IDA | RESEARCH NOTES

THE WELCH AWARD 2013

6 PREDICTED AND MEASURED MULTIPATH IMPULSE RESPONSES

12 TRANSNATIONAL DRUG TRADE EXPANSION IN AFRICA

19 TEXT CLASSIFICATION FOR RETRIEVAL OF INFORMATION ON CRITICAL TECHNOLOGIES

25 HUMAN, SOCIAL, CULTURAL, AND BEHAVIORAL MODELING FOR MILITARY APPLICATIONS

31 POTENTIAL MILITARY CONFLICT BETWEEN NORTH AND SOUTH KOREA

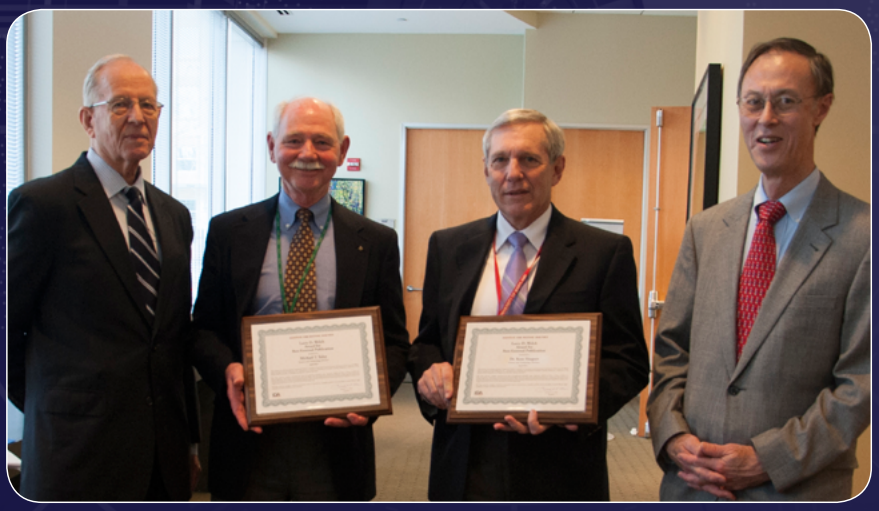
36 ESTIMATING HAZARDOUS RELEASES IN URBAN ENVIRONMENTS

42 ASSESSING COMBAT RISK AND COMPENSATION

47 CARBON FOOTPRINT OF SHALE GAS

53 EFFECTS OF SPACE DEBRIS ON SPACECRAFT

Winter 2013



Recipients of the Welch Award – authors Michael Tuley and Kent Haspert – flanked by General Larry D. Welch and Dr. David S. C. Chu, President of IDA.

IDA is the Institute for Defense Analyses, a non-profit corporation operating in the public interest.

IDA's three federally funded research and development centers provide objective analyses of national security issues and related national challenges, particularly those requiring extraordinary scientific and technical expertise.

IDA | THE WELCH AWARD

This issue of IDA *Research Notes* is dedicated to the 2013 winner and finalists of IDA's Larry D. Welch Award for best external publication. Named in honor of former IDA president and U.S. Air Force Chief of Staff, General Larry Welch, the award recognizes individuals who exemplify his high standards of analytic excellence through their publication in peer-reviewed journals or other professional publications, including books and monographs. The articles in this edition are executive summaries of the original published pieces. We have identified and credited the original publication for each article and, where possible, we have included a link and Quick Response (QR) code to the full piece.

Winner

Examining the issue of predicting multipath radar responses, **Kent Haspert** and **Michael Tuley** present a modified version of the original analytical approaches for evaluating multipath effects in the article "Comparison of Predicted and Measured Multipath Impulse Responses," published in the journal *IEEE Transactions on Aerospace and Electronic Systems*.

Finalist

Ashley Bybee examined "The Twenty-First Century Expansion of the Transnational Drug Trade in Africa," in the *Journal of International Affairs*, Fall/Winter 2012, Vol. 66, No. 1. She observed that not only has West Africa emerged over the last decade as a transit hub for Latin American drug trafficking organizations, the drug trade is expanding to other parts of the continent, with Africans assuming more proactive roles.

Finalist

The research team of **Arun Maiya**, **Francisco Loaiza-Lemos**, and **Robert Rolfe** authored "Supervised Learning in the Wild: Text Classification for Critical Technologies," for the 2012 Military Communications (MILCOM) Conference in Orlando, Florida. They studied how documents pertaining to critical technologies can be located and identified from among massive and heterogeneous collections of largely unimportant files on user workstations, laptops, and file servers.

Finalist

Sue Numrich and **Peter Picucci** wrote "New Challenges: Human, Social, Cultural and Behavioral Modeling," a chapter in an edited volume *Engineering Principles of Combat Modeling and Distributed Simulation*, published by John Wiley & Sons, Inc., 2012. The authors examine how human, social, cultural, and behavior (HSCB) elements are being introduced into the combat modeling community.

Finalist

Kongdan Oh Hassig and **Ralph Hassig's** article "Military Confrontation on the Korean Peninsula," appeared in *Joint Force Quarterly*, Issue 64, 1st Quarter 2012. The authors identify the inherent differences between the two societies and describe the danger of the military asymmetries.

The research team of **Nathan Platt**, **Steve Warner**, **Jeffry Urban**, and **James Heagy** published "Effects of Meteorological Data Thresholding on the Quality of Urban HPAC Predictions of the Joint Urban 2003 Field Trials," in the *International Journal of Environment and Pollution*, Vol. 44, Nos. 1/2/3/4, 2011. The authors analyzed ways to estimate the effects of hazardous releases on the underlying population, a necessary step for potential planning, emergency response, and recovery efforts.

Alexander Gallo, **Brandon Gould**, **Maggie Li**, **Shirley Liu**, and **Stanley Horowitz** (with Saul Pleeter, Curtis Simon, and Carl Witschonke) authored "Risk and Combat Compensation," Chapter 9 of the Supporting Research Papers for the Eleventh Quadrennial Review of Military Compensation (QRMC) Report, 2012. The authors assessed the incongruities and inequities of Service members' compensation in hazardous duty areas.

Christopher L. Weber and **Christopher Clavin** in "Life Cycle Carbon Footprint of Shale Gas: Review of Evidence and Implications" published in *Environmental Science & Technology*, 2012, Vol. 46, examined the implications of the production of natural gas from shale deposits on our energy outlook.

Joel Williamsen (with William Schonberg and Hilary J. Evens) authored "An Improved Prediction Model for Spacecraft Damage Following Orbital Debris Impact," an American Institute for Aeronautics Paper for AIAA Structures, Dynamics and Materials Conference, August 2012. In their paper, they identified new equations for predicting holes and cracks that would be produced following meteoroid or orbital debris penetration of spacecraft.

The winner of the 2013 Welch Award is Kent Haspert and Michael Tuley's paper published in *IEEE Transactions*, a summary of which is included in this *Research Notes*. The original article describes how the authors modified the WWII approach to multipath analysis to allow its predictions to be compared precisely to wideband measurements from the direct and reflected paths.

Multipath describes the phenomena that occur when a radar wave propagates to a target and back by more than one optical path. The coherent nature of radar means that these separate waves interfere with each other with consequent impact on both signal strength and wavefront shape. While some radar systems are able to exploit multipath, most of the time its effects are pernicious. A particularly vexing case arises from the problem of defending ships from sea-skimming anti-ship missiles, where multipath often leads to both delayed detection and erroneous tracking of the threat. That fact alone has prompted defense laboratories to engage in extensive studies of multipath effects since the first military applications of radar.

Appreciation of the fundamental physics underlying multipath effects dates to the 19th century, and the analytic theory was well developed during World War II. Nevertheless, it has not been trivial to devise and conduct experiments that provide sufficient precision to enable quantitative comparison between theory and measurement. In general, it has been simpler to find conditions that enable comparison in the *specular* case, that is, the case of reflections from a smooth mirror-like surface. Conducting experiments under *diffuse* multipath conditions (here think of the moon's glistening reflection in a rippled water surface) has been much more challenging. This situation has improved with the development of wide-band radar sources capable of resolving the small differences in transit time between two or more wave paths. Such a system was used to conduct a series of multipath experiments at the Atlantic Test Range in October 2008. One notable feature of these measurements, suggested by the IDA team, was the use of two separate narrow-beam antennas, one aimed directly at the target and the other aimed at the reflecting surface. This enabled the direct and multipath-reflected waves to be isolated in separate receiver channels. Despite this innovation, the analysis of the resulting data proved anything but straightforward. The quality of the collected data presented an opportunity for a precise validation of theory, but not before the IDA authors successfully accounted for several minute details of the experiment. Indeed, the sheer doggedness with which Kent and Mike pursued the data analysis demonstrated their confidence both in their methods and in the quality of the experimental data.

The multipath model described herein applies to both specular and diffuse multipath components, although the more difficult diffuse case was strongly favored in the particular experimental conditions. After accounting for intruding "real-world" factors in the data collection, IDA was able to show very close agreement between the predicted and observed multipath. This is not, however, the end of the effort. Partly because of the success of the work cited here, additional multipath measurements were undertaken to help the sponsoring government program office gain an even fuller understanding of the effect of multipath on their operational radar applications. In fact, the modified predictions validated in the current paper helped define the test conditions used in these subsequent measurements. Kent and Mike have now prepared a follow-on paper on these tests that compares experimental and theoretical data while considering both the magnitude and phase variations produced by multipath.

By Dr. James N. Ralston
IDA Research Staff Member

PREDICTED AND MEASURED MULTIPATH IMPULSE RESPONSES

Kent Haspert and Michael Tuley

The fundamental concepts used to evaluate multipath effects date back over 50 years. Today's technology can support wide-bandwidth communications and radar systems that were not available or considered when these multipath concepts were being formulated.

The following article summarizes a slightly modified version of the original analytical approaches for evaluating multipath effects and compares the predicted multipath to data collected from a wideband instrumentation radar. The multipath model presented here covers the specular (coherent) and the diffuse (noncoherent) components of multipath. The test data were collected for conditions that strongly favor diffuse multipath, but the experimental technique supported detection of any unanticipated specular contributions. Because the purpose of this validation effort was to perform an in-depth examination of multipath effects, the demanding test conditions revealed a couple of real-world effects that had to be addressed. After incorporating these effects into the analytical multipath formulations, IDA researchers were able to show very close agreement between the predicted and observed multipath.

The multipath model presented here covers the specular (coherent) and diffuse (noncoherent) components of multipath. The test data were collected for conditions that strongly favor diffuse multipath.

Multipath can degrade radio frequency (RF) transmission by adding unwanted reflected signals to the desired direct-path signal. The unwanted signals are delayed and will consequently have a different phase than the direct-path signal. Depending on the wavelength, geometry, and surface conditions, the total multipath signal can have an amplitude approaching that of the direct-path signal. Moreover, multipath can appear to come from a specular reflection point, from a diffuse glistening surface, or partly from both. For a stationary transmitter-receiver-reflecting surface geometry, the specular reflection will have a constant phase difference with respect to the direct-path signal. The diffuse reflection consists of a collection of signals of varying amplitudes and phases. The combined multipath returns can therefore add constructively or destructively with the direct-path signal to produce a stronger or weaker (or even vanishing) total received signal. When one or both of the platforms move, the received signal will fade in and out. This fading can affect RF communications and radar target detection/tracking.

MULTIPATH MODEL

For air-to-air communications, Figure 1 shows a direct path between the transmitter and the receiver and several other paths

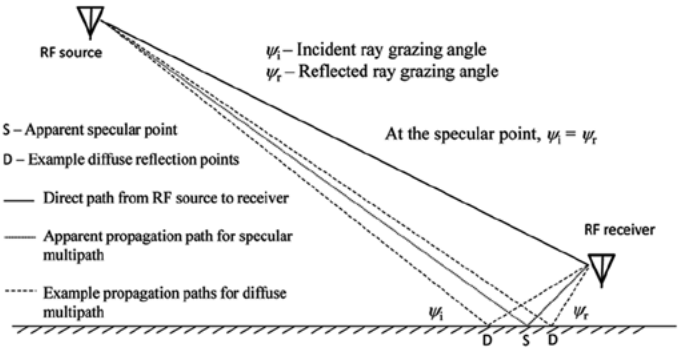


Figure 1. Multipath Geometry.

between the transmitter and the receiver that involve scattering of RF signals off the earth's surface.[1] The propagation paths involving scattering are longer than the direct path. The path from the transmitter to the specular point (point S in the figure) and then from the specular point to the receiver defines the minimum multipath distance. However, scattering can also occur from points on either side of the specular point (i.e., those points labeled D in the figure), and these multiple paths have a longer total path length.

Specular and Diffuse Multipath

Smooth, mirror-like surfaces produce specular multipath. Although the specular multipath appears to come from a single point, it actually involves coherent integration over a wide portion of the surface. Rough surfaces produce diffuse multipath. In this case, the reflected RF signals over the surface add noncoherently and typically appear weaker than specular multipath. Diffuse multipath appears to come from a sizable region, known as a glistening surface, and it is spread in path delay (Beckmann and Spizzichino 1987; Barton 2005, 279–290). While the peak amplitude of the diffuse multipath can sometimes be relatively weak, the total strength of the diffuse multipath can become signifcant after noncoherently combining the energy coming from a large glistening surface.

In 1953, Ament developed a formula that predicts the specular reflection coeffcient, ρ_s (Ament 1953, 142–146). This formula was subsequently modified in 1984 by Miller (Miller, Brown, and Vegh 1984, 114–116) and takes the following form:

$$Ps = 2\left(\frac{2\pi\sigma_h \sin(\psi)}{\lambda}\right)^2 \quad (1)$$

$$\rho_s = \exp\left[-Ps\right] I_0\left(Ps\right), \quad (2)$$

where σ_h is the root mean square surface height variation, ψ is the grazing angle, λ is the radar's wavelength, and $I_0(\,)$ represents the modified Bessel function of the first kind.

System Response for Specular Multipath

The following equation expresses the strength of the specular return relative to the strength of the direct-path return as a function of four factors (ρ_s, $\sqrt{G_{ant}}$, $|\Gamma_p|$, and ρ_{veg}):

$$\frac{\text{Specular voltage}}{\text{Direct path voltage}} = \rho_s\sqrt{G_{ant}}|\Gamma_p|\rho_{veg}, \quad (3)$$

where ρ_s comes from Eq. (2), $\sqrt{G_{ant}}$ is the ratio of antenna gains between the specular and the direct paths, Γ_p is the Fresnel reflection coefficient,[2] and ρ_{veg}

[1] Due to reciprocity, the locations of the transmitter and receiver can be interchanged without affecting the theory.

[2] The Fresnel reflection coefficient is a complex value that is a function of the surface conductivity and dielectric constant, the wavelength, the grazing angle, and the polarization. The subscript *p* denotes the polarization.

represents a vegetation absorption factor. The system impulse response consists of a unit impulse at the time of arrival of the direct path pulse and a second pulse with a magnitude given by Eq. (3) delayed according to the following formula:

$$\text{Specular range delay} = R_1 + R_2 - R, \quad (4)$$

where R_1 is the distance from the transmitter to the specular point S in Figure 1, R_2 is the distance from the specular point to the receiver, and R is the direct-path distance between the transmitter and the receiver.

Glistening Surface and Diffuse Multipath

A rough surface produces scattering, but the surface roughness makes the phase relationships among the reflections across the surface unpredictable (or at least infeasible to address in anything other than a statistical sense). Barton extended the rough surface formulation of Beckmann and Spizzichino, and this extension provided the basis for formulating a single equation that describes the diffuse multipath return from an arbitrary small patch of area dA lying somewhere on the glistening surface. The last term in Eq. (5) differs from the one provided by Barton to incorporate a more rigorously derived representation of the surface shadowing that can occur at low grazing angles. This shadowing reduces the effective area for accumulating multipath diffuse energy. The resulting equation becomes

$$\frac{\text{Diffuse voltage from } dA}{\text{Direct path voltage}} = \sqrt{\frac{1}{4\pi}\left(\frac{R}{R_1 R_2}\right)^2 \frac{1}{\beta_0^2}\exp\left(-\frac{\beta^2}{\beta_0^2}\right) dA} \cdot |\Gamma_v| \cdot \rho_{veg} \cdot \sqrt{G_{ant}} \cdot \rho_{roughness}\sqrt{S_f}, \quad (5)$$

where:

$\frac{1}{4\pi}\left(\frac{R}{R_1 R_2}\right)^2$ is the one-way spreading loss;

$\frac{1}{\beta_0^2}\exp\left(-\frac{\beta^2}{\beta_0^2}\right)$ is the expected bistatic radar cross section per unit area (σ_h) of the diffuse patch defined by dA, where β_0 is the mean square value of the surface slope over the region of interest and β is the angle between the bisector of the R_1 and R_2 rays and the local vertical; $\rho_{roughness}$ is the roughness factor (potential maximum diffuse return multiplier); and

$\sqrt{S_f}$ is the shadowing factor (probability that dA is actually seen by both the transmitter and receiver) based on previous research (Smith 1967, 668–671).

The antenna factor term in Eq. (5) is different from the specular antenna factor in that it varies over the glistening surface, as do most of the other terms in this equation.

Impulse Response Curve for Diffuse Multipath

Figure 2 shows an example of a one-way impulse response curve (IRC) for the combination of the direct path, the total (specular and diffuse) multipath, and the diffuse-only multipath. The first portion of the diffuse-return IRC combines the energies associated with those portions of the glistening surface that have a path delay within one radar range-resolution bin of the specular path delay. This example assumes

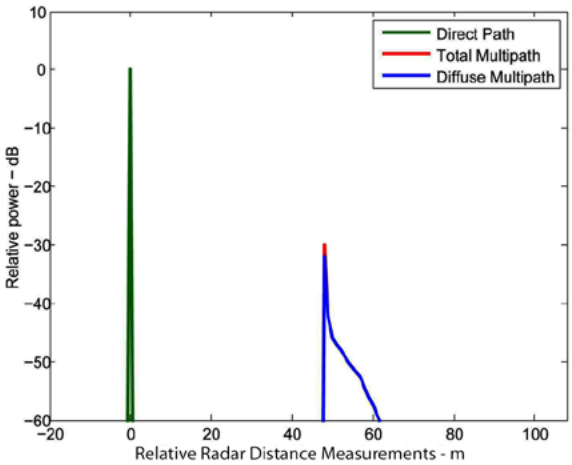


Figure 2. Example IRC.

that the specular energy is about 25% of the diffuse energy occurring in the same range bin. Additional portions of the glistening surface have longer path delays, and they produce a one-way diffuse IRC that results in a series of diminishing response values. The figure shows only a slight addition to the diffuse multipath due to the addition of the specular component.

