all the models at 1 km is about 50 % greater than observed. Observed cloud height was estimated only at 0.5 km distance where there were two short towers with samplers at heights of 0.3, 3, and 6 m. The larger of the observed height for the two towers and the predicted heights to 20 ppm by seven models are listed in Table 12. The median model-predicted height is only 8 % less than the observed height at 0.5 km, and the model predictions range from 0.7 to 1.9 times the observed height.



673

674 **Figure 7.** Comparison for Trial 6 of model-predicted and observed cloud widths (to 20
675 ppm) as a function of downwind distance. The X points are from samplers.

676

677

678 **Table 12.** Model predictions and observations for 20 ppm cloud height (in m) (7 models)
679 at time of arc max C for Trial 6 for 0.5 km downwind distance. **Black** is prediction and
680 observation within a factor of 2; **Blue** is underprediction larger than a factor of two (none
681 shown); **Red** is overprediction larger than a factor of two (none shown).

x (km)	0.5
ADAM	10.7
DRIFT	8.4
ESCAPE	8.1
HPAC	21.7
IDDM	8.3
PUMA	10.8
TRACE	11.6
Median for all models	**10.7**
Observed MR	**11.6**

682

683 Table 13 extends the factor of two analyses of widths and heights for Trial 7
684 (previously presented in Tables 6 and 7) to Trials 1 and 6. The table includes data
685 from all three trials. Counts of the over, under and within factor of 2 predictions
686 by all models for these three trials and for all distances available are listed for
687 both cloud width and cloud height. Note that not all models provided predictions
688 for all combinations. It is seen that 74 % of the model predictions are within a
689 factor of 2 of the observations (70 % for width and 84 % for height). There are
690 more overpredictions (22 %) than underpredictions (5 %) by greater than a factor
691 of 4. Many models mostly predict dimensions within a factor of two of observa-
692 tions. It was found that there are 5 models that have a tendency to overpredict
693 dimensions for 25 % or more of the comparisons. The overprediction tendency is
694 perhaps to be expected given that the models are formulated to represent an en-

695 semble mean outcome, whereas these tests are individual realizations with no en-
696 sembles available.

697 **Table 13.** Summary of numbers of times the observed cloud width or height (to 20
698 and 200 ppm) is overpredicted (**red**) or underpredicted (**blue**) by more than a factor
699 of two, or predictions were within a factor of two (**black**) of observations for the
700 three trials and the six downwind distances. "x" means no prediction. Percentages
701 for each model's predictions are in parentheses.

	number red (overpredicted)		**number black (within factor of 2)**		**number blue (underpredicted)**	
	Width	**Height**	**Width**	**Height**	**Width**	**Height**
ADAM		1 (10%)	18 (100%)	8 (80%)		1 (10%)
CUDM		x	13	x	5	x
DRIFT	1 (5.6%)		17 (94.4%)	10 (100%)		
ESCAPE	7 (38.9%)		11 (61.1%)	9 (90%)		1 (10%)
HPAC		3 (30%)	18 (100%)	7 (70%)		
IDDM		1 (10%)	18 (100%)	9 (90%)		
PHAST	16 (88.9%)	x	2 (11.1%)	x		x
PMSS	7 (38.9%)	x	7 (38.9%)	x	4 (22.2%)	x
PUMA		2 (20%)	18 (100%)	8 (80%)		
TRACE	15 (83.3%)	2 (20%)	3 (16.7%)	8 (80%)		

RAILCAR ALOHA*	3 (33.3%)	x	6 (66.7%)	x		x
RAILCAR QUIC*		x	7 (77.8%)	x	2 (22.2%)	x
Sum	**49 (24.7%)**	**9 (12.9%)**	**138 (69.7%)**	**59 (84.3%)**	**11 (5.6%)**	**2 (2.9%)**

* - Results reported at 10 ppm, not 20 ppm used by all other models and observations

5. Major Conclusion and Further Comments

In conclusion, we find that the general performance of the 17 models in this comparison with JR II field observations of chlorine clouds was good. The observations lie within the range of the model predictions, which cover about 1 ½ orders of magnitude for all trials and distances. Certainly no major biases or excessive scatter are found, and there is no trend in model performance with downwind distance.

This model comparison exercise focused on the dispersion models. The modelers were not asked to apply emissions models or meteorological models; instead, we provided emissions and meteorology inputs based on extensive research by the JR II scientific team. In particular, the tank observations and the rainout observations required analysis in order to optimize the emissions estimates. Even in a controlled research environment, it is tricky to estimate emissions for chlorine releases from holes in pressurized liquid tanks.