VALIDATION TESTING

The original multipath formulations that underlie the methodology did not explicitly consider their applicability to wide-bandwidth systems. Therefore, this effort addresses the foregoing theoretical formulations in the context of wideband RF systems. For this validation effort, we chose to use a wideband instrumentation radar observing a point scatterer (i.e., a calibration sphere), with a relatively rough surface between the radar and the point scatterer, and then directly observing the one-way multipath IRC.

Atlantic Test Range Experimental Configuration

The Atlantic Test Range experiments used two nominally identical high-gain X-band radar antennas sitting near the coastline of the Chesapeake Bay and a tethered sphere a few miles off shore. One of the antennas tracked the tethered sphere and bounced RF energy off it. The other antenna pointed downward and observed the glistening surface. The downward-looking antenna only received radar signals (i.e., it did not transmit) but had its local oscillator synchronized with the upward-looking antenna so it could coherently process the received multipath returns. Figure 3 illustrates the test configuration. Because the calibration sphere appears as a point source, the experiment directly measured the one-way IRC over the sphere-surface-receiver path. The relatively short ranges and low grazing angles generate an IRC with relatively little range delay, which necessitated the use of a wide-bandwidth radar to resolve the structure of the IRC.

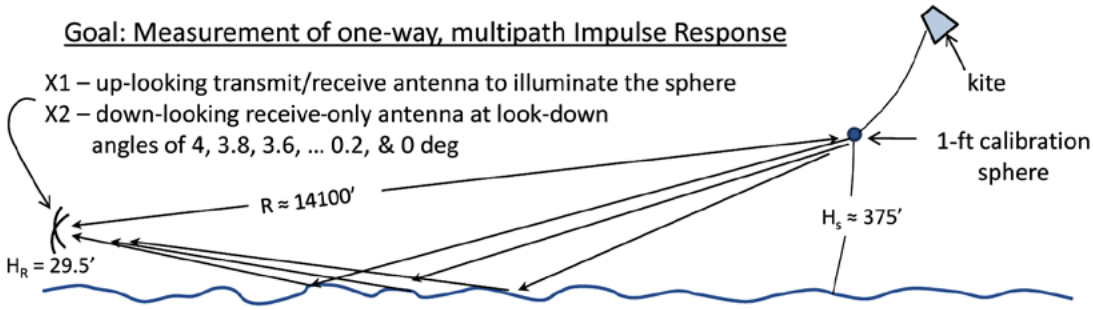


Figure 3. Atlantic Test Range Test Configuration.

During the test, the winds varied between 20 and 30 knots out of the north. These winds created approximately 3.5-foot waves with a period of about 3 sec. The high-gain, approximately 10-foot diameter radar antennas, have 3-dB beamwidths of about 0.66º. The upward-looking transmitting and receiving antenna, designated X1, tracked the tethered sphere as the winds blew it around. The sphere's height varied from about 310 to 475 foot, and the range varied from about 13,900 to 14,200 ft. The downward-looking, receive-only antenna, designated X2, stepped through 21 look-down angles between 4º and 0º to scan the glistening surface. The downward-looking antenna recorded data for 30 second at each look-down step and ran through this range of look-down angles twice. The test conditions produced $\sigma_h \approx 27$ cm and $\beta_0 \approx 0.055$ radians.

Data-Analysis Procedures

The test conditions suggest that the diffuse multipath should totally dominate the specular multipath.[3] However, we wanted to validate this theoretical prediction by analyzing the data in such a way that any specular multipath significantly above the predicted value would become readily apparent. Also, the diffuse multipath model assumes a statistically random Gaussian surface. During the surface's approximately 3-second wave period, the diffuse glistening surface should appear random, in accordance with the Gaussian assumption inherent in the diffuse multipath formulation. Therefore, we needed to use relatively long coherent processing intervals, which therefore meant that we had to perform motion compensation to remove the effects of the winds moving the tethered sphere during the data measurement intervals.

RESULTS

Figure 4 shows the initial comparison of the measured and predicted IRCs after convolving the predicted multipath IRC with the inherent IRC of the measurement radar. The solid line indicates the predicted IRC, and the dashed line represents the measured IRC. While these results generally look good for positive values of relative range, the figure shows significant energy in the negative-axis portions of the measured data. This early arriving energy seemed to violate the basic principle of the speed-of-light limit for RF propagation. However, upon closer investigation, we discovered that the apparent discrepancy was most likely caused by some return energy diffracting off the edges of the Cassegrain antenna's subreflector. Figure 5 illustrates this "short-circuit" path. After accounting for this effect and combining all of the measurement data, the result shown

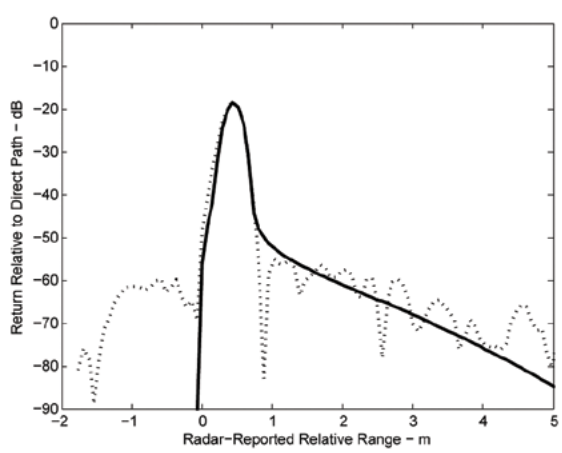


Figure 4. Initial Impulse Response Result.

[3] With $\rho_S = 0.1787$, the specular power, which is related to the square of this value, becomes about 3% of the diffuse power.

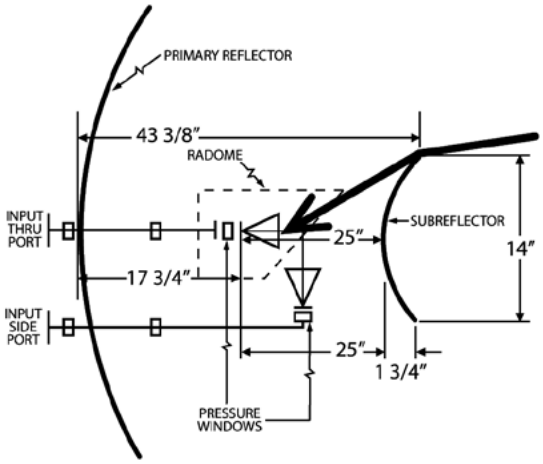


Figure 5. Antenna Diffraction Path.

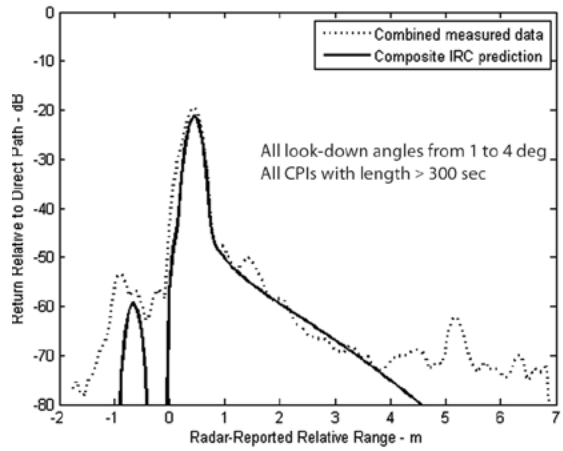


Figure 6. Combined Theoretical and Observed IRCs.

in Figure 6 was obtained, thereby demonstrating that the basic multipath formulations appear correct even when extended to wideband applications.

Dr. Haspert holds a Ph.D. in Electrical Engineering from the University of Maryland.

Mr. Tuley hold an M.S. in Electrical Engineering from the Georgia Institute of Technology.

The full article was published in the journal *IEEE Transactions on Aerospace and Electronic Systems.*

AN APPROACH FOR EVALUATING MULTIPATH EFFECTS

http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5937259&isnumber=5937244

REFERENCES

Ament, W. S. 1953. "Toward a Theory of Reflection by a Rough Surface." *Proceedings of the IRE* 41 (1) (January): 142–146.

Barton, David K. 2005. *Radar System Analysis and Modeling.* Norwood, MA: Artech House.

Beckmann, Petr, and Andre Spizzichino. 1987. *The Scattering of Electromagnetic Waves from Rough Surfaces.* Norwood, MA: Artech House.

Miller, A. R., R. M. Brown, and E. Vegh. 1984. "New Derivation for the Rough-Surface Reflection Coefficient and for the Distribution of Sea-Wave Elevations." *IEE Proceedings H* 131 (2) (April): 114–116.

Smith, B. 1967. "Geometrical Shadowing of a Random Rough Surface." *IEEE Transactions on Antennas and Propagation* 15 (5) (September): 668–671.

TRANSNATIONAL DRUG TRADE EXPANSION IN AFRICA

Ashley N. Bybee

In the last decade, West Africa emerged as a major transit hub for Latin American Drug Trafficking Organizations (DTOs) transporting cocaine to Western Europe. Sadly, the fears of many observers have been confirmed as the insidious effects of the drug trade have begun to take effect in many West African states. Consumption is on the rise, and narco-corruption now undermines the rule of law and legitimate economic growth necessary for development and stability. One of the most alarming trends is the number of new fronts on which the illicit drug trade is growing. Its geographic expansion beyond the relatively confined region of West Africa is now endangering East and Southern Africa. The arrival of new drugs to the region—heroin and Amphetamine-Type Stimulants (ATS, commonly referred to as synthetic drugs)—has been accompanied by the discovery of local manufacturing facilities to process them. Finally, the growing level of involvement by Africans—who initially served as facilitators but now appear to be taking a more proactive role—raises concerns that a new generation of African DTOs is rising in the ranks.

BACKGROUND

Africa's geographic proximity between the source zone and final markets for many illicit goods has contributed to its exploitation by Drug Trafficking Organizations (DTOs) (Mazzitelli 2007, 1075–1084). In the last decade, the emergence of West Africa as a major transit point for Latin American cocaine en route to Europe has taken the international community by surprise (*Confronting Drug Trafficking in West Africa* 2009; "Spanish Drug Enforcement Chief …" 2012). Europe is now experiencing increasing levels of cocaine entering its borders from West Africa and, from its experience as both a transit zone and a final market for heroin, is all too familiar with the public health and safety issues associated with addiction, crime, prostitution, and violence that typically accompany drug trade. The United States fears the existence of several anti-American terrorist groups operating in Africa fueled with drug revenues that could cultivate a crime-terrorist nexus, thus threatening its interests in yet another region (Wechsler 2012). Furthermore, since the DTOs operating in West Africa also traffic drugs into the United States, the Americans view the proceeds of these Africa-based operations as bolstering

the strength of those DTOs operating on its southern border with Mexico (*Transnational Drug Enterprises* … 2009). All parties are concerned with the destabilizing impacts of the drug trade (including its potential to finance local insurgencies), the expansion into new illicit markets (such as arms), and the undermining effect on programs that improve governance, anticorruption, and accountability, and raise public trust in government institutions.[1]

Today, these concerns are more relevant than ever, with new drugs, in greater quantities, and an ever-increasing number of transit states now characterizing Africa's drug trade. Whereas Africa's involvement in the global drug trade before the mid-2000s was generally limited to West African heroin distribution networks, the last several years have witnessed an unrelenting expansion of the drug trade throughout the continent.

MORE DRUGS, NEW DRUGS, AND NEW TRANSIT STATES

Historically, three factors facilitated African involvement in the heroin trade: the existing smuggling routes in a resource-rich continent, global economic downturns that left many Africans impoverished and jobless, and the massive African (primarily Nigerian) diaspora with a presence in 80 countries around the world (Akyeampong 2005, 429–447; Wannenburg 2005, 5–16). Although these roots of African involvement in the international drug trade are well documented, what has not been adequately documented is the 21st century expansion of the drug trade in Africa, aside from a few notable efforts (Gastrow 2011). The drug trade has expanded geographically across the continent, necessitating the participation of new African and transnational organized crime groups. The number of new drugs now found in Africa represents yet another front for the expansion of drug trade in Africa.

Cocaine

In 2004, the United Nations Office on Drugs and Crime (UNODC) first recorded a significant increase in cocaine seizures in West Africa (United Nations Office on Drugs and Crime 2005, 71). By 2006, the issue had reached such an alarming level that the UNODC Executive Director Antonio Maria Costa characterized the region as "under attack" by these DTOs. By 2008, the UNODC reported that 35 percent of cocaine shipped to Europe from Latin America transited West Africa (United Nations Office on Drugs and Crime 2008). As a result, reports, commentaries, and opinions that implied a potential proliferation of "narco-states" in West Africa began to emerge (Einarsdóttir 2007; Vulliamy 2008; Sourt 2009).

Since 2008, the West African route for trafficking cocaine to Europe began to decline following the international attention focused on the region after

[1] Non-attribution interviews with the author, November 2009 to November 2010. Throughout this article, the insights and perspectives of various individuals interviewed by the author while conducting field research are presented as evidence to support arguments and explain certain trends. Due to the sensitive nature of this topic, they requested that their names be withheld from any published material.

the assassinations of two high-level government officials in Guinea-Bissau (United Nations Office on Drugs and Crime 2010, 242). However, other regions, in particular East Africa, have reported increases in cocaine seizures. The UNODC reported a fourfold increase in cocaine seizures in East Africa in 2009/2010 as compared to the 2005/2006 levels (United Nations Office on Drugs and Crime 2012, 40). Cocaine is also turning up more in Southern Africa, where it has rarely been seen before.

Despite this adoption of alternate cocaine trafficking routes, one should not assume that the Western route has been abandoned permanently. It is highly likely that many traffickers, whose operations had previously been entirely dependent on support from state officials, are lying low while they evaluate their position with the new political elite and the effectiveness of counternarcotics programs in the region.[2]

Heroin

The last 5 years have witnessed a distinct rise in heroin seizures across Africa (United Nations Office on Drugs and Crime 2012, 2). East and West Africa have long been used as staging grounds for Southwest Asian heroin en route to Europe and the United States. West African—and especially Nigerian—criminal organizations have historically been deeply involved in this illicit trade and continue to be the most active African heroin traffickers today (Ellis 2009). Yet, the amounts transiting West, East, and even Southern Africa in recent years are far greater than those that have been observed in previous decades, and the countries now reporting significant seizures are countries where heroin was previously uncommon. For example, increased heroin seizures have been reported in Egypt (from 159 kg in 2009 to 234 kg in 2010), in Kenya (from 8.5 kg to 35 kg), in Nigeria (from 104 kg to 202 kg), and in Tanzania (from 7.9 kg to 191 kg) (United Nations Office on Drugs and Crime 2012, 34). The UNODC attributes this rise in heroin seizures to the traffickers' needs to seek out alternate routes to avoid detection by law enforcement and suggests that heroin markets are expanding in this region. Further, academic research appears to corroborate this claim, noting the mounting use of heroin and other injectable drugs in Kenya, Mauritius, Mozambique, Seychelles, and Tanzania.

AMPHETAMINE-TYPE STIMULANTS (ATS) AND PRECURSOR CHEMICALS

The most recent trend in drug trafficking through Africa has been the significant uptick in the transshipment of ATS, including precursor chemicals, since late 2009. The U.S. government, echoed by the UNODC, noted an alarming rise in the importation of precursor chemicals that appears to exceed the legitimate demand of African countries, even considering the growth of the pharmaceutical industry in the region (Wechsler 2012). The U.S. government and UNODC believe that precursor chemicals (pseudoephedrine and ephedrine) are imported into the region for the illicit production of methamphetamine and then exported to Far East Asian markets, most notably Japan and the Koreas, after transiting Southeast Asian countries such as Malaysia (Wechsler 2012; United Nations Office on Drugs

[2] Non-attribution interview with the author, Dakar, Senegal, November 2009.

and Crime 2012, 57, 80). This illicit trade is believed to be controlled by Mexican DTOs who initially established ties with local criminal groups in Mozambique, the Democratic Republic of the Congo, Ghana, and Nigeria (*Countering Narcotics Threats in West Africa* (Harrigan) 2012). In the last few years, these DTOs have found additional partners in Cape Verde, Egypt, and Kenya (*Countering Narcotics Threats in West Africa* (Harrigan) 2012; United Nations Office on Drugs and Crime 2012, 57). Perhaps the most worrisome development on this front was the discovery of several operational meth labs in Nigeria and South Africa in 2011 and 2012, revealing that Africa has advanced beyond simply a transit hub and is now an active producer of illegal drugs (*Countering Narcotics Threats in West Africa* (Harrigan) 2012).

IMPACTS ON STATES AND POPULATIONS

The impacts of this growing drug trade on African states and their populations are beginning to manifest themselves. Consumption is a growing concern. The UNODC estimates that more than 2.5 million people are drug users in West and Central Africa (*Countering Narcotics Threats in West Africa* (Carson) 2012). The corrosive effect that narco-corruption has on government officials and nascent democratic institutions is also alarming (Gastrow 2011, 8). The concern that drug revenues could finance international terrorism in Africa has already been realized. Terrorist organizations often have ties to illicit revenue streams, such as those provided by drugs.

One former senior official from the Drug Enforcement Administration (DEA) stated, "Of the 43 Foreign Terrorist Organizations listed by the State Department … 19 have clearly established ties to drug trafficking, and many more are suspected of having such ties" (Braun 2008). Given that Hezbollah, Hamas, and Al-Qaeda in the Islamic Maghreb (AQIM) are known to have an operational presence in Africa, one can see why this issue has become a concern to many Western governments. Several high-profile law enforcement investigations have revealed the involvement of the Fuerzas Armadas Revolucionarias de Colombia (FARC), Hezbollah, and AQIM in facilitating the movement of drugs through Africa.

In one operation, the DEA revealed a bizarre and complex trade-based money laundering scheme; the proceeds of drugs sold in Europe (which had transited West Africa) were mixed with legitimate used-car sale profits in Africa and sent to a Hezbollah-affiliated bank through exchange houses. Cocaine moved from South America to Europe and the Middle East through West Africa. Drug proceeds were remitted back to the United States through the Lebanese Canadian Bank (LCB), which is known to be a financial institution used for money laundering with links to Hezbollah and Iran. These proceeds were washed by mixing them with the proceeds of used cars bought in the United States and sold in Africa, where sale proceeds were also wired to the LCB (DEA News 2011; Rotella 2011). Thus, the role West Africa plays in the international drug trade is incontrovertibly funding some terrorism.

LOOKING FORWARD

Looking ahead, one can assume that the same conditions that gave rise to the 21st century expansion of the drug trade in Africa will continue to invite such illicit activity in the years to come. Given the counternarcotics community's main focus on opiates produced in Southwest Asia and the preponderance of resources focused on that region, the African route will likely grow unless considerable counternarcotics resources are invested in the continent. That is not to say, however, that the international law enforcement community has ignored this problem. The UNODC, the Economic Community of West African States (ECOWAS), and an array of international donors have made great strides in acknowledging the growing problem of drug trafficking and have implemented practical measures to stem the flow of illegal drugs. Nonetheless, as African subregional organizations further integrate economically and create structures such as customs unions in the East African Community, licit and illicit economies can be expected to become further intertwined, dependent on each other, and therefore stronger.

Unfortunately, the experiences of other transit states, such as Jamaica, Tajikistan, Turkey, and Mexico, demonstrate that, over time, the deleterious effects of the drug trade will not diminish. Consumption, addiction, and drug-funded insurgencies will become commonplace. The overarching concern of many, however, is the growth of indigenous African DTOs. DEA investigations and UNODC data reveal a growing number of West Africans involved in cocaine and heroin trafficking in regions where they were not common previously. For example, the UNODC reports that Nigerian groups have become active in exporting cocaine from Brazil to destinations in Africa and Europe (United Nations Office on Drugs and Crime 2012, 84). In July 2011, a DEA investigation culminated in the arrest of several Ghanaian nationals at Dulles International Airport for smuggling in heroin destined for New York City ("Members of Ghana-based Heroin Smuggling Ring …" 2011). In January 2012, a Tanzanian heroin trafficker was arrested in Mozambique en route to South Africa (*Agencia de Informacao de Mocambique* 2012). In 2010, a Mozambican businessman, Mohamed Bachir Suleman, was even added to the U.S. Treasury Department's list of "drug kingpins," along with several other West Africans ("Treasury Sanctions Entities Owned …" 2010). Additional examples abound, all of which demonstrate that international criminal groups are cultivating deeper ties with indigenous African criminal networks, who are taking on a more active role in the trade. Much the way that Mexican DTOs replaced many of the functions of Colombian DTOs in the 1990s, it appears that African DTOs could be taking on many of the functions initially carried out by Latin Americans and enjoying a more central role, along with decision-making power, in West African operations.

Dr. Bybee holds a Ph.D. in Public Policy (International Security) from the School of Public Policy, George Mason University.

The full article was published in the *Journal of International Affairs* 66 (1) (Fall/Winter 2012): 69–86.

THE TWENTY-FIRST CENTURY EXPANSION OF THE TRANSNATIONAL DRUG TRADE IN AFRICA

http://jia.sipa.columbia.edu/twenty-first-century-expansion-transnational-drug-trade-africa

REFERENCES

Agencia de Informacao de Mocambique. 2012. "Mozambique: Tanzanian Drug Trafficker Detained," January 9. http://allafrica.com/stories/201201100240.html.

Akyeampong, Emmanuel. 2005. "Diaspora and Drug Trafficking in West Africa: A Case Study of Ghana." *African Affairs* 104 (416): 429–447. http://users.polisci.wisc.edu/schatzberg/ps362/Akyeampong2005.pdf.

Beyrer, Chris, Kasia Malinowska-Sempruch, Adeeba Kamarulzaman, Michel Kazatchkine, Michel Sidibe, Stephanie A. Strathdee. 2010. "Time to Act: A Call for Comprehensive Responses to HIV in People Who Use Drugs." *Lancet* 376 (9740): 551–553.

Braun, Michael. 2008. "Drug Trafficking and Middle Eastern Terrorist Groups: A Growing Nexus" (speech). Washington DC: Washington Institute for Near East Policy, July 25. http://www.washingtoninstitute.org/policy-analysis/view/drug-trafficking-and-middle-eastern-terrorist-groups-a-growing-nexus.

Confronting Drug Trafficking in West Africa. 2009. *Before the U.S. Senate Committee on Foreign Relations Subcommittee on African Affairs*, 111th Cong. (statement of Thomas Harrigan, Assistant Administrator and Chief of Operations, Operations Division, Drug Enforcement Administration, June 23). http://www.foreign.senate.gov/imo/media/doc/HarriganTestimony090623a1.pdf.

Countering Narcotics Threats in West Africa. 2012. *Before the Senate Caucus on International Narcotics Control*, 112th Cong. (statement of Johnnie Carson, Assistant Secretary of State for African Affairs, May 6). http://www.drugcaucus.senate.gov/hearing-5-16-12/Amb%20Carson%20Testimony.pdf.

Countering Narcotics Threats in West Africa. 2012. *Before the Senate Caucus on International Narcotics Control*, 112th Cong. (statement of the Honorable Thomas Harrigan, Deputy Administrator, Drug Enforcement Administration, May 6). http://www.justice.gov/dea/pr/speeches-testimony/2012-2009/051612_testimony.pdf.

Countering Narcotics Threats in West Africa. 2012. *Before the Senate Caucus on International Narcotics Control*, 112th Cong. (statement of William Wechlser, Deputy Assistant Secretary of Defense, Counternarcotics and Global Threats, May 6). http://www.drugcaucus.senate.gov/hearing-5-16-12/DASD%20Wechsler%20Statement.pdf.

"DEA News: Civil Suit Exposes Lebanese Money Laundering Scheme for Hizballah." 2011. *DEA Press Release*, December 15. http://www.justice.gov/dea/pubs/pressrel/pr121511.html.

Einarsdóttir, Jónína. 2007. "Partnership and Post-War Guinea-Bissau." *African Journal of International Affairs* 10 (1 and 2): 93–112.

Ellis, Stephen. 2009. "West Africa's International Drug Trade." *African Affairs* 108 (431): 171–196.

Gastrow, Peter. 2011. *Termites at Work: Transnational Organized Crime and State Erosion in Kenya*. New York: International Peace Institute.

Mazzitelli, Antonio. 2007. "Transnational Organized Crime in West Africa: The Additional Challenge." *International Affairs*, 83 (6): 1071–1090. http://onlinelibrary.wiley.com/doi/10.1111/j.1468-2346.2007.00674.x/pdf.

"Members of Ghana-based Heroin Smuggling Ring Charged International Traffickers Used Dulles International Airport as Entry Point for East Coast Distribution of Heroin from West Africa." 2011. *Immigrations and Customs Enforcement* (ICE) Press Release, July 14. http://www.ice.gov/news/releases/1107/110714alexandria.htm.

Rotella, Sebastian. 2011. "Government Says Hezbollah Profits from U.S. Cocaine Market Via Link to Mexican Cartel." *ProPublica*, December 13. http://www.propublica.org/article/government-says-hezbollah-profits-from-us-cocaine-market-via-link-to-mexica.

Sourt, Caroline. 2009. "Africa's First Narco-State?" *The Guardian*, March 9.

"Spanish Drug Enforcement Chief: Africa a Preferred Route for SouthAm Cocaine." 2012. *Fox News Latino*, September 9. http://latino.foxnews.com/latino/news/2012/09/09/spanish-drug-enforcement-chief-africa-preferred-route-for-southam-cocaine/.

Transnational Drug Enterprises: Threats to Global Stability and U.S. National Security from Southwest Asia, Latin America, and West Africa. 2009. *Before the U.S. House Committee on Oversight and Government Reform Subcommittee on National Security and Foreign Affairs*, 111th Cong. (statement of Douglas Farah, Senior Fellow at the International Assessment and Strategy Center, October 1).

"Treasury Sanctions Entities Owned by Drug Kingpin Mohamed Bachir Suleman Treasury Action Targets Narcotics Trafficking Network in Mozambique, Builds on President Obama's Drug Kingpin Identification." 2010. *U.S. Department of Treasury Press Release*, June 1. http://www.treasury.gov/press-center/press-releases/Pages/tg729.aspx.

United Nations Office on Drugs and Crime. 2005. *World Drug Report 2005*. Vienna, Austria: United Nations Publications. http://www.unodc.org/pdf/WDR_2005/volume_1_web.pdf.

United Nations Office on Drugs and Crime. 2008 *World Drug Report 2008*. Vienna, Austria: United Nations Publications. http://www.unodc.org/documents/wdr/WDR_2008/WDR_2008_eng_web.pdf.

United Nations Office on Drugs and Crime. 2010. *World Drug Report 2010*. Vienna, Austria: United Nations Publications. http://www.unodc.org/documents/wdr/WDR_2010/World_Drug_Report_2010_lo-res.pdf.

United Nations Office on Drugs and Crime. 2012. *World Drug Report 2012*. Vienna, Austria: United Nations Publications. http://www.unodc.org/documents/data-and-analysis/WDR2012/WDR_2012_web_small.pdf.

Vulliamy, Ed. 2009. "How a Tiny West African Country Became the World's First Narco State." *The Guardian Observer*, March 9. http://www.guardian.co.uk/world/2008/mar/09/drugstrade.

Wannenburg, Gail. 2005. "Organized Crime in West Africa." *African Security Review* 14 (4): 5–16. http://www.issafrica.org/topics/organised-crime/01-dec-2005-organised-crime-in-west-africa-gail-wannenburg.

TEXT CLASSIFICATION FOR RETRIEVAL OF INFORMATION ON CRITICAL TECHNOLOGIES

Arun S. Maiya, Francisco Loaiza-Lemos, and Robert M. Rolfe

The research effort described here addresses the problem of locating documents pertaining to critical technologies (e.g., restricted, proprietary, or sensitive technical information) from among a massive and highly heterogeneous collection of largely unimportant files. The authors' solution is a system that uses supervised machine learning to detect such critical documents. To address difficult or ambiguous instances, the text classifier is supplemented with an automated keyword search (i.e., discriminative terms (keywords) are extracted in an automated fashion from the training set and matched against documents during the classification process). The effectiveness of this hybrid approach is demonstrated through a series of validation tests and case studies.

How does one locate and identify documents pertaining to critical technologies from among a massive and highly heterogeneous collection of largely unimportant files? We refer to such sensitive documents as *critical documents*—documents that may contain restricted or proprietary information. (Documents relevant to a critical technology of interest are referred to as *positive*, and documents unrelated to the critical technology are referred to as *negative*.) We explore the extent to which we can detect critical (or positive) documents on user workstations, laptops, and file servers (an environment that we refer to as "in the wild," given the high level of heterogeneity among files on workstation hard drives). Such a capability has applications to several areas, including consequence management of cyber security breaches. So, how does one go about finding such "needles in haystacks" or even determining whether any "needles" exist on a particular system?

THE TROUBLE WITH KEYWORD SEARCHES

The standard approach to making such a determination is to use *keyword searches* (or queries). The documents returned by the search are then manually reviewed by a human to determine and confirm their relevancy to a given critical topic.

Such keyword searches to locate critical documents pose several challenges. For instance, the optimal set of keywords needed to home in on a particular set of critical documents can be difficult to determine—even for subject matter experts (SMEs). Use of an ill-chosen keyword set can result in the omission of critical documents (i.e., false negatives) and the

inclusion of non-relevant documents (i.e., false positives). Moreover, functions based largely on the number of keyword hits are not always good indicators of a document's richness of content with respect to a particular critical technology.

SUPERVISED MACHINE LEARNING

To investigate the use of *supervised machine learning* to detect and locate critical documents, we employ the use of Linear Support Vector Machines (LinearSVMs) as our main learning algorithm (Fan et al. 2008). Beginning with a set of labeled documents (or training set)

$$D = \{(d_1, y_1), \ldots, (d_n y_n)\},$$

we will build our document classifier, where y_1 is the topic label of document d_1 for all i.[1] Each document in the training set D is represented as a bag of words (i.e., a multi-set structure). To use machine learning, we must transform each document

$$d_i \in \{d_i\}_{i=1}^{n}$$

into a *feature vector* $\vec{x}_1$ for use in machine learning. Each component of the vector represents a term (e.g., word) from the vocabulary of D, and the component value is assigned using a well-studied weighting scheme referred to as TF-IDF [term frequency–inverse document frequency]. More information on such representations can be found in Manning, Raghavan, and Schütze 2008 and Wu et al. 2008. A LinearSVM algorithm accepts

$$\widehat{D} = \{(\vec{x}_1 y_1), \ldots, (\vec{x}_n y_n)\}$$

as input. A decision boundary (or hyperplane) that best divides the positive documents from the negative documents is sought—a process called *training* or *learning*—as shown in Figure 1. New documents supplied to the algorithm are plotted in this space and classified as positive or negative based on which side of the hyperplane they fall. The distance of a new document from the hyperplane can be employed by the classifier as a score of confidence in the assignment of documents to the positive (or negative) category.

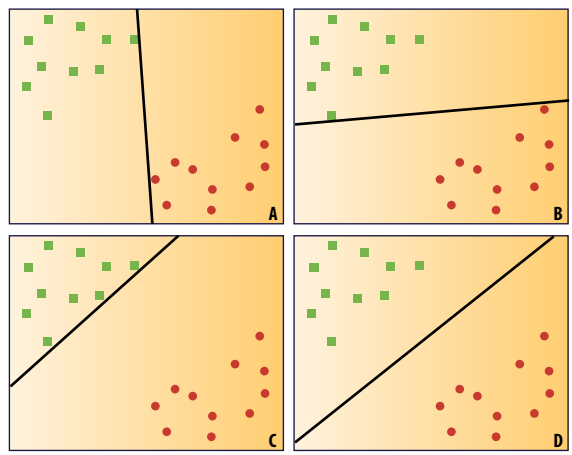


Figure 1. SVM Classifiers Are Essentially Linear Separators That Best Divide Existing Data.

Note for Figure 1: In this simple two-dimensional illustration, it is the linear separator in (D) that best divides the points (and is, thus, chosen).

An ideal classifier can potentially address the aforementioned problems associated with standard keyword searches. For instance, a classifier trained to achieve both high precision (i.e., low false positives) and high recall (i.e., low false negatives) minimizes mistakes and omissions. Moreover, *confidence* scores can be assigned to documents to indicate relevancy to a given topic. Machine learning, however,

[1] For instance, $y_0 = 1$ if document d (0) is a critical document, and $y_0 = 0$ otherwise.

faces its own unique challenges that must be addressed.

SCARCITY OF POSITIVE EXAMPLES

For a variety of reasons, SMEs typically can supply only a small set of *positive* example documents pertaining to a critical technology. If supplied positive examples are not fully representative of the positive class, the recall suffers (i.e., false negatives increase) because the classifier might not recognize the positive documents. Thus, we must find alternative means to assist analysts in homing in on these false negatives. To that end, we extract discriminative keywords from the training set, using the concepts of *entropy* and *information gain* from the field of information theory (Manning, Raghavan, and Schütze 2008). The *entropy H* of a set of labeled documents *D* measures impurity as follows:

$$H(D) = -p^{+} log_2(p^{+}) - p^{-} log_2(p^{-}),$$

where p^{+} and p^{-} are the proportions of positive and negative documents in *D*.[2] For instance, if all documents are positive (or negative), then *H* (*D*) =0, while a perfectly even split of positive and negative documents has entropy of 1. The information gain, *IG*, of a word, w, in training set *D*, then, is the expected entropy reduction due to segmenting on w:

$$IG(w, D) = H(D) - \frac{|D^{w}|}{|D|} H(D^{w}) - \frac{|\overline{D^{w}}|}{|D|} H(\overline{D^{w}}),$$

where D^{w} is the set of documents in *D* that contain word w. Thus, words with the highest information gain in a training set are expected to be the most discriminative keywords.

During the text classification process, we match the top 15 keywords against each document analyzed. The final results, then, not only show the assigned class of the document (e.g., positive or negative), but also show keyword matches as an aid to the analysts. We sort documents assigned to positive classes by confidence score and sort documents assigned to the negative class by cosine similarity (Manning, Raghavan, and Schütze 2008) to the keyword set (both sorts are in descending order). In this way, positive documents (those correctly labeled as positive and incorrectly labeled as negative by the classifier) will ideally tend to be ranked toward the top of the list and can be reviewed more readily by the analysts.

BUILDING THE SET OF NEGATIVE EXAMPLES

We now turn our attention to the construction of sets of *negative examples*. We initially complement the aforementioned set of positive examples with a small selection of random Department of Defense (DoD) technical reports that are unrelated to the topic of interest (i.e., negative examples). This is our *initial* but incomplete version of training set D. The negative class is characterized by significantly larger size and variance than that of the set of critical documents in which we are interested. A workstation hard drive, for instance, can contain a multitude of different files, including operating system files and personal documents. We must *grow* the pool of negative

[2] Note that log_2 (0) is taken to be 0.

documents to better capture the size and variance of the negative class. We focus on sampling only the most *informative* negative examples.

What are *informative* examples? In Figure 1, the placement of the hyperplane is wholly dependent on those points (i.e., documents) that are closest to the "border" between the positive and negative class. Such "border-line" examples are referred to as *support vectors* and are used by the LinearSVM learning algorithm to approximate the optimal placement of the hyperplane. This fact yields a key insight: examples that are *not* located near the border can be discarded since they have no effect on the learned classifier or its ultimate performance. The retained examples near the hyperplane are referred to as *informative*. This fact also leads us to an intriguing approach to improving the set of negative examples: using the initial (or most recently trained) classifier to seek out negative examples that are either near the hyperplane of a classifier (i.e., ambiguous) or completely misclassified by a classifier. In the process, we are able to improve the classifier iteratively, in effect, by *exploiting* its own inadequacies. This process, known as *active learning* (Ertekin et al. 2007), is realized in three distinct ways:

- **Sampling from government machines.** We sample informative negative examples from random hard drives (e.g., unclassified file servers or workstations) that contain a mix of operating system files, files associated with installed software, personal files, and technical and non-technical government documents *unrelated* to the critical technology of interest.

- **Sampling from the web.** We take the discriminative terms extracted from the training set described previously and use them to conduct keyword searches on Google Search and Google Scholar. Our aim is to locate documents that contain terms associated with the critical technology in question but, at the same time, are wholly unrelated to the critical technology.