The deposition modules in codes could not be assessed, because, at the JR II desert site (a salt playa), chlorine deposition could not be measured. Based on fundamental chemistry, deposition to a dry desert such as at JR II was expected to be minimal. However, at a different site with vegetation and/or organic soil, the

effects of removal of chlorine by deposition are expected to be significant (e.g., removing half of the chlorine in the first 1 or 2 km travel distance). The DHS is currently sponsoring laboratory experiments aimed at developing better deposition estimates for different vegetation and soils.

We believe that the over-arching goal to facilitate assessment and improvement of all models used for analysis of hazardous gas releases is off to a very good start. We encourage continued collaborative studies involving experts in the field and dialog regarding lessons learned, improvement of models, and identification of the most important scientific gaps related to such modeling to aid plans for future research.

Acknowledgements - The research carried out by the first, second and third through sixth authors has been sponsored primarily by the U.S. Defense Threat Reduction Agency (DTRA). Other authors are funded by their own organizations.

References

Anfossi, D., Tinarelli, G., Trini Castelli, S., Nibart, M., Olry, C., Commanay, J., 2010. A new Lagrangian particle model for the simulation of dense gas dispersion. Atmospheric Environment, 44, 753-762.

Bauer, T.J., 2013. Comparison of chlorine and ammonia concentration field trial data with calculated results from a Gaussian atmospheric transport and dispersion model. Journal of Hazardous Materials, 254-255, 325-335.

Bauer, T.J., 2015. Software user's manual for the RAILCAR 4.1 toxic industrial chemical source characterization program. NSWCDD/TR-15/135, Dahlgren Division Naval Surface Warfare Center, Dahlgren, VA USA, 82 pp.

Bauer, T.J., Creekmore, S.I., Cunningham, A.R., Jones, A.M., Santos-Koos, L.M., 2018. QUIC-RAILCAR and ALOHA-RAILCAR model validation against the JR II field da-

754 ta set. NSWCDD/TR-18/xxx, Dahlgren Division Naval Surface Warfare Center,
755 Dahlgren, VA USA, 82 pp.
756 Björnham, O., Burman, J., 2016. PUMA Dense Gas Modelling. Report number: FOI-R--
757 4281—SE, Swedish Research Defence Agency, 164 90 Stockholm, Sweden, 35
758 pp.Bjornham O, Grahn H, Burman J, 2020. Comparison between the predictive results
759 of the two dispersion models PUMA and LPELLO with the JR II field data. AE Spe-
760 cial Issue on JR II
761 Chang, J, Hanna, S., 2004, Air quality model performance. Meteorol. and Atmos. Phys-
762 ics**,** vol 87, 167-196.
763 Chang, J., Mazzola, T., 2020. Analysis of observed concentrations from the sampler net-
764 work at JR II, for use in evaluating dispersion models. Atm. Environ. JR II Special Is-
765 sue.
766 DTRA, 2008: HPAC Hazard Prediction and Assessment Capability Version 5.0 SP1.
767 DTRA DVD, Defense Threat Reduction Agency, 8725 John J. Kingman Rd., MS
768 6210, Ft Belvoir VA.
769 Fabbri, L., Binda, M., Bruinen de Bruin, Y. 2017. Accident Damage Analysis Module
770 (ADAM) – Technical Guidance, Publications Office of the European Union, Luxem-
771 bourg, EUR 28732 EN, 2017, ISBN 978-92-79-71879-3, doi: 10.2760/719457.
772 Fabbri, L., Wood, M., 2019. Accident Damage Analysis Module (ADAM): Novel Euro-
773 pean Commission tool for consequence assessment—Scientific evaluation of perfor-
774 mance. Process Safety and Environmental Protection, Volume 129, September 2019,
775 249-263.
776 Fabbri L, Wood M, Azzini I, Rosati R, 2020. Global sensitivity analysis of the ADAM
777 dispersion model: Jack Rabbit II test case. Atm. Environ. JR II Special Issue.
778 Fox, S., Hanna, S., Mazzola, T., Spicer, T., Chang, J., Gant, S., 2020. Overview of Jack
779 Rabbit II (JR II) field experiment and summary of the methods used in the dispersion
780 model comparisons. Atm. Environ. JR II Special Issue.
781 Gant S, Tickle G, Kelsey A, Tucker H, 2020. DRIFT dispersion model predictions for the
782 Jack Rabbit II model inter-comparison exercise. Atm. Environ. JR II Special Issue on
783 JR II.