- **Subsampling for balance.** At the conclusion of the previous two steps, we typically arrive at a pool of negative examples that is significantly larger than the set of positive examples. Such imbalance is known to deteriorate the accuracy of machine learning classifiers. Our final step, then, is to filter down the constructed pool of negative examples. We correct this balance by again leveraging hyperplane distance to subsample only the most informative examples from the training set. Interestingly, in the process of subsampling, the training set becomes *balanced* (and redundancy is removed), as sets of examples near the border tend to be more balanced than the dataset as a whole (Ertekin et al. 2007).

These three phases of augmenting the set of negative training examples can be repeated in an iterative fashion to arrive at the final training set.

RESULTS

For this evaluation and to demonstrate our approach, we focus on the binary classification problem of finding documents that pertain to a very particular critical technology. Due

to the sensitive nature of our problem domain, we will simply refer to this critical technology as Technology X. We constructed a training set for Technology X, using the methodology discussed previously. Our final training set consisted of 51 positive examples of Technology X and 53 negative examples, for a total of 104 training examples.

Validation Tests

Using our constructed training set, we first performed *10-fold cross-validation*, a standard technique used to estimate the extent to which the performance of a trained classifier will generalize to new, independent datasets (i.e., real-world settings) (Manning, Raghavan, and Schütze 2008). As shown in Table 1, the LinearSVM classifier performed exceedingly well overall, especially given the relatively small number of training examples available. For comparison purposes, we also show results for the Naive Bayes learning algorithm (Manning, Raghavan, and Schütze 2008), which did not perform as well in this domain.

Table 1. 10-Fold Cross-Validation Results: Mean Precision, Recall, and F-Score (with Standard Errors) for Technology X.

Algorithm	Mean Precision	Recall	F-Score
LinearSVM	0.98±0.02	0.93±0.03	0.95±0.02
Naive Bayes	0.77±0.03	0.92±0.03	0.83±0.02

Note for Table 1: Such results validate our choice of LinearSVM.

Case Study: A Confirmed Positive Case

We acquired an external hard drive onto which the home directories of six users were copied. Exhaustive and time-consuming reviews by analysts indicated that this drive contained at least a few critical documents pertaining to Technology X (embedded in a plethora of work-related documents that were unrelated to Technology X). In total, the drive consisted of 25,007 documents, mostly technical in nature. Upon execution of our system, the classifier identified 15 documents as pertaining to Technology X. A manual review by an SME confirmed that all 15 documents did, in fact, pertain to Technology X (i.e., they were *true positives*).

Next, to develop an approximation for ground truth, we had two analysts conduct a manual analysis of the drive using their existing practices: manual keyword searches using ad hoc queries. The manual search took a total of 12 person-hours, while the classifier (and automated keyword search) typically complete its processing in seconds. The two analysts manually identified 21 documents as being related to Technology X. This set included the 15 documents that the classifier had identified, which left six remaining documents that required further examination. A review by an SME revealed that three of the six were, in fact, *unrelated* to Technology X (*true negatives* for the classifier). One of the files was determined to be a legitimate critical document (a *false negative* for the classifier). The final two files contained a small amount of information about Technology X but, on the whole, were unrelated to the technology (referred to as having embedded information). Our *automated keyword search* was specifically developed for detecting such documents, in addition to detecting *false negatives*.

When sorting the documents assigned to the negative class in descending order by cosine similarity, we found that the sole *false negative* and one of the two *embedded documents* were

within the top 20 (out of nearly 25,000 documents), while the second *embedded document* was within the top 70. Finally, we found that using *only* automated keyword searches was wholly inadequate since some of the content-richest documents pertaining to Technology X did not appear anywhere near the top and the signal-to-noise ratio was quite low.

This finding further motivates our use of text classification as the primary retrieval technique in this hybrid approach. The results are summarized in Figure 2 as a Venn diagram.

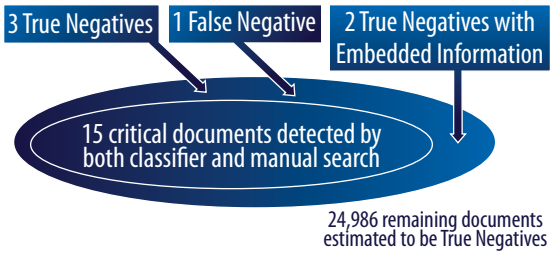


Figure 2. Venn Diagram Showing the Results for Our Case Study.

Note for Figure 2: The inner ellipse is the document set returned by the classifier, while the outer ellipse is the document set returned by a tedious manual search. True/False and Positive/Negative assignments are from the perspective of the classifier.

Dr. Maiya holds a Ph.D. in Computer Science from the University of Illinois at Chicago.

Dr. Loaiza-Lemos holds a Ph.D. in Chemistry from Princeton University and a Juris Doctor from George Mason University School of Law.

Dr. Rolfe holds a Ph.D. in Physics from the University of California, Los Angeles.

The full article was a presentation at the 2012 Military Communications (MILCOM) Conference, Orlando.

SUPERVISED LEARNING IN THE WILD: TEXT CLASSIFICATION FOR CRITICAL TECHNOLOGIES

http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6415660&isnumber=6415559

REFERENCES

Ertekin, Seyda, Jian Huang, Léon Bottou, and Lee Giles. 2007. "Learning on the Border: Active Learning in Imbalanced Data Classification." In *Proceedings of the Sixteenth ACM Conference on Information and Knowledge Management CIKM '07*, edited by Mario J. Silva, Alberto H. F. Laender, Ricardo A. Baeza-Yates, Deborah L McGuinness, Bjørn Olsen. Øystein Huag, and Andre O. Falcao, pages 127–136, Lisbon, Portugal, November 6–10. doi: 10.1145/1321440.1321461.

Fan, Rong-En, Kai-Wei Chang, Cho-Jui Hsieh, Xiang-Rui Wang, and Chih-Jen Lin. 2008. "LIBLINEAR: A Library for Large Linear Classification." *Journal of Machine Learning Research* 9 (August): 1871–1874. http://jmlr.org/papers/volume9/fan08a/fan08a.pdf.

Manning, Christopher D., Prabhakar Raghavan, and Hinrich Schütze. 2008. *Introduction to Information Retrieval*. Cambridge, England: Cambridge University Press. http://nlp.stanford.edu/IR-book/html/htmledition/irbook.html.

Wu, Ho Chung, Robert Wing Pong Luk, Kam Fai Wong, and Kui Lam Kwok. 2008. "Interpreting TF-IDF Term Weights as Making Relevance Decisions." *ACM Transactions on Information Systems* 26 (3) (June): 1–37. doi: 10.1145/1361684.1361686.

HUMAN, SOCIAL, CULTURAL, AND BEHAVIORAL MODELING FOR MILITARY APPLICATIONS

Sue Numrich and Peter Picucci

As a result of a more diversified mission space for military operations, a strong push for the inclusion of human, social, cultural, and behavior (HSCB) elements into traditional military combat modeling has occurred over the last decade. The research effort discussed here was targeted at providing the combat modeling community an introduction to the difficulties of incorporating those elements into combat modeling as well as some of the means to overcome them. The utility of developing a high-level conceptual model using the United States Marine Corps' (USMC) "Operational Culture" concept as an example case is considered. Further consideration is given to the insertion of HSCB factors through a discussion of two high-level modeling architectures (Conflict Modeling, Planning and Outcome Experimentation (COMPOEX) and High-Level Architecture (HLA)) and the differing means that support such inclusion.

BACKGROUND

The U.S. military has an expansive mission set[1] (U.S. Department of Defense 2010a). In addition to maintaining the capacity to engage in major military conflict (the domain of most combat models), the U.S. military must also be prepared to undertake counter-insurgency (COIN) operations (U.S. Department of Defense 2010b), be first responders in humanitarian assistance and disaster relief (HA/DR) missions, provide support for stabilization, security, and transition to control of indigenous government, and aid in essential reconstruction to assist governments under stress—all of which require an awareness of socio-cultural factors that exceeds the military's current capabilities to effectively model.

M – A – L – O

Previous work (Picucci and Numrich 2010) suggested that military modeling efforts incorporating socio-cultural attributes need to focus on four characteristics: **M**ission, **A**rea of responsibility, **L**evel of command, and **O**perator expertise:

[1] While the U.S. military has always undertaken a diverse and expansive set of missions, Department of Defense Directive (DoDD) 3000.05 (November 2005) established that "stability operations are a core U.S. military mission that the Department of Defense shall be prepared to conduct and support. They shall be given priority comparable to combat operations and be explicitly addressed and integrated across all DoD activities" (U.S. Department of Defense 2005).

- **Mission.** Specific socio-cultural concerns of the military will vary depending on the broad categorization of mission (HA/DR missions require different modeling capabilities than COIN missions) and on the characteristics and goals of the specific mission (not all HA/DR missions will emphasize the same concerns). Relevant incorporation of those concerns can occur but requires coordination modeling and operator communities to avoid efforts that are unresponsive to expectations.
- **Area (of Responsibility).** While the social sciences are concerned with the discovery of covering laws, it is generally acknowledged that said laws find unique expressions within different communities and populations. The discovery of a specific variable relationship in one state or community does not guarantee that the same relationship holds constant in another, with the natural implication that a model that works well as an evaluative tool for an HA/DR mission in one instance will not necessarily be applicable in another instance.
- **Level (of Command).** It must be determined whether the model is being constructed for relevance at the tactical, operational, or strategic level. This understanding is made difficult by the knowledge that socio-cultural factors at the micro and the macro levels can impact decision-making processes, perceptions, and beliefs. Turnaround time requirements, data availability, reach-back, and subject matter expert support can vary widely from the strategic to tactical levels, thus impacting the type of modeling effort that should be pursued.
- **Operator (Expertise).** Finally is the need to understand those who will be employing these tools and their level of socio-cultural expertise and comfort. A particular difficulty in developing tools that are useful to a wide range of expertise levels is the need to include socio-cultural modeling elements that are understandable and transparent for those who have limited socio-cultural experience but sophisticated enough so that highly proficient individuals will not dismiss the contributions as superficial.

While the four MALO components represent crucial nodes upon which any given model design team must reach a shared understanding, they also have more broad-based implications—one of which is that models that incorporate socio-cultural elements will be less generalizable. A task as ubiquitous as the identification of local power brokers can require the isolation of fundamentally different factors, depending on whether one is seeking to identify those relevant to an HA/DR mission or those relevant to a COIN effort. Likewise, the direct and indirect power relationships within any given society are often highly specific. Building a general model across these domains introduces clear risks. Therefore, determining the bounds of this generalizability and verifying the appropriateness of particular instantiations are critical issues.

Entities and Behaviors

For combat models, alterations in the battlespace are measured in

changes in the performance of units within that space and effects that manifest themselves as visible impacts on the behavior of units. By contrast, much of socio-cultural modeling focuses on incremental changes not in behavior, but in opinion, belief, and decision-making calculus. Further, although behavioral changes do manifest, they often do so only once certain "tipping points" have been reached (Gladwell 2000).Those tipping points shift in response to other, often macro-level, indicators. As a consequence, observable behavioral shifts are rarely directly traceable to specific, singular activities in which the military may engage.

Further difficulties arise from the dynamic nature of the entities that have to be instantiated within existing combat models. Tank, aircraft, or unit behaviors largely change only in the degree of capability they have. Social entities, however, can and do alter their sets of potential behaviors depending on social context and on how the social context is perceived. Further, while the performance of a given military platform can be impacted by the objective external conditions, those capabilities are rarely impacted by the human perception of those conditions. For social interactions, the opposite is often true: the objective social condition is often less important than how that condition is perceived by the social entity being modeled.

Data Concerns

Including the human dimension in military models increases the demands that are already placed on data. In socio-cultural assessments, many more of the data must be extracted from unstructured text from diverse sources. In most cases, no authoritative source is available for a given data element, and differing sources and extraction techniques can lead to widely varying model outputs. Further, socio-cultural data of most interest to modelers are likely to have inconsistencies or be nonexistent in areas of most critical concern. Data on perceptions are also perishable and can rapidly shift in response to actions and words either done or not done, with only some of these shifts being of significance (either persistent or extreme enough to provoke behavioral changes), thus complicating issues of data collection for operational usage.

Common Vocabulary

The ability to collect, store, and use data effectively depends on a common vocabulary or taxonomy that permits all participants, regardless of background, to understand and correctly interpret the information that is being exchanged among humans first and then among processors in a computational environment. No such common vocabulary exists among the various social sciences, and the situation only becomes more complicated when the military and the existing combat modeling community are added to the conversation. If model builders from different disciplines talk past each other while examining the problem and use different methodologies to create their models, combining these models through the exchange of formatted data will probably not create a meaningful system. Achieving this semantic consistency implies that no conflicting assumptions are buried in the entire group of models being brought together. The process of ferreting out

those inconsistencies involves building a common vocabulary for discourse and for creation of a data structure.

Conceptual Models

A means of establishing this vocabulary occurs through a coordinated construction of a conceptual model. One such effort is the Operational Culture for the Warfighter (Salmoni and Holmes-Eber 2008). Written from the perspective of anthropologists working with Marines, the authors use three different descriptive models—ecological, social structure, and symbolic—from anthropology to create a framework designed to help Marines understand, work with, and potentially influence people from different cultures. Each of models looks at any given socio-cultural group from a different perspective and provides a common terminology and lexicon for understanding that perspective. Inherent in this conceptual model is the understanding that none of the descriptive models provide the entire picture and that each may have different value to the warfighter under differing conditions and missions.

Architectures

Even in the presence of a commonly accepted vocabulary, establishing semantic understanding among the reassembled pieces is a significant problem in all forms of modeling. Establishing an appropriate, agreed-upon architecture is a means of addressing this concern. Architectures have a very specific role in addressing problems. They are solution enablers and are effective to the extent that they create an environment in which appropriate modules can communicate and interact. Both architectures discussed—Conflict Modeling, Planning and Outcome Experimentation (COMPOEX) and Human, Social, Cultural, and Behavior (HSCB) Capability within a High-Level Architecture (HLA) Federation—provide lexical and syntactic consistency but rely on human processes to establish semantic consistency.

COMPOEX

COMPOEX is a multi-resolution architecture, developed to permit the analyst to explore many potential futures and to gain an understanding of how different circumstances impact the outcomes of specific actions. The Defense Advanced Research Projects Agency (DARPA) used the architecture to enable rapid course-of-action analyses for military planning. The common vocabulary that provides the framework for COMPOEX is based on the political, military, economic, social, infrastructure and information (PMESII) systems concept of the environment, a choice that reflects the orientation of military analysts. A more frequently used architectural approach is semi-automated force modeling built to the specifications of the HLA.

HSCB Capability within an HLA Federation

HLA-enabled simulations have a common synthetic world in which entities of different types interact with one another and with the elements of the synthetic battlespace. The complications introduced by adding human dynamics involve determining what entities to add and at what levels the interactions should take place. For each population group included, a representative number of individuals would have to be

modeled together with their culturally specific decision structure and the artificial intelligence that creates the interaction between those entities, their environment, and the other entities in the environment. HSCB-aware agents could be programmed to react to the socio-cultural environment using constructs similar to those used by traditional vehicles to react to the physical environment. Environmental data would be supplied to all entities by external servers and, for the HSCB-aware entities, would be interpreted by their own internal behavioral logic.

Figure 1 shows a notional concept of the HLA architecture, separating the situational and perceptual data. The HLA architecture should support the development of an agent-server system using the concept of situational and perceptual data. Designers, however, would have to exercise restraint in choosing what to simulate in this fashion, at what level of aggregation the simulation should be composed, and which theories of behavior should animate the agents.

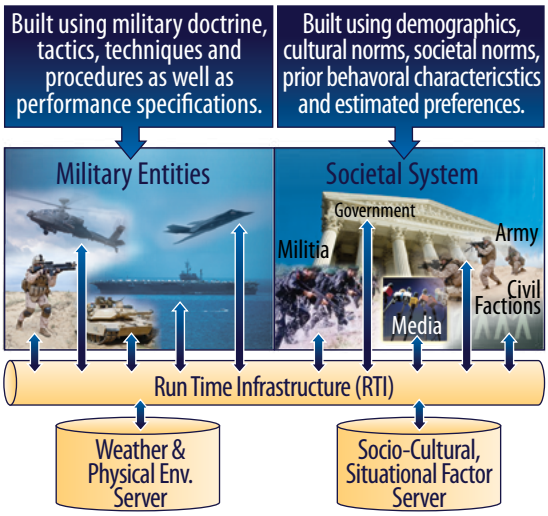


Figure 1. Socio-Cultural Server and Societal Agents in an HLA Federation.

Source: Adapted from Silverman et al 2006.

CONCLUSION

Fundamental challenges exist for the inclusion of HSCB factors into military combat modeling. The inclusion of the HSCB factors is compounded by a lack of understanding of those challenges. Although socio-cultural modeling and simulation requires that capabilities be brought together from a broad range of scientific disciplines, the connective tissue to make them work together seamlessly is lacking. The disciplines from which the capabilities arise have different vocabularies, different ways of interpreting events, and different concepts of measurement, data, and models—all which differ from what the military user expresses as his needs. Taking MALO limitations into account can help the social scientist and the military user to communicate more effectively about the nature of the problem and has the potential for providing suitable tools designed to improve understanding about the socio-cultural environment of concern to the military.

Conceptual models are a crucial element of this process, provide the beginnings of a shared vocabulary, can be made responsive to MALO concerns, and serve to temper expectations of utility. Existing modeling architectures can be responsive to the incorporation of HSCB factors. However, the novel computational methods and engineering solutions required cannot alone conquer this problem space, which is inherently interdisciplinary and multifaceted. The time spent in interdisciplinary and multidisciplinary contexts to clarify the problems of socio-cultural modeling will produce dividends in the time and resources saved in the processes of meeting the many challenges facing researchers, modelers, and military in this critical domain (Numrich and Tolk 2010).

Dr. Numrich holds a Ph.D. in Physics from the American University and did post-graduate work at Cambridge University.

Dr. Picucci holds a Ph.D. in Political Science from the University of Kansas.

The full article appeared as chapter 26 of an edited volume *Engineering Principles of Combat Modeling and Distributed Simulation*, published by John Wiley & Sons, Inc., 2012.

NEW CHALLENGES: HUMAN, SOCIAL, CULTURAL, AND BEHAVIORAL MODELING

http://onlinelibrary.wiley.com/doi/10.1002/9781118180310.ch26/summary

REFERENCES

Gladwell, Malcom. 2000. *The Tipping Point: How Little Things Can Make a Big Difference*. Little, Brown and Company.

Numrich, S. K., and Tolk, Andreas. 2010. "Challenges for Human, Social, Cultural and Behavioral Modeling." *SCS M&S Magazine* (2010-01 January): 1–9. http://www.scs.org/magazines/2010-01/index_file/Files/article_Numrich-Tolk.pdf.

Picucci, P. M., and Numrich, S. K. 2010. "Mission-Driven Needs: Understanding the Military Relevance of Socio-Cultural Capabilities." In *Proceedings of the 2010 Winter Simulation Conference* (454–465), edited by B. Johansson, S. Jain, J. Montoya-Torres, J. Hugan, and E. Yücesan, Piscataway, NJ, December 5–8. http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5679139.

Salmoni, Barak A., and Paula Holmes-Eber. 2008. *Operational Culture for the Warfighter: Principles and Applications*. Quantico, VA: Marine Corps University Press.

Silverman, Barry G., Gnana Bharathy, Kevin O'Brien, and Jason Cornwell. 2006. "Human Behavior Models for Agents in Simulators and Games: Part II Gamebot Engineering with PMFserv." *Presence: Teleoperators and Virtual Environments* 15 (2) (April): 163–185. http://www.mitpressjournals.org/doi/pdf/10.1162/pres.2006.15.2.163.

U.S. Department of Defense. 2005. Military Support for Stability, Security, Transition, and Reconstruction (SSTR) Operations. Department of Defense Directive 3000.05 (November 28). http://www.fas.org/irp/doddir/dod/d3000_05.pdf.

U.S. Department of Defense. 2010a. Quadrennial Defense Review Report. Washington, DC: Secretary of Defense. http://www.defense.gov/qdr/images/QDR_as_of_12Feb10_1000.pdf.

U.S. Department of Defense. 2010b. Irregular Warfare: Countering Irregular Threats Joint Operating Concept. Version 2.0. Washington, DC: U.S. Joint Forces Command and U.S. Special Operations Command. http://www.dtic.mil/futurejointwarfare/concepts/iw_joc2_0.pdf.

POTENTIAL MILITARY CONFLICT BETWEEN NORTH AND SOUTH KOREA

Kongdan Oh Hassig

The two Koreas have suffered through a long history of military confrontation, and there is little reason to expect that relations will improve in the near future. Over the past few years, both Koreas have strengthened their armed forces, and, thanks to the 2010 North Korean attacks in the West Sea, this military buildup is likely to continue. In light of this reality, the best that can be hoped for is to limit the violence that sometimes accompanies this confrontation and continue to seek ways to end the confrontation.

TWO STATES IN CONFRONTATION

The two Koreas have been involved in hostile confrontation since the peninsula was divided after the end of Japanese colonization in 1945. Although outbreaks of violence make news headlines, the core issue on the Korean peninsula is political confrontation, reinforced by social and economic differences. Military confrontation will continue until the two Koreas have found a way to eliminate the oppositional aspects of their political, economic, and social systems.

The essence of the *political* confrontation is that North Korea—the Democratic People's Republic of Korea (DPRK)— is a dynastic one-party dictatorship; South Korea—the Republic of Korea (ROK)—is a multiparty democracy, and both governments claim jurisdiction over the entire Korean peninsula. The *economic* confrontation has its roots in the incompatibility of the South Korean government's capitalism and the North Korean government's continued commitment to centrally managed socialism—although most North Koreans work illegally outside the system in a society rife with corruption. Not only are the two economic systems different, but the economic conditions in the two Koreas are also growing wider all the time in favor of the South. The *social* confrontation of the two Koreas is best exemplified by the dramatic differences in the amount of individual freedom permitted by the two Korean governments.

THE DANGER OF MILITARY ASYMMETRIES

When states confront each other, they are not necessarily fighting. In fact, most of the time, confrontation is passive. It could even be argued that as long as two armies openly face each other, a kind of balance exists in that the respective forces are deployed in such a way that any attack is likely to be met by a successful counterattack.

A lack of balance or symmetry in military forces could, however, lead one state to believe that it holds a military advantage that could be exploited by an attack. A comparison of the two Koreas reveals numerous asymmetries, with some seeming to benefit the North and others seeming to benefit the South.

Nuclear Weapons

The North has a few small nuclear weapons that it threatens to employ in an all-out war. The South, lacking such weapons, takes shelter under the U.S. nuclear umbrella, a canopy with thousands of nuclear weapons. One might expect that North Korea would view the nuclear balance as decidedly in South Korea's favor, but that is not the whole story because the North can decide when and if it wants to use nuclear weapons whereas the South cannot. Moreover, the leaders of North Korea, especially top military officers, are probably less concerned about the consequences of using nuclear weapons than are the Americans.

Conventional Forces

Conventional forces are unbalanced in terms of type, quantity, and quality. U.S. forces available to assist South Korean forces further complicate any calculations of balance (International Institute for Strategic Studies (IISS) 2008, 387–391). South Korea has fewer active-duty soldiers than does North Korea (687,000 vs. 1.1 million); fewer special forces (20,000 vs. 200,000); and fewer tanks (2,700 vs. 3,500), artillery pieces (5,000 vs. 10,000+), and combat aircraft (555 vs. 590). South Korea also has fewer submarines (12 vs. 63) and fewer ships (130 vs. 350), although it has more large ships (44 vs. 8). In terms of quality and training, South Korea holds a decided advantage in all weapons systems except small coastal combat boats. How the two forces would fare in various battle scenarios is difficult to say, but, in a sustained conflict, especially with the support of U.S. forces, virtually all observers outside of North Korea believe the South would ultimately destroy the North's forces.

Military Policy

Given the likelihood that North Korea would lose a lengthy war, its military policy is offensive in nature, stressing the need to attack a potential aggressor before coming under attack itself. This preference for preemption adds an important destabilizing element to the balance of forces on the peninsula.

Value of Targets

South Korea is filled with high-value targets, the best case being Seoul, which is within range of thousands of North Korean artillery pieces. In this sense, the superiority of South Korea's economy counts as a wartime liability. North Korea's cities are smaller, and military and civilian facilities are in sad need of repair. A good example would be North Korea's largest building—the unfinished Ryugyong Hotel—whose destruction would be an absolute boon to the North Koreans by saving them the cost of tearing it down.

Value Placed on Human Life

The value of individual lives is discounted in a dictatorship like North Korea's. Decisions about war and peace are made by the leaders as they consider what will benefit them personally. Witness how well Kim Il-sung survived his disastrous decision to launch the Korean War and how Kim Jong-il made it through the Arduous March period of

the 1990s[1] when hundreds of thousands of Koreans died of starvation. The Kim Jong-un regime might be willing to lose millions of its people in a war if it felt it could improve its own security. In South Korea, a decision that proved costly to the people would immediately be followed by a repudiation of the government and quite possibly punishment of its leaders.

Military Alliances

The two Koreas have very different military alliances. The ROK-U.S. alliance is solid, and, from the beginning of any large-scale conflict, the U.S. forces would presumably play an important role. Since the relationship that the DPRK has with China is not a military alliance, the North Koreans probably would not expect the Chinese to come in on their side as they did during the Korean War. This lack of support influences North Korea's wartime options, virtually forcing its army to launch a strong first strike and then hunker down and hope that the Chinese can persuade the Americans and South Koreans to abandon their counterattack.

Satisfaction with the Status Quo

South Koreans are doing well and want only to live in peace and continue to pursue prosperity. North Korea is by nature a revolutionary country: neither the leaders nor the masses can be satisfied with the status quo. The regime has frequently told its people that reunification must be accomplished by war, although perhaps this claim is simply meant to boost morale. In any case, most military provocations come from the North rather than the South, and North Korea is probably the state that will decide if and when future military confrontations take place. Table 1 shows some of the post-Korean War military actions on the peninsula.

Table 1. Post-Korean War Military Actions on the Peninsula in Descending Order of Seriousness.

Event	Details and Years
Open attacks	North Korean airplanes against ROK or U.S. airplanes or ships (1965, 1968, 1969, 1999, 2002, 2003); torpedoing of the Cheonan (an ROK naval ship (2010)), artillery attack on Yeonpyeong Island (2010)
Commando raids	Against the Presidential Mansion (1968); on the east and west coasts (1968, 1969, 1975, 1980, 1981, 1985)
Submarine incursions detected	1996, 1998
Military infiltration across the demilitarized zone (DMZ)	1969, 1970, 1976, 1979, 1980, 1992, 1995
North Korean intrusions across the military demarcation line	1996, 1997
Assassination missions against ROK authorities	1974, 1983
Tunneling under the DMZ	1974, 1975, 1978, 1990
Airplane hijackings	1958, 1969, attempt in 1971; Korean Air Lines (KAL) bombing in 1987
Kidnappings and boat hijackings	Too frequent to list; according to the ROK government, 3,835 South Koreans have been abducted since the end of the Korean War, with 517 still held in North Korea.

Sources: From various sources, including Nanto 2003 and "Estimates of South Korean Abductees . . ." 2010.

[1] The term "Arduous March" became a metaphor for the famine following a state propaganda campaign in 1993.

RESPONDING TO PROVOCATIONS

In the short term, South Korea's goal must be to limit the North Korean regime's propensity for resorting to military force. The basic principles for discouraging bad behavior are well known: Provocations should be followed promptly by punishment strong enough to reduce the chances that a similar provocation will be launched in the future. How strong the punishment needs to be is always a matter of guesswork, but past experience can provide guidelines. It is clear, for example, that condemnations from the United Nations have no effect on North Korea and thus do not count as punishment; sanction resolutions are likewise largely ineffective.

It is not easy to respond promptly and appropriately to provocations, especially when those provocations can come at any time, in almost any form, and from almost any direction. Because it costs too much to be ready to respond immediately to all possible attacks, delayed responses must be accepted as a practical alternative. If a military response is delayed too long, however, it begins to look like a separate attack and not only might draw international criticism but also could be treated by North Korea as a new provocation inviting another attack.

North Korea is a military-oriented state primed for war. Launching a military attack on North Korea (apart from a defensive response) is playing to its strength. On the other hand, North Korea is perennially poor, and the North Korean leaders need to keep their people ignorant and under control. South Korean responses in the form of economic punishment and information warfare may be more useful in discouraging North Korean attacks than bombing a few military installations, and these non-kinetic forms of response would be less likely to trigger further North Korean military action.

FINALLY ENDING CONFRONTATION

Military conflict is rooted in political, economic, and social confrontations. North Korea's provocations are launched to achieve political goals, with the paramount political goal being regime survival. Thus, to eliminate conflict, it is necessary to eliminate the North Korean regime. Hoping that the regime will change or "reform" is a poor option because the North Koreans leaders are well aware of what happens to dictators who adopt reforms (see Table 2). Until the North Korean political system changes, South Korea's best hope for peace is to maintain a strong deterrent force.

Table 2. Fates of Socialist Dictators.

Leader	Country	Fate
János Kádár	Hungary	Deposed 1988; died 1989
Erich Honecker	East Germany	Deposed 1989; arrested for corruption and manslaughter
Gustáv Husák	Czechoslovakia	Deposed 1989; expelled from party 1990; died 1991
Todor Zhivkov	Bulgaria	Deposed 1989; expelled from party; arrested for embezzlement
Wojciech Jaruzelski	Poland	Deposed 1990; charged with crimes committed while defense minister
Nicolae Ceausescu	Romania	Deposed 1989; executed

Dr. Kongdan Oh Hassig holds a Ph.D. in Asian Studies from the University of California, Berkeley. She coauthored the article with Ralph Hassig, an adjunct professor of psychology at the University of Maryland University College.

The full article was published in the journal *Joint Forces Quarterly* 64 (1st Quarter, January 2012): 82–90.

MILITARY CONFRONTATION ON THE KOREAN PENINSULA

http://kms1.isn.ethz.ch/serviceengine/Files/ISN/127220/ipublicationdocument_singledocument/e24154e5-2eef-43f3-b087-0ca5917ec061/en/Feb11_Oh_Mil_Confrontation_Korea.pdf

REFERENCES

"Estimates of South Korean Abductees from South Korea's Ministry of Unification." 2010. Yonhap News Agency, October 4. http://english.yonhapnews.co.kr.

International Institute for Strategic Studies (IISS). 2008. The Military Balance: 2008. London: Taylor and Francis Group.

Nanto, Dick K. 2003. North Korea: Chronology of Provocations, 1950–2003. Report for Congress. Washington, DC: Congressional Research Service.

ESTIMATING HAZARDOUS RELEASES IN URBAN ENVIRONMENTS

Nathan Platt, Steve Warner, Jeffry T. Urban, and James F. Heagy

During the Joint Urban 2003 (JU03) field experiment, low-altitude wind measurements were made within the city center, and high-altitude measurements were made outside of the city center. One would expect that model predictions that include meteorological observations as inputs from as close as feasible to the prediction location would result in superior predictions, and more observations might be assumed to be better than fewer observations. Contrary to this expectation, Hazard Prediction and Assessment Capability (HPAC) predictions based on wind measurements from the rooftop of the tallest building in the city were relatively poor. Evidence that this poor performance may be related to the altitude at which wind measurements were taken in the urban area comes from the Mock Urban Setting Test (MUST) field experiment, where the best HPAC predictions were generated from wind data taken within the boundaries of the obstacle array but at a height of six times that of the obstacles. Wind measurements made below 70 to 100 meters above ground level (AGL) within the city center led to worse predictions of nighttime releases of the Joint Urban 2003 (JU03) field experiment, perhaps due to the modeling system's inability to account adequately for low-level fluctuating wind measurements.

Releases of hazardous materials in densely populated urban areas continue to be of concern to the nation. Estimates of the effects of hazardous releases on the underlying population are required to aid planning, emergency response, and recovery efforts.

BACKGROUND

The potential effects of airborne releases of hazardous materials in densely populated urban areas continue to be of concern to the nation. Estimates of the effects of hazardous releases on the underlying population are required to aid planning, emergency response, and recovery efforts. These estimates require accurate knowledge of the atmospheric concentrations of dispersed material in time and space. Estimates of when and where relatively low-level concentration thresholds for human health effects can be exceeded are especially desired. Recent field experiments, designed to help us improve our hazard prediction and assessment capability, have included the release and subsequent sampling of environmentally inert tracer gases in urban environments. Tracer gases were released in Salt Lake City in 2000 (Allwine et al. 2002) and during the Mock Urban Setting Test (MUST) at Dugway Proving Ground, Utah (Biltoft 2002). Data from these experiments have been used to evaluate atmospheric transport and dispersion (AT&D) models (Warner, Platt, and

Heagy 2004b; Warner et al. 2006). Under the joint sponsorship of the U.S. Department of Defense (DoD), the Defense Threat Reduction Agency (DTRA), and the U.S. Department of Homeland Security (DHS), a series of tracer gas releases was carried out in Oklahoma City between June 28 and July 31, 2003 (Allwine et. al 2004). This field experiment, referred to as Joint Urban 2003 (JU03), included 10 time periods in which the inert tracer gas sulfur hexafluoride (SF6) was released in downtown Oklahoma City.

Intuitively, one would expect that improved hazard predictions will result from AT&D model predictions that include meteorological observations as inputs from as close as feasible to the release location and measured as often as possible. In terms of releases in a city, this expectation has led to the suggestion of using wind measurements at altitudes that are inside the "urban canopy," the region where the direction and speed of the wind are substantially influenced by the large buildings. These measurements might include rooftop observations (e.g., the roof of the relatively tall Latter Day Saints (LDS) building during the Salt Lake City field experiment (Warner, Platt, and Heagy 2004b)), or vertical wind profile measurements (e.g., the Botanical Gardens mini-SOnic Detection And Ranging (SODAR) instrument, which was used during the JU03 field trials).

In past studies, we have obtained somewhat contradictory results. The predictions of contaminated areas based on the LDS building rooftop measurements did not compare as well to the observed data as did the predictions based on measurements from meteorological instruments upwind of the central business district. The hypothesized reason for this result is that the LDS rooftop wind observations included fluctuations in the wind direction that were considered non-representative of winds elsewhere in the city (e.g., they were an artifact of the unique LDS rooftop location). On the other hand, model predictions of the MUST field experiment (dispersion through an array of cargo containers intended to represent a scaled-down city), which were based on meteorological measurements obtained directly above the container array, were improved relative to predictions based on measurements from other meteorological instrument locations (Warner et. al 2006). In this case, it was hypothesized that the wind measurements, which were made at an altitude that was six times the height of the containers, sampled the relatively unperturbed wind flow above the array and therefore were more globally representative of the wind flow across the extent of the array. For the measurements taken within an urban area, low-altitude measurements can be affected by obstacles (such as buildings) within the urban canopy, while the upper altitude observations presumably measure the relatively unperturbed winds.

Comparison Methodology

The Hazard Prediction and Assessment Capability (HPAC) modeling system (DTRA 2001) is composed of a suite of software modules that can generate source

terms for hazardous atmospheric releases, retrieve and prepare meteorological information for modeling purposes, model the transport and dispersion of the hazardous release over time, and plot and report the results of these calculations. IDA is providing analytic support of DTRA's efforts to validate HPAC and its individual submodels. A significant portion of this effort is directed toward the urban AT&D modules.

A large number of vertical wind profile meteorological measurements and ground-level tracer gas measurements were made in the vicinity of Oklahoma City during the JU03 field experiment. The source of vertical wind measurements closest to the city and the inert gas releases, the Botanical Gardens mini-SODAR instrument, was located about 200 meters upwind of the upwind edge of the tracer sampling network. This instrument provided wind speed and direction measurements starting at 15 meters above ground level (AGL) and extending to approximately 200 meters AGL, in 5-meter increments. These wind measurements were recorded every 15 minutes. To study the effects of low-altitude meteorological observations within the urban canopy on urban HPAC prediction quality, we varied the "cut-off" altitude below which wind measurements were not included as inputs to the model.