784 Gant, S., Weil, J., Delle Monache, L., McKenna, B., Garcia, M., Tickle, G., Tucker, H.,
785 Stewart, J., Kelsey, A., McGillivray, A., Batt, R., Witlox, H., Wardman, M., 2018.
786 Dense gas dispersion model development and testing for the Jack Rabbit II phase 1
787 chlorine release experiments. Atm. Environ. 192, 218-240.
788 Gomez F, Makke L, Ribstein B, Armand P, Nibart M., 2020. Simulations of a dense gas
789 model of a chlorine release with a Lagrangian Particle Dispersion Model (LPDM).
790 Atm. Environ. JR II Special Issue.
791 Hanna, S.R., 2018, Meteorological recommendations for input to dispersion models ap-
792 plied to JR II trials. Report P198-Met, Hanna Consultants, Kennebunkport, ME, 41 p.
793 Hanna, S.R., 2020. Meteorological data recommendations for input to dispersion models
794 applied to Jack Rabbit II trials. Atm. Environ. JR II Special Issue.
795 Hanna, S.R., 2020. B&M 1988 Workbook nomograms for dense gas modeling applied to
796 the JR II chlorine release trials. Atm. Environ. JR II Special Issue.
797 Hanna, S.R., Britter, R.E., Chang, J.C., Argenta, E., 2012, The Jack Rabbit chlorine re-
798 lease experiments: Implications of dense gas removal from a depression and downwind
799 concentrations. J. Haz. Mat., vol. 213-214, 406-412.
800 Hanna, S.R., Chang, J.C., 2012. Acceptance criteria for urban dispersion model evalua-
801 tion. Meteorol. and Atmos. Physics, 116 (No. 3), 133-146.
802 Hanna, S.R., Chang, J.C., Huq, P., 2016. Observed chlorine concentrations during Jack
803 Rabbit I and Lyme Bay field experiments. *Atm. Environ*., vol. 125, 252-256.
804 Hanna, S.R., Chang, J.C, Strimaitis, D.G., 1993. Hazardous gas model evaluation with
805 field observations. Atmos. Environ., 27A, 2265-2285.
806 Hanna, S.R., Dharmavaram, S., Zhang, J., Sykes, I., Witlox, H., Khajehnajafi, S., Koslan,
807 K., 2008, Comparison of six widely-used dense gas dispersion models for three recent
808 chlorine railcar accidents. Proc. Safety Prog., vol. 27, 248-259.
809 Havens, J., Spicer, T.O., 1990. LNG Vapor Dispersion Prediction with the DEGADIS
810 Dense Gas Dispersion Model. Report GRI 89/0242, Gas Research Institute (GRI).
811 Hoot, T.G., Meroney, R.N., 1974. The behavior of negatively buoyant stack gases.
812 CEP73-74RNM-TGH33, Presented at the 67th Annual meeting of the Air Pollution
813 Control Association, Denver, C), 9-13 June, 21 p.