For this analysis, we compared tracer gas predictions and observations paired in space and time using a measure of effectiveness (MOE) (Warner, Platt, and Heagy 2004a) that allowed for assessments of the model's ability to predict the contaminated ("hazardous") region (i.e., the region in which the tracer gas concentration exceeded a threshold of interest). This MOE has two dimensions and is described in Figure 1. The x axis corresponds to the ratio of the area of overlap between the predicted and observed hazardous regions to the area of the observed region. The y-axis corresponds to the ratio of the area of the overlap between the predicted and observed hazardous regions to the area of the predicted hazardous region. After algebraic rearrangement, one recognizes that the x-axis corresponds to 1 minus the fraction of the observed hazardous

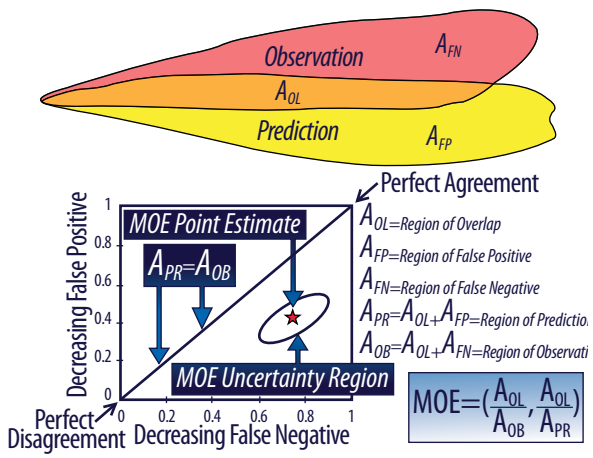


Figure 1. MOE Description.

region that corresponds to false negative predictions and that the y-axis corresponds to 1 minus the fraction of the predicted hazardous area that corresponds to false positive predictions. The MOE is given as

$$MOE = (x, y) = \left(\frac{A_{OV}}{A_{OB}}, \frac{A_{OV}}{A_{PR}}\right) = \left(\frac{A_{OB} - A_{FN}}{A_{OB}}, \frac{A_{PR} - A_{FP}}{A_{PR}}\right) = \left(1 - \frac{A_{FN}}{A_{OB}}, 1 - \frac{A_{FP}}{A_{PR}}\right)$$

where A_{FN} = area of the region of false negative predictions, A_{FP} = area of the region of false positive predictions,

A_{OV} = area of the region of overlap between observed and predicted hazardous areas, A_{PR} = area of the predicted hazardous region, and A_{OB} = area of the observed hazardous region.

This MOE includes directional effects (i.e., a good prediction according to this MOE will match the actual geographic location of the hazard and not just the shape and size of hazardous area). A perfect model prediction leads to no false negative and no false positive and, hence, perfect overlap between the predicted and observed hazardous areas, corresponding to an MOE value of (1,1). MOE values along the "diagonal" of the MOE space imply that the areas of the predicted and observed hazardous areas are equal (i.e., A_{PR} = A_{OB}), even if their locations differ.

RESULTS AND CONCLUSIONS

Figure 2 shows MOE plots for HPAC predictions of the area over which the atmospheric tracer gas concentration exceeds 250 parts per trillion (ppt) of tracer gas using "altitude thresholding" of the Botanical Gardens mini-SODAR measurements of the vertical wind profile. Altitude thresholding corresponds to discarding wind measurements below a given threshold altitude to explore whether low-altitude winds may adversely affect the tracer gas dispersion predictions. On the left side of the figure, the MOE plots depict statistically resampled MOE values (corresponding to approximate 99% confidence regions) obtained for the 12 nighttime releases. On the right side of the figure, the MOE plots depict statistically resampled MOE values obtained for the 17 daytime releases. For the daytime releases, discarding low-altitude mini-SODAR wind measurements has little effect on the MOE values—the MOE confidence region (i.e., the confidence "ellipses" for the predictions generated from wind data at different altitude thresholds largely overlap with each other). For the nighttime releases, however, the MOE values depend strongly on the altitude below which the wind measurements are discarded, and the associated confidence regions for the MOE estimates do not overlap. The MOE values improve as the threshold altitude increases, and the worst predictions (i.e., when the MOE confidence region is closest to (0,0)) are those that are based on the full set of mini-SODAR wind measurements (all altitudes).

An important implication derived from this research is that, for prediction of nighttime releases, SODAR wind measurements below approximately 70 to 100 meters within the urban canopy should not be used as inputs for HPAC predictions of the JU03 field experiment. For the daytime releases, all available vertical wind measurements from the SODAR could be used to obtain

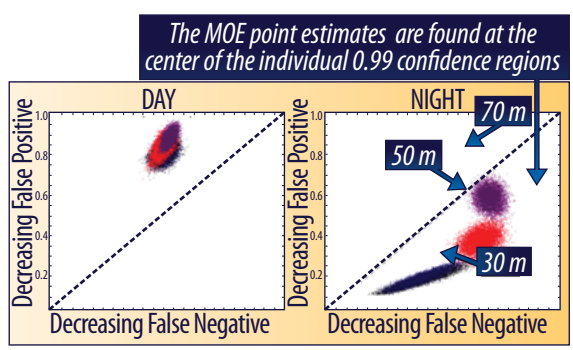


Figure 2. MOE Results as a Function of Vertical Profiler Data Cut-off Altitude: Day vs. Night.

reasonable predictions of the JU03 field experiment. Three other SODAR instruments near the city collected substantial vertical wind profile measurements during the JU03 experiment. We conducted similar analyses using data from all three instruments and obtained similar results. Also, we conducted similar evaluations using a Los Alamos-developed ("QUIC-URB/QUIC-PLUME") urban AT&D model and verified that withholding low-altitude mini-SODAR data improves model agreement with observations. We hypothesize that these results are due to changes in the height of the atmospheric boundary layer above Oklahoma City during the night, similar to the well-known change in atmospheric stability category between daytime and nighttime over flat (non-urban) terrain.

Figure 3 depicts wind vector profiles measured by the Botanical Gardens mini-SODAR as a function of time and altitude. The left panel depicts a set of wind speeds and directions observed during a typical daytime release, while the right panel does the same for a typical nighttime release. For the nighttime release, the wind direction varies significantly at altitudes below 50 meters AGL, while the wind speed drops to almost zero at about 30 meters AGL. For the daytime release, the wind direction is quite consistent with changes in the altitude, while the wind speed increases relatively smoothly with increasing altitude (with the exception of the very lowest measurements at 15 meters AGL). The behavior of the wind profile for the daytime release is similar to the behavior of the wind profile for the nighttime release above 50 meters AGL. We surmise that the HPAC meteorological preprocessor unsuccessfully attempts to capture the intricacies of the nighttime meteorology involving the unpredictable wind direction at the lower altitudes, resulting in predicted hazardous areas that have a wide spread around the release site and a significant overprediction of the hazard near the release site.

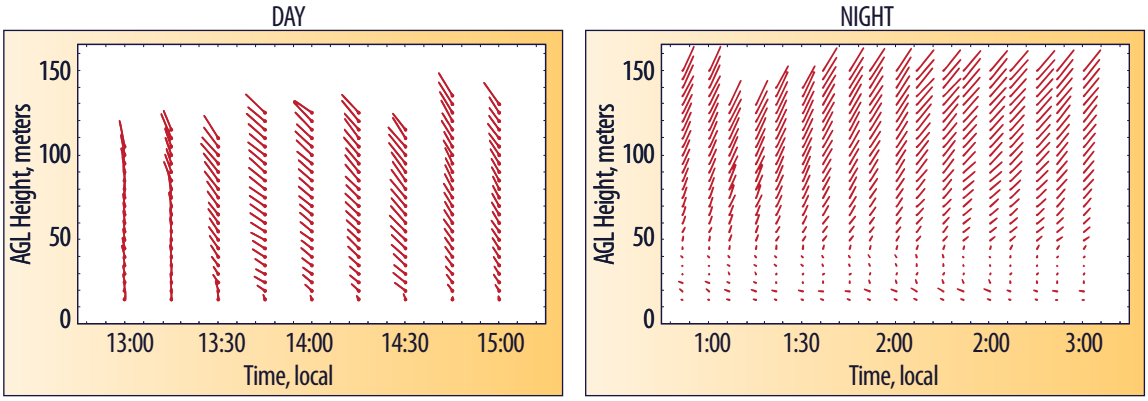


Figure 3. Wind Vectors as a Function of Altitude: Day vs. Night.

ACKNOWLEDGMENTS

This effort was supported by the Defense Threat Reduction Agency, with Dr. John Hannan as the project monitor.

Dr. Platt holds a Ph.D. in Applied Mathematics from Brown University.

Dr. Urban holds a Ph.D. in Chemistry from the University of California, Berkeley.

Dr. Heagy holds a Ph.D. in Physics from Drexel University.

Dr. Warner holds a Ph.D. in Inorganic Chemistry from the Massachusetts Institute of Technology.

The full article was published in the *International Journal of Environment and Pollution* 44 (1/2/3/4): 280–287.

EFFECTS OF METEOROLOGICAL OBSERVATIONS ON THE QUALITY OF URBAN HPAC PREDICTIONS OF THE JOINT URBAN 2003 FIELD TRIALS

http://www.inderscience.com/offer.php?id=38428

REFERENCES

Allwine, K. J., J. H. Shinn, G. E. Streit, K. L. Clawson, and M. Brown. 2002 "Overview of URBAN 2000: A Multiscale Field Study of Dispersion Through an Urban Environment." *Bulletin of the American. Meteorological Society* 83 (4): 521–536. doi:10.1175/1520-0477(2002)083<0521:OOUAMF>2.3.CO;2.

Allwine, K. J., M. J. Leach, L. W. Stockham, J. S. Shinn, R. P. Hosker, J. F. Bowers, and J. C. Pace. 2004. "Overview of Joint Urban 2003: An Atmospheric Dispersion Study in Oklahoma City." Paper (J7.1) presented at *The 84th AMS Annual Meeting, Joint Session 7, Joint Urban 2003 Field Study and Urban Mesonets*, Seattle, WA, January 10–16. https://ams.confex.com/ams/84Annual/techprogram/session_16251.htm.

Biltoft, C. A. 2002. *Customer Report for Mock Urban Setting Test*. Technical Report No. WDTC FR-01-121. Dugway, UT: U. S. Army Dugway Proving Ground.

DTRA. 2001. *The Hazard Prediction and Assessment Capability (HPAC) User's Guide, Version 4.0.3*. Technical Report HPAC-UGUIDE-02-U-RAC0. Prepared for the Defense Threat Reduction Agency by Science Applications International Corporation.

Warner, Steve, Nathan Platt, and James F. Heagy. 2004a. "User-Oriented Two-Dimensional Measure of Effectiveness for the Evaluation of Transport and Dispersion Models." *Journal of Applied Meteorology* 43 (1) (January): 58–73. doi: http://dx.doi.org/10.1175/1520-0450(2004)043<0058:UTMOEF>2.0.CO;2.

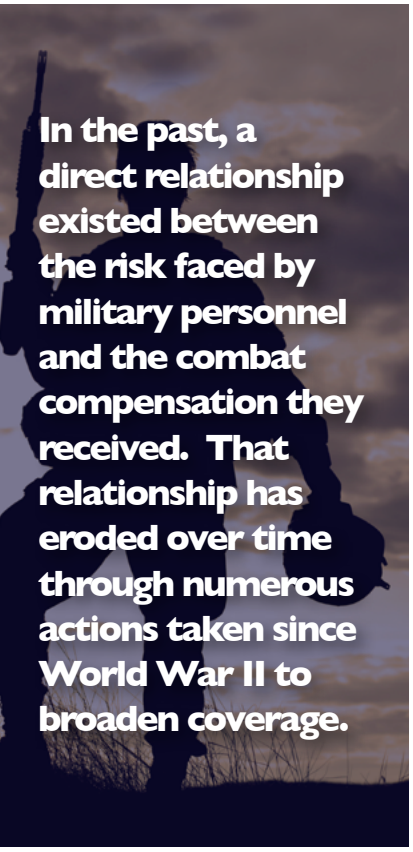
Warner, Steve, Nathan Platt, and James F. Heagy. 2004b. "Comparisons of Transport and Dispersion Model Predictions of the URBAN 2000 Field Experiment." *Journal of Applied Meteorology* 43 (6) (June): 829–846. doi: http://dx.doi.org/10.1175/1520-0450(2004)043<0829:COTADM>2.0.CO;2.

Warner, Steve, Nathan Platt, and James F. Heagy, Jason E. Jordan, and George Bieberbach. 2006. "Comparisons of Transport and Dispersion Model Predictions of the Mock Urban Setting Test Field Experiment." *Journal of Applied Meteorology* 45 (10) (October): 1414–1428. doi: http://dx.doi.org/10.1175/JAM2410.1.

ASSESSING COMBAT RISK AND COMPENSATION

Stanley A. Horowitz, Alexander Gallo, Brandon Gould, Maggie Li, Shirley Liu

The Eleventh Quadrennial Review of Military Compensation (QRMC) was chartered to review four areas of the military compensation system, one of which was compensation for service performed in a combat zone, combat operation, or hostile fire area, or while exposed to a hostile fire event. The QRMC asked IDA to examine the extent to which variations in combat compensation reflected differences in risk.



BACKGROUND

Combat compensation is an important element in the remuneration of military personnel. The principal justification for combat compensation is to recognize military personnel who face significant combat risk. In the past, a direct relationship existed between the risk faced by military personnel and the combat compensation they received.[1] That relationship has eroded over time through numerous actions taken since World War II to broaden coverage. Today, some service members in declared combat zones are subject to little risk and receive all the elements of combat compensation, while others in hostile situations but not in combat zones do not fully receive combat compensation. By examining casualty rates, both killed-in-action and wounded-in-action, we find that many areas in designated combat zones give rise to very little risk (e.g., Saudi Arabia and the United Arab Emirates and ships in the combat zone during recent operations in Iraq). More than half of the countries in combat zones have zero casualty rates. Surveys show that military members recognize that their combat-zone deployments often are not dangerous.

Eligibility for combat compensation is determined by the designation and management of combat zones. Military members who are deployed to areas of combat or combat support operations receive hostile fire pay/imminent danger pay (HFP/IDP) and the combat zone tax exclusion (CZTE). HFP/IDP provides $225 for any month or part of a month that the member is deployed to a combat zone or to a designated imminent danger area. Under the CZTE, all pays and bonuses received by an enlisted member or warrant officer in a designated combat zone are excluded from federal and state

[1] For example, Badge Pay was initially awarded only to front-line units in World War II.

income taxes. Officers in the same situation in 2011 were allowed to exclude up to $7,714.80 per month from their tax returns. HFP/IDP cost the Department of Defense (DoD) $789 million in 2009. The Department of the Treasury's cost for CZTE was $3.6 billion—approximately 4.5 times the cost of HFP/IDP.

FINDINGS

While all military members, regardless of rank, who are deployed to a combat zone receive the same amount of HFP/IDP, the CZTE benefit affects service members differently. The tax exclusion lowers the individual's income tax obligations and creates eligibility for various tax credits and deductions; however, depending upon an individual's circumstances (e.g., marital status, filing status, family size, or medical deductions), the value of the CZTE is quite variable. IDA collaborated with the Department of the Treasury to determine—for the first time—the value of the CZTE to the individual service member. For 2009, the average total value of the CZTE was $5,990, with the value at the first percentile at $280 and the value at 99th percentile at $22,430—almost 100 times the value at the lower end. More than half of those deployed to a combat zone received at least $4,660 in federal tax savings and benefits. One unexpected aspect of CZTE-related compensation is that some officers, due to their ability to exclude this income, also qualified for the Earned Income Credit (EIC), established to help low-income families.

First enacted in 1975, the EIC has been expanded by tax legislation numerous times to become a significant anti-poverty tool. The EIC was created to supplement the income of low-income households by creating a refundable tax credit that varies with the number of dependents and income. Figure 1 illustrates the structure of the EIC program. The EIC has three component parts—a phase-in that increases with increased income; a plateau, where benefit levels are constant as income increases; and a phase-out that decreases benefits with increased income and a maximum amount, where benefits are exhausted. Figure 1 shows four different relationships, ranging from "No Children" to "Three or More." Increasing family size (up to a family size of five) increases the benefit and the phase-in and phase-out levels.

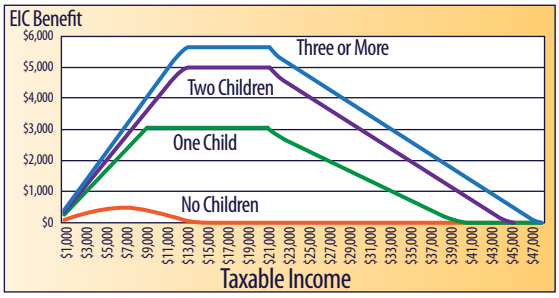


Figure 1. EIC Benefit and Income.
Source: Department of the Treasury, Internal Revenue Service.

Present policy allows members to decide whether to count income received in a combat zone.[2] Consider an E-4 serving in a combat zone for 6 months who has more than 4 years of service (YoS) and two children. Pay for this E-4 is approximately $2,200 per month. Assuming no spousal income,

[2] This is an all-or-nothing decision. Members cannot choose to count a portion of their income.

this E-4's monthly pay, plus $325 per month in HFP/IDP and hardship duty pay (HDP), results in an annual total of $30,300.[3] EIC for this family would be $3,160. With CZTE, the member can opt to exclude $15,150 in income earned while deployed to a combat zone, thereby reducing the member's income by half and increasing EIC to the maximum of $5,036.[4]

Now consider an O-6 with more than 20 YoS and two children. The O-6's annual basic pay, HFP/IDP, and HDP is $113,048. Adding to this average *non-taxable* housing and subsistence allowances in the amount of $31,656, annual income for an average O-6 is $144,704. If this O-6 is deployed for 12 months to a combat zone, he or she can elect to exclude $92,532 of his or her income from the EIC calculation. Because allowances are not taxable, the O-6's net income is reduced to $20,516 and, consequently, he becomes entitled to the maximum EIC payment of $5,036.[5] The O-6, whose total compensation is about five times that of the E-4, can receive more in EIC than an E-4 stationed in the United States.

Combat compensation received by members within a combat zone varies widely. Since HFP/IDP does not vary by pay grade, any variation in combat compensation benefit is the result of differences in CZTE benefit.

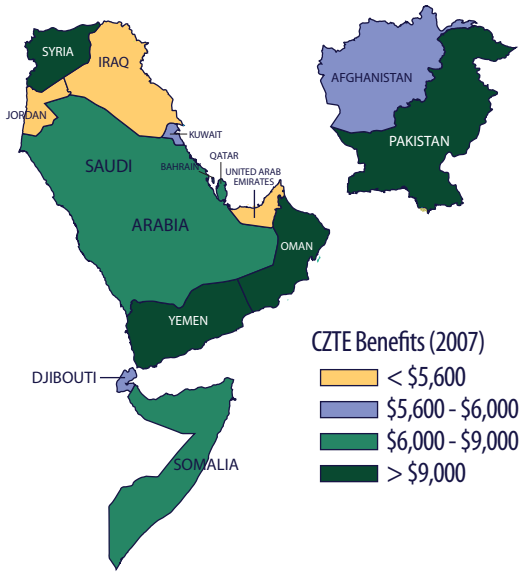


Figure 2. CZTE Savings by Country 2007.
Source: Department of the Treasury, Office of Tax Analysis.

Figure 2 presents the average CZTE savings by country. The average CZTE savings is calculated by pay grade and weighted by the number of members of that pay grade deployed, by country.[6] The greater the percentage of officers and the greater the seniority, the higher, in general, will be the average CZTE benefit.

Figure 3 shows casualty rates by country for 2007. Comparison of the average benefit with the average casualty rate indicates many anomalies. For example, Oman, a country with a zero casualty rate, has an average benefit that is almost 50 percent higher than Afghanistan, the country with the highest casualty rate.

[3] This E-4 would also receive $18,514.44 in housing and subsistence allowances as part of his or her regular military compensation (RMC) of $47,218.66 per year.

[4] Actual EIC benefit is based upon family income. EIC benefits could increase or decrease based upon income exclusion and spousal income.

[5] The relationship between time deployed and CZTE benefit for an O-6 is similar to the previous discussion for the E-4.

[6] Calculation of CZTE savings is based upon data provided by the Department of the Treasury. The Internal Revenue Service analyzed the taxes of everyone who benefited from the CZTE and calculated the extent of the benefit.

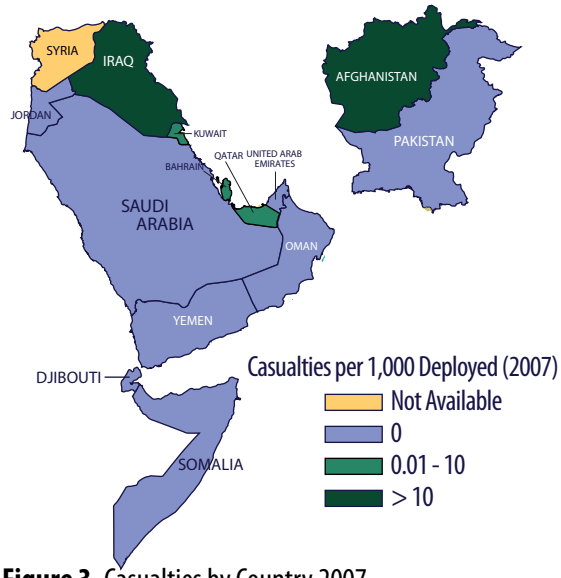


Figure 3. Casualties by Country 2007
Source: Defense Manpower Data Center.

Overall, we find virtually no correlation across countries within combat zones between casualty rates and average combat compensation. Countries with zero casualty rates tend to have the highest average benefit primarily because of their pay grade structure. Furthermore, junior enlisted personnel (along with junior officers) have the highest incidence of death and injury, but benefit the least from the CZTE.

SUMMARY

The divergence between the risks that military members face when deployed and the associated compensation can be brought into better alignment in several ways. The designation of combat zones is difficult to initiate and even more difficult to terminate. As a result, combat zones include areas in which no combat takes place and areas in which the threat of hostilities is minimal/nonexistent. Better management of what is and what is not a combat zone could eliminate combat compensation from being paid to service members in areas in which the risk or threat of danger is minimal or nonexistent. As a result, compensation could be limited to members actually exposed to danger.

CZTE benefits, the major component of combat compensation, depend to a large extent upon individual circumstance and the vagaries of the tax code, which are totally unrelated to risk. Because of the complexity of determining the CZTE benefit, it is not likely that members know the actual amount of benefit nor can they compare these rewards with the risks of combat. CZTE benefits could be made more uniform by substituting a refundable income tax credit for the present system of income exclusion.

A major part of the current CZTE benefit is eligibility for the EIC, a program designed to assist low-wage households. Income exclusion allows field-grade officers and senior enlisted personnel—in some cases, officers with basic pay and allowances in excess of $150,000 per year—to be eligible for this program. Basing EIC eligibility on all income (which includes that excluded for tax purposes by the CZTE) would restrict EIC payments to those households that qualify within the original intent of the program.

The DoD's stated philosophy is for compensation to increase with increased danger or risk. This goal cannot be achieved within the current structure of CZTE. A closer relationship than that which currently exists between risk and compensation could be attained in a variety of

ways. For example, the CZTE could be eliminated in favor of a tiered, refundable tax credit available to those in designated areas. Another possibility would be for DoD to adopt a "true" combat pay for members actually in a combat environment. This combat pay could either be a supplement to other kinds of combat compensation or a substitute.

Mr. Horowitz holds an M.A. in Economics from the University of Chicago.

Mr. Gallo holds an M.A. in Global Security Studies from Johns Hopkins University.

Mr. Gould holds an M.A. in Public Policy from the University of Virginia, Frank Batten School of Leadership and Public Policy.

Ms. Li holds a B.A. in Economics from Princeton University.

Dr. Liu holds a Ph.D. in Economics from the State University of New York at Stony Brook.

The full article was published in Chapter 9 of the *Report of the 11th Quadrennial Review of Military Compensation: Main Report*

COMBAT RISK AND COMBAT COMPENSATION

http://militarypay.defense.gov/reports/qrmc/11th_QRMC_Supporting_Research_Papers_Files/SR12_Chapter_9.pdf

CARBON FOOTPRINT OF SHALE GAS

Christopher L. Weber and Christopher Clavin

The recent increase in the production of natural gas from shale deposits has changed energy outlooks in the United States and the broader world. Shale gas may have important benefits for climate change if it displaces more carbon-intensive oil or coal. But recently, significant attention has been on the potential for upstream methane emissions to neutralize these reduced greenhouse gas emissions from combustion. We examined six recent studies to produce a Monte Carlo uncertainty analysis of the carbon footprint of both shale and conventional natural gas production. Comparison shows the most likely upstream carbon footprints of these different types of natural gas production are largely similar, with overlapping 95 percent uncertainty ranges of 11.0 to 21.0 g CO_2e/MJ_{LHV} for shale gas and 12.4 to 19.5 g CO_2e/MJ_{LHV} for conventional gas. However, because this upstream footprint represents less than 25% of the total carbon footprint of gas (with combustion representing the remainder), the efficiency of producing heat, electricity, transportation services, or other function is of equal or greater importance. Better data are clearly needed to reduce the uncertainty in natural gas's carbon footprint, but understanding system-level climate impacts of shale gas, through shifts in national and global energy markets, may be more important and requires more detailed energy and economic systems assessments.

BACKGROUND

Recent advances in horizontal drilling and hydraulic fracturing technology have made it technically and economically possible to access vast deposits of natural gas in shale deposits located across the United States and elsewhere (Birol et al. 2011; EIA 2011a, 4). Shale gas production grew 48 percent per year from 2006 to 2010 in the United States, and growing estimates of recoverable resources have altered U.S. and world energy outlooks (EIA 2011b, 2). Many have praised shale gas's impact on jobs, domestic energy prices, and climate change.

Despite the potentially positive impacts of shale gas development, however, it has also been criticized for several reasons, including but not limited to its impacts on water quality, air quality, and climate change. One hotly debated concern in the recent literature is the potentially high life-cycle greenhouse gas emissions (i.e., carbon footprint[1] (BSI 2008)) associated with shale

[1] The term "carbon footprint" is a standardized term in environmental impact analysis that refers to the overall life cycle (i.e., cradle to grave) impact of a product, service, or organization on climate change.

gas production due to fugitive methane during production. A study by Howarth, Santoro, and Ingraffea (2011) suggested that fugitive methane emissions from shale gas yielded a higher overall carbon footprint for shale gas compared to coal, although the authors' methods and modeling choices have been criticized by other authors (Stephenson, Valle, and Riera-Palou 2011; Burnham et al 2012; Cathles et al. 2012). Since the publication of this study, several authors have performed similar life-cycle carbon footprint studies using different data and assumptions (Hultman et al. 2011; Jiang et al. 2011; Skone, Littlefield, and Marriott 2011; Stephenson, Valle, and Riera-Palou 2011; Venkatesh et al. 2011; Burnham et al. 2012).

The goal of this analysis is threefold: first, to compare the original study to five subsequent studies with consistent assumptions; second, to compare these recent estimates of the life-cycle carbon footprint of shale gas to conventional onshore natural gas production; and third, to examine the impacts of recent Environmental Protection Agency (EPA) regulations (New Source Performance Standards (NSPS)) on methane emissions from natural gas that require reduced emission completions (RECs). The policy implications of these studies and their uncertainties are also discussed.

METHODS

We compared the results of six comparable studies of shale gas and conventional natural gas carbon footprints by harmonizing study assumptions, units, and scopes. A Monte Carlo simulation (sample size 10,000) was conducted using a combination of the inputs taken from across the studies and further created category subtotal (preproduction, production, and transmission) and total carbon footprint estimates for conventional and shale gas production. Among many other issues, the research harmonized the different studies in terms of assumed processes, well lifetimes, units, and allocation of emissions between natural gas and co-produced oil. More details on these methods can be found in the journal publication.

RESULTS

Upstream Comparison

Figure 1 shows the results of the upstream (i.e., from production of raw natural gas to delivery at a power plant) carbon footprint at the category level (preproduction, production, and transmission) for each study and the best estimate simulated through Monte Carlo analysis.

With respect to shale gas, most of the studies' best estimates for shale gas carbon footprint fall within a narrow range of 13 to 15 g CO_2e/MJ_{LHV}.[2] The exceptions are Stephenson (low) and Howarth (high). Howarth's total for shale gas, which Figure 1 shows as the

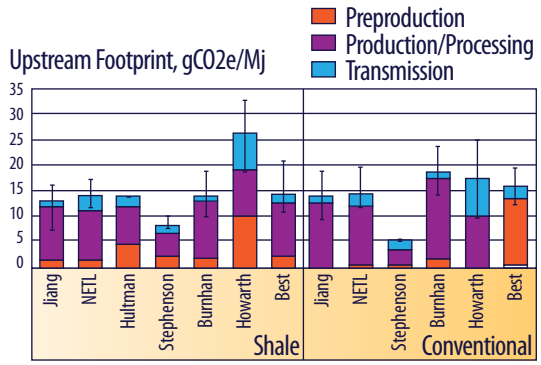


Figure 1. Summary of Upstream Carbon Footprint E stimates and Uncertainty Ranges for Each Study, Including Authors' Best Estimates.

[2] This unit represents the total greenhouse gas emissions normalized to the unit of energy in the fuel.

midpoint of a low and a high estimate, is well outside the range of uncertainty estimated by the other authors due to two extremely high estimates for well completion and fugitive emissions in transmission, as discussed previously and by several authors (Barcella, Gross, and Rajan 2011; Stephenson, Valle, and Riera-Palou 2011; Burnham et al. 2012).

An equally important observation, however, is the relative difference in the upstream carbon footprint between conventional gas production and shale gas production. As Figure 1 shows, this difference is much smaller than the uncertainty in either estimate. Our alternative scenario with NSPS required REC estimates the carbon footprint at 12.7 g CO_2e/MJ_{LHV}, with an uncertainty range of 9.9 to 15.6 g CO_2e/MJ_{LHV}. Of the studies that compare the two gas sources, some estimated a higher total for conventional gas than shale gas, whereas Howarth and Stephenson estimated a higher total for shale gas. The balance between the two depends on the assumptions associated with liquids unloading, which most authors assume applies only to conventional wells, and the higher one-time emissions associated with well completion and workovers for shale gas. Substantially reducing the emissions associated with completions and workovers from shale wells reduces the mean and the uncertainty range associated with shale gas carbon footprint.

WELL-TO-WIRE COMPARISON

While emissions from the upstream component of the natural gas supply chain may be significant, they must be placed into context of the overall life cycle of natural gas, which, in most cases, ends in combustion for some purpose—natural gas-fired electricity, commercial or home heating, industrial energy use, and so forth. The well-to-wire analyses conducted by all studies reviewed here (other than Howarth) put these emissions in the context of the life cycle for natural gas-fired vs. coal-fired electricity.

Figure 2 shows well-to-wire carbon footprints[3] (including the best estimate upstream values for shale gas as shown in Figure 1) across the range of power generation efficiencies in the studies, including typical single-cycle power plants, high-efficiency combined-cycle plants, and the average mix of gas plants from 2001 to 2010. Importantly, despite the somewhat high uncertainty in upstream emission estimates (96 g CO_2e/kWh at 37 percent power-plant efficiency), the difference between different types of natural gas power plants (less efficient steam turbine vs. highly efficient combined-cycle plants, 171 g CO_2e/kWh) accounts for a much greater difference. Over the last decade, the increase in fleet average efficiency, in part due to increased use of combined-cycle plants, has already pushed down the

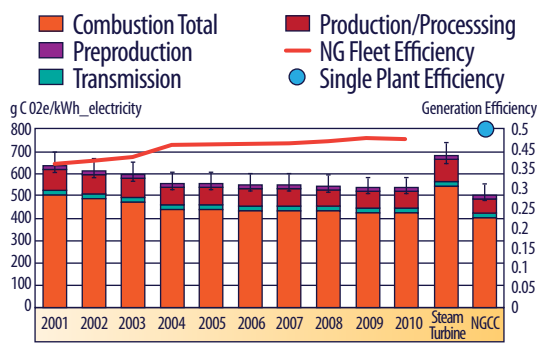


Figure 2. Well-to-Wire Carbon Footprint for Shale Gas Across the Range of Power Generation Efficiencies.

[3] Well-to-wire footprints include both upstream emissions and the creation of electricity (i.e., "wire") from the fuel in question.

life-cycle carbon footprint of shale gas by more than 100 g CO_2e/kWh. This life-cycle emission decrease is significant because it is larger than the total upstream uncertainty by more than 25 percent (76 g CO_2e/kWh in 2010) and is nearly as large as the upstream emissions altogether (111 g CO_2e/kWh in 2010).

DISCUSSION AND POLICY IMPLICATIONS

Our review of several studies shows that although the carbon footprint of shale gas is highly uncertain, it is also difficult to distinguish from conventional onshore gas production. While reducing the life-cycle emissions associated with conventional and unconventional gas should be a goal for policy, the evidence to date suggests that the carbon footprint of shale gas is not of considerably greater concern than previously known issues with the natural gas system.

One of the traditional uses of carbon footprint assessments has been to identify "hot spots" where the environmental impacts of a product can be reduced (Acquaye et al. 2011). Many of the upstream greenhouse gas emissions associated with natural gas can be controlled effectively and economically through flaring (thus converting high-Global Warming Potential (GWP) methane to CO_2) or capture of fugitive emissions through best practices, such as RECs. The studies analyzed here provide significant guidance on where to focus to reduce emissions effectively from shale and conventional gas production.

Nevertheless, despite these uncertainties, the upstream emissions associated with both types of gas are less significant because they represent less than a quarter of total life-cycle emissions when combustion in typical power plants is included. When examining the entire life cycle, it is clear that the uncertainty in upstream emissions is less significant than using the resulting gas in the most efficient manner possible. Within the electricity sector, the average gas power generation fleet is expected to continue recent efficiency increases given the relatively low use of advanced combined-cycle gas plants today.

However, alternative future energy pathways could lead to alternative uses for the large quantities of shale gas expected to be extracted in the coming years, including power generation, transportation (through compressed natural gas or electricity), industrial usage, and exports (Birol et al. 2011; Venkatesh et al. 2011; Burnham et al. 2012). These uses are not mutually exclusive, but not all can simultaneously be expanded. Any natural gas used to displace coal-fired electricity will not be available to displace oil in the transportation sector. It is within this broader scope of systems issues that the true impacts of such a large energy shift must be analyzed. Such questions can be answered only by the use of energy systems and climate models that take into account the results of carbon footprint analyses and the economics and technological development pathways (Wigley 2011; EIA 2012, 6; Jacoby, O'Sullivan, and Paltsev 2012).

Dr. Weber holds a Ph.D. in Civil and Environmental Engineering (CEE) and Engineering and Public Policy (EPP) from Carnegie Mellon University.

Mr. Clavin holds a M.A. in Engineering and Project Management from the University of California, Berkeley.

The original article was published in *Environmental Science & Technology, a journal of the American Chemical Society*, April 2012.

LIFE CYCLE CARBON FOOTPRINT OF SHALE GAS: REVIEW OF EVIDENCE AND IMPLICATIONS

http://pubs.acs.org/doi/abs/10.1021/es300375n

REFERENCES

Acquaye, Adolf A., Thomas Wiedmann, Kuishang Feng, Robert H. Crawford, John Barrett, Kohan Kuylenstierna, Adian P. Duffy, S. C. Lenny Koh, and Simon McQueen-Mason. 2011. "Identification of 'Carbon Hot-Spots' and Quantification of GHG Intensities in the Biodiesel Supply Chain Using Hybrid LCA and Structural Path Analysis." *Environmental Science and Technology* 45 (6) (February): 2471–2478. http://pubs.acs.org/doi/pdfplus/10.1021/es103410q.

Barcella, Mary Lashley, Samantha Gross, and Surya Rajan. 2011. *Mismeasuring Methane: Estimating Greenhouse Gas Emissions from Upstream Natural Gas Development. Cambridge*, MA: IHS CERA, Inc. http://www.ihs.com/images/Mismeasuring-Methane-feb-2013.pdf.

Birol, Faith, John Corben, et al. 2011. *World Energy Outlook 2011: Are We Entering a Golden Age of Gas? Special Report*. Paris, France: International Energy Agency, Office of the Chief Economist. http://www.worldenergyoutlook.org/media/weowebsite/2011/WEO2011_GoldenAgeofGasReport.pdf.

BSI. 2008. *PAS 2050:2008 Specification for the Assessment of the Life Cycle Greenhouse Gas Emissions of Goods and Services*. London: British Standards Institute. http://clients.squareeye.net/uploads/brompton/documents/PAS2050.pdf.

Burnham, Andrew, Jeongwoo Han, Corrie E. Clark, Michael Wang, Jennifer B. Dunn, and Ignasi Palou-Rivera. 2012. "Life-Cycle Greenhouse Gas Emissions of Shale Gas, Natural Gas, Coal, and Petroleum." *Environmental Science and Technology* 46 (2): 619–627. http://pubs.acs.org/doi/pdf/10.1021/es201942m.