814 Khajehnajafi, S., Pourdarvish, R., 2011, Correlations for mass transfer from a liquid spill:
815 comparisons and recommendations. Process Safety Progress, Vol. 30, No.2
816 Kukkonen, J., Nikmo, J, Riikonen, K, 2017. An improved version of the consequence
817 analysis model for chemical emergencies, ESCAPE. Atm. Environ. 150 198-209.
818 DOI: 10.1016/j.atmosenv.2016.11.050
819 Mazzola, T., Hanna, S., Spicer, T., 2019. Jack Rabbit II Coordinated Model Comparisons
820 to Data, Initial Specification for Modelers. Nine pages prepared by SAIC dated 19
821 April 2019.
822 Meel, A., Khajehnajafi, S., 2012, A comparative analysis of two approaches for pool
823 evaporation modeling: Shrinking vs. Non-shrinking pool area. Process Safety Progress,
824 Vol. 31, No.3.Nicholson, D., Hedrick, A., Lian, N., Schmidt, E.,. 2017, Final Test Re-
825 port for Jack Rabbit (JR) II. WDTC-SPD-FTR-001, US Army Dugway Proving
826 Ground, Dugway, UT, USA, 120 p.
827 NOAA/HMRAD and EPA/CEPPO, 1992. ALOHA User's Manual and Theoretical De-
828 scription. Reports available from NOAA/HMRAD, 7600 Sand Point Way NE, Seattle,
829 WA 98115 and on CAMEO/ALOHA web site.
830 Oldrini, O., Armand, P., Duchenne, C., Olry, C., Tinarelli, G., 2017. Description and
831 preliminary validation of the PMSS fast response parallel atmospheric flow and disper-
832 sion solver in complex built-up areas. Environ. Fluid Mech., 17 (3) 1-18.
833 SAFER Systems, 2014. Description of modeling algorithms, TRACE version 10.0., Help
834 File available upon request, 310 N Westlake Blvd #200, Westlake Village, CA 91362
835 Sigg, R., Lindgren, P., von Schoenberg, P., Persson, L., Burman, J., Grahn, H.,
836 Brännström, N., Björnham, O., 2018. Hazmat risk area assessment by atmospheric dis-
837 persion modelling using Latin hypercube sampling with weather ensemble. Meteorol.
838 Applic. 25, 575-585.
839 Simpson S, Miner S, Mazzola T, Meris R, 2020. HPAC model studies of selected Jack
840 Rabbit II releases and comparisons to test data. Atm. Environ. JR II Special Issue.
841 Spicer, T., Miller, D., 2017, Quantifying the mass release rate for flashing two phase re-
842 leases. Presented at Am. Inst. Chem. Eng. 2017 Spring Meeting, 13th Global Congress
843 on Process Safety, San Antonio, Texas, USA, March 26-29. 34 p.

844 Spicer, T.O., Tickle, G., 2020. Simplified Source Description for Atmospheric Dispersion
845 Model Comparison of the Jack Rabbit II Chlorine Field Experiments. Atm. Environ.
846 JR II Special Issue
847 Sykes, R.I., Parker, S.F., Henn, D.S., Chowdhury, B., 2016. *SCIPUFF Version 3.0 Tech-*
848 *nical Documentation*. Sage Management, 15 Roszel Road, Suite 102, Princeton, NJ,
849 403 pp.
850 Tinarelli, G., Mortarini, L., Trini Castelli, S., Carlino, G., Moussafir, J., Olry, C., Ar-
851 mand, P., Anfossi, D., 2012. Review and validation of Micro-Spray, a Lagrangian par-
852 ticle model of turbulent dispersion. In Lagrangian Modeling of the Atmosphere (Ed. J.
853 Lin, D. Brunner, C. Gerbig, A. Stohl, A. Luhar and P. Webley), Geophysical Mono-
854 graph, American Geophysical Union, 200, 311-327, Washington DC.
855 Thomson, D.J., 1987. Criteria for the selection of stochastic models of particle trajectories
856 in turbulent flows. J. Fluid Mech., 180. 529-556
857 Trini Castelli, S., Armand, P., Tinarelli, G., Duchenne, C., Nibart, M, 2018. Validation of
858 a Lagrangian particle dispersion model with wind tunnel and field experiments in ur-
859 ban environment. Atmos. Env., 193, 273-289.
860 VDI, 1990. Dispersion of Heavy Gas Emissions by Accidental Releases, Verein
861 Deutscher Ingenieure (VDI) 3783 Part 2, 44 pp.
862 Weil, J.C., 2019. Integral Dense-gas Dispersion Model (IDDM), NCAR report prepared
863 for Defense Threat Reduction Agency, Fort Belvoir, VA, 7 pp.
864 Weil, J.C., Alessandrini, S., 2020. Integral Dense-gas Dispersion Model (IDDM). Atm.
865 Environ. JR II Special Issue.
866 Wilson, J.D., Yee, E., Ek, N., D'Amours, R., 2009. Lagrangian simulation of wind
867 transport in the urban environment. Q. J. R. Meteorol. Soc., 135: 1586-1602.Witlox,
868 H.W.M., 2017. Theory Unified Dispersion Model. DNV GL-Software, 161 pp.
869 Witlox, H.W.M., Harper, M., 2014. Modeling of time- varying dispersion for releases
870 including potential rainout. Process Safety Progress 33(3), 265-273.
871 Witlox, H.W.M., Holt, A., 1999. A unified model for jet, heavy and passive dispersion
872 including droplet rainout and re-evaporation, International Conference and Workshop
873 on Modelling the Consequences of Accidental Releases of Hazardous Materials,
874 CCPS, San Francisco, California, September 28 – October 1, 315-344.