Cathles, Lawrence M., III, Larry Brown, Milton Taam, and Andrew Hunter. 2012. "A Commentary on 'The Greenhouse-Gas Footprint of Natural Gas in Shale Formations' by R. W. Howarth, R. Santoro, and Anthony Ingraffea." *Climatic Change* 113 (2) (July): 525–535. http://link.springer.com/article/10.1007%2Fs10584-011-0333-0#.

EIA. 2011a. *Review of Emerging Resources: U.S. Shale Gas and Shale Oil Plays*. Washington, DC: U.S. Energy Information Administration, U.S. Department of Energy. http://www.eia.gov/analysis/studies/usshalegas/pdf/usshaleplays.pdf.

EIA. 2011b. *Annual Energy Outlook 2011—With Projections to 2035*. Washington, DC: U.S. Energy Information Administration, Office of Integrated and International Energy Analysis, U.S. Department of Energy. http://www.columbia.edu/cu/alliance/documents/EDF/Wednesday/Heal_material.pdf.

EIA. 2012. *Effect of Increased Natural Gas Exports on Domestic Energy Markets*. Washington, DC: U.S. Energy Information Administration, U.S. Department of Energy. http://energy.gov/sites/prod/files/2013/04/f0/fe_eia_lng.pdf.

Howarth, Robert W., Renee Santoro, and Anthony Ingraffea. 2011. "Methane and the Greenhouse-Gas Footprint of Natural Gas from Shale Formations." *Climatic Change* 106 (4): 679–690. http://link.springer.com/content/pdf/10.1007%2Fs10584-011-0061-5.pdf.

Hultman, Nathan, Dylan Rebois, Michael Scholten, and Christopher Ramig. 2011. "The Greenhouse Impact of Unconventional Gas for Electricity Generation." *Environmental Research Letters* 6 (4) (October–December): 044008 (9pp). http://iopscience.iop.org/1748-9326/6/4/044008/fulltext/.

Jacoby, Henry D., Francis M. O'Sullivan, and Sergey Paltsev. 2011. "The Influence of Shale Gas on U.S. *Energy and Environmental Policy*." Economics of Energy and Environmental Policy 1 (1): 37–51. http://globalchange.mit.edu/files/document/MITJPSPGC_Reprint_12-1.pdf.

Jiang, Mohan, W. Michael Griffin, Chris Hendrickson, Paulina Jaramillo, Jeanne VanBriesen, and Aranya Venkatesh. 2011. "Life Cycle Greenhouse Gas Emissions of Marcellus Shale Gas." *Environmental Research Letters* 6 (3): 034014 (9pp). http://iopscience.iop.org/1748-9326/6/3/034014/pdf/1748-9326_6_3_034014.pdf.

Skone, Timothy J., James Littlefield, and Joe Marriott. 2011. *Life Cycle Greenhouse Gas Inventory of Natural Gas Extraction, Delivery and Electricity Production*. Report No. DOE/NETL-2011/1522. Morgantown, WV: National Energy Technology Laboratory, U.S Department of Energy. http://www.netl.doe.gov/energy-analyses/pubs/NG-GHG-LCI.pdf.

Stephenson, Trevor, Jose Eduardo Valle, and Xavier Riera-Palou. 2011. "Modeling the Relative GHG Emissions of Conventional and Shale Gas Production." *Environmental Science and Technology* 45 (24): 10757–10764. http://pubs.acs.org/doi/pdf/10.1021/es2024115.

Venkatesh, Aranya, Paulina Jaramillo, W. Michael Griffin, and H. Scott Matthews. 2011. "Uncertainty in Life Cycle Greenhouse Gas Emissions from United States Natural Gas End-Uses and its Effects on Policy." *Environmental Science and Technology* 45 (19): 8182–8189. http://pubs.acs.org/doi/pdf/10.1021/es200930h.

Wigley, T. 2011. "Coal to Gas: The Influence of Methane Leakage." *Climactic Change* 108 (3) (October): 601–608. http://link.springer.com/content/pdf/10.1007%2Fs10584-011-0217-3.pdf.

EFFECTS OF SPACE DEBRIS ON SPACECRAFT

Joel E. Williamsen

In this paper, new equations are presented for predicting holes and cracks that would be produced following meteoroid or orbital debris penetration of International Space Station (ISS) modules. When these new hole- and crack-size equations are used in survivability assessments, the fidelity of the ISS risk predictions are expected to increase dramatically.

Nomenclature

A = empirical constant controlling the maximum (asymptotic) hole size or crack length values

B = empirical constant controlling the effect of obliquity on hole size or crack length values

C = empirical constant controlling the rate of increase of hole size or crack length growth towards its asymptotic value

D_{bl} = ballistic limit of shield type, inches

D_h = predicted hole diameter, inches

L_{tt} = predicted maximum tip-to-tip crack length, inches

S = standoff distance from the front of the bumper to the back of the rear wall

S_2 = standoff distance from the back of the intermediate bumper to the back of the rear wall

θ_p = obliquity of impacting particle to shield face, radians

Vp = velocity of the impacting particle, km/sec

The approach used to compute and reduce the consequences of a meteoroid (and later, an orbital debris particle) penetration and its link to catastrophic failure (defined here as a crew fatality) has changed over time. The advent of large space structures, such as the ISS, has allowed scenarios in which many meteoroid and orbital debris (MOD) penetrations could be survivable. As the orbital debris population and the associated penetration threat increased in the 1980s and early 1990s, National Aeronautics and Space Administration (NASA) engineers began to develop a tool that would allow them to determine the percentage of ISS penetrations that might be survivable for the crew and the ISS and to assess possible changes in station design or operations to improve that percentage.

The advent of large space structures, such as the ISS, has allowed scenarios in which many meteoroid and orbital debris (MOD) penetrations could be survivable.

Given these developments, NASA developed the *MSCSurv* computer code for quantifying a so-called "R factor": the ratio (R) of orbital debris penetrations that would cause either one or more crew losses or the long-term (potentially irrevocable) loss of spacecraft habitability to all orbital debris penetrations (Williamson 1994). The overall probability of no catastrophic failure (PNCF) is computed using the following equation:

$$PNCF = PNP^{R}, (1)$$

where PNP, the probability of no penetration, is given by

$$PNP = \exp(-N). (2)$$

PNP, which is determined using *Bumper*,[1] is a function of particle flux, module surface area, exposure time, and shield ballistic limit (Christiansen 2003). NASA has recently expended significant effort (Hyde and Christiansen 2007) to review *Bumper* and benchmark it to other MOD risk assessment codes used by some ISS international partners.

The R factor is also a function of the parameters noted previously, plus hypervelocity impact (HVI) damage level, crew operating parameters, and ISS equipment characteristics. By altering the input parameters regarding crew operations, internal arrangement of the ISS modules, and other design factors, an analyst can compare the safety of various existing or proposed modes of ISS operation.

A key component of the process, *MSCSurv*, follows to calculate R and requires use of damage prediction equations to calculate hole size and crack length following an on-orbit module wall penetration by an orbital debris particle (Evans, Blacklock, and Williamsen 1997). Considering the significance of these calculations, it is imperative that the equations used are as accurate as possible.

NEW GENERIC HOLE- AND CRACK-SIZE PREDICTION EQUATIONS

Figure 1, which shows a new generic-hole-diameter-vs.-projectile-diameter curve, is broken up into three regions. Each region corresponds to a certain type of projectile response and pressure wall hole growth pattern. The first region is shaded to indicate where hole diameter modeling is currently available. The shape of the curve shown in each of these three regions is based on the following considerations:

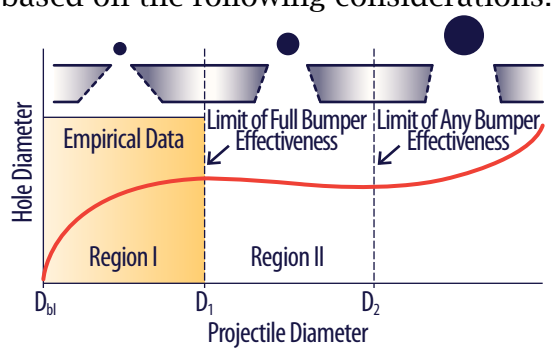


Figure 1. Generic Pressure Wall Hole Diameter as a Function of Projectile Diameter.

Source: Williamsen and Schonberg 2012.

- Initially, the hole diameter (and the cracking) phenomena are governed by the nature of the debris cloud loading on the module pressure wall. This case corresponds to Region I of the curve shown

[1] *Bumper* is a code NASA and contractors have used since the early 1990s to perform meteoroid and debris risk assessments.

in Figure 1. In this region, the projectile is completely shattered upon impact, and the degree of fragmentation increases with increasing projectile diameter. As a result, spread of the debris cloud created by the initial impact also increases, as does the effective diameter of the hole in the pressure wall.

- However, at a certain projectile diameter (labeled D_1 in Figure 1), the projectile is too large for it to be completely shattered by the outer bumper or shielding system. Hence, for projectile diameters beyond this point (i.e., in Region II), the amount of projectile fragmentation decreases with increasing projectile diameter, as does the spread of the debris cloud and the size of the hole in the pressure wall.

- Finally, in Region III (i.e., for $D_p >> D_2$), the projectile diameter is very large when compared to the ballistic limit diameter value, perhaps even as large as the spacing between the bumper and the pressure wall. In such cases, the multi-wall system can be expected to act as a single thin plate. Hence, the form of that part of the curve should be similar to that generated by single-plate hole-diameter equations. In this case, the pressure wall hole diameter again increases with increasing projectile diameter.

The new Williamsen-Schonberg (W-S) model consists of a single hole-size and single crack-size equation that can be applied to all of the ISS wall configurations tested previously, as well as other configurations within a similar range of shield design parameters (such as wall thicknesses, bumper areal densities, and standoff distances). Each hole- and crack-size equation consists of three parts: (1) a data-based equation for Region I of Figure 1, (2) an interpolation equation for Region II between the data-based equation for Region 1 and the single-wall equation for Region III, and (3) a single-wall equation for Region III that begins at that projectile diameter where the bumper ceases to be effective in fragmenting an impacting projectile.

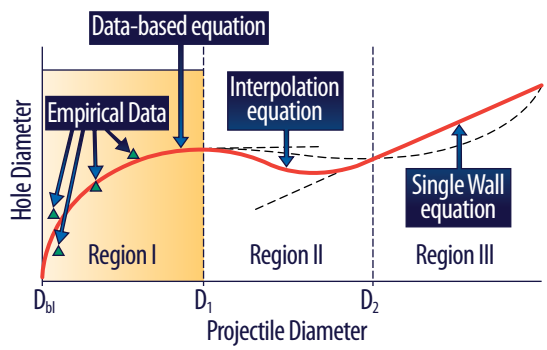


Figure 2. Representation of a Three-Part Hole-Diameter or Crack-Length Equation.

Figure 2 presents a sketch of this three-part equation (thick solid line) that is intended to model the response as outlined previously (still shown as the thinner line with dashes and dots). This figure also includes some generic empirical data to support the premise that the first part of each three-part equation is empirically based. The thinner dashed lines are shown only to indicate extensions or precursors of the data-based and single-wall equations, respectively, and are not actually used by the model in the regions where they are drawn.

REGION I

The following considerations, based on analysis of several hundred perforated rear walls from dual- and multi-wall target systems impacted by

high-speed projectiles, were used in determining the form of the equation for this region:

- **Effect of bumper-to-rear-wall stand-off.** The larger the stand-off distance between the outer bumper and the rear wall, the bigger the resulting hole or crack.

- **Effect of rear wall thickness.** Rear-wall thickness affects the rate of growth of the hole as well as its final size—the thicker the rear wall, the smaller the hole. A thin rear wall also results in a big hole rather quickly. Rear walls of multi-wall systems that have strong or massive intermediate bumpers fail after bulging first, not by piercing or by through-holes. As a result, in such systems, a big hole happens fairly quickly once the ballistic limit is exceeded.

- **Effect of bumper thickness and intermediate bumper.** As the ratio of projectile-diameter-to-bumper thickness increases, the debris cloud becomes increasingly concentrated. The more massive an intermediate shield, the larger the hole or crack, and the further away from the rear wall it is located, the less of an effect it has (i.e., a light, thin intermediate bumper far away from the rear wall has little or no effect, while a more massive intermediate bumper closer to the rear wall would have more of an effect, producing larger holes in the rear wall). A more massive intermediate bumper decreases obliquity effects (see next bullet item) because as debris clouds move through intermediate bumpers, their trajectories become more and more normal, which allows them to act more in concert again.

- **Effect of trajectory obliquity.** Oblique impacts without intermediate bumpers tend to drive down hole size and crack length. This effect is amplified in walls with large standoff distances. The larger standoff distance allows the in-line and normal debris clouds that were created by the initial impact on the bumper to separate, which, in turn, causes the more damaging particles in the debris clouds to act more independently (i.e., without the damage enhancement of the fuller impulse that would exist if the two clouds acted together), resulting in smaller holes.

- **Effect of impact velocity.** The obliquity effect is more pronounced at lower impact velocities. As velocity increases, the obliquity effect is less and less. At higher velocities, rear-wall failure is from bulging and not from piercing or by through-holes. Again, bulging failures make bigger holes more quickly.

Based on these considerations, the following equation forms were used to model hole diameter and crack length in Region I, as indicated previously in Figure 1.

Hole diameter in Region I:

$$D_h = A_h (V_p/6.5)\cos^{B_h}\theta_p\{1-\exp[-C_h(D_p/D_{BL}-1)]\} \quad (3)$$

Crack length in Region I:

$$L_{tt} = A_L (V_p/6.5)\cos^{B_L}\theta_p\{1-\exp[-C_L(D_p/D_{BL}-1)]\} \quad (4)$$

Expressions for A_h, B_h, C_h, A_L, B_L, and C_L were obtained by first determining an appropriate baseline set of values for these parameters for wall systems considering 6.5 km/s impacts and then adjusting these baseline values to fit the test data for the other wall systems according to the phenomenological considerations outlined previously.

REGION II

In this region, the hole-diameter and crack-length equations for Region I are extended to larger projectile diameters to account for the decreasing effectiveness of bumpers in breaking up the debris cloud. The hole-size values between $D_p = D_1$ (i.e., the Region I/Region II interface) and $D_p = D_2$ (i.e., the Region II/Region III interface) are obtained by interpolating between the values predicted by empirical equation developed for Region I (and extended into Region II) and the values obtained using a single-wall hole-size prediction equation (as described in the following section). The values of D_1 and D_2 are likely to be dependent on the material properties and geometric parameters of the shield design and on impact velocity and trajectory obliquity.

The boundary between Region I and Region II is presumed to occur when the projectile is too large for it to be shattered completely by the outer bumper or shield. A review of the literature on the effectiveness of fragment projectiles on thin plates revealed that aluminum bumper plates begin to be effective in protecting a rear wall against perforation by aluminum projectiles as the t_b/D_p ratio is increased above ~ 0.05 (with an optimal t_b/D_p ratio of ~ 0.25). For this study, it is assumed that the Region I/II boundary occurs at D_p/t_b = ~ 20.

Beyond $D_p = D_2$, the projectile is so large relative to the thickness of the bumper that the bumper has hardly any effect on the impact projectile. In such a case, the diameter of the hole in the bumper plate is likely to be just a fraction larger than the diameter of the projectile itself. For example, if we assume that the ineffectiveness of the bumper is said to occur when the hole in the bumper plate is only 10 percent larger than the diameter of the projectile, then invoking the following equation from Gehrig (1970) for the size of a hole in a thin plate, we have:

$$1.1D_p = 0.45D_p[V_p(t_b / D_2)^{2/3} + 2] \quad (5)$$

Solving for D2 yields

$$D_2 / t_b = (V_p / 0.444)^{3/2} \quad (6)$$

Therefore, for an impact velocity of 6.5 km/s, D_2/t_b = ~ 55. Thus, for a given value of t_b, the corresponding value of D_2 can be readily calculated. Further study could undoubtedly lead to improvements of the assumed values of D_1/t_b and D_2/t_b.

REGION III

In this region, the hole-diameter and crack-length equations for Region II are extended to larger projectile diameters. If we restrict our discussions to normal impacts, the following equation can be used to predict the hole diameter for the initial impact on the bumper and for the subsequent impact on the inner or rear wall of the dual-wall system (Maiden and McMillian 1964):

$$D_h = 0.45D_p\left[V_p\left(t_b/D_p\right)^{2/3} + 2\right]. \quad (7)$$

COMPARISON TO PRIOR DAMAGE PREDICTION EQUATIONS IN REGION I

Figures 3 and 4 present comparisons of hole-size and crack-length predictions as given by the previous Schonberg-Williamsen (S-W) model and the new W-S model presented herein. Also presented in these figures are the data used to derive the empirical S-W model equations so that we can also see how well the new W-S model equations compare against them. The results shown in these figures are representative of those seen in the full suite of comparative plots, which can be found in Squire et al. (2011).

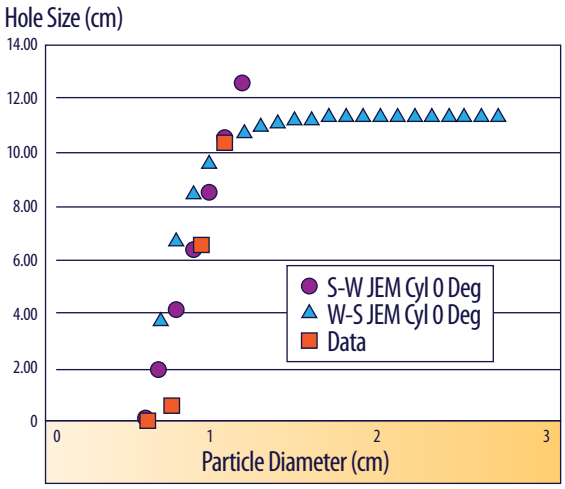


Figure 3. Predictions of Hole Size, JEM Cylinder Wall System, V = 6.5 km/s, 0-deg Trajectory Obliquity.

In these figures, we see that the S-W model either did not fit the data well (e.g., Figure 4), or, in those cases where it did, its plot was irregular in that it did not show the expected asymptote as projectile diameter increased beyond the ballistic limit (e.g., Figure 3). In these cases, the new W-S model either fit the data just as well or better but also displayed the proper asymptotic behavior. Taking these results and others like it under consideration, we believe that the W-S hole and crack size model is an improvement over the S-W model used in previous versions of *MSCSurv*.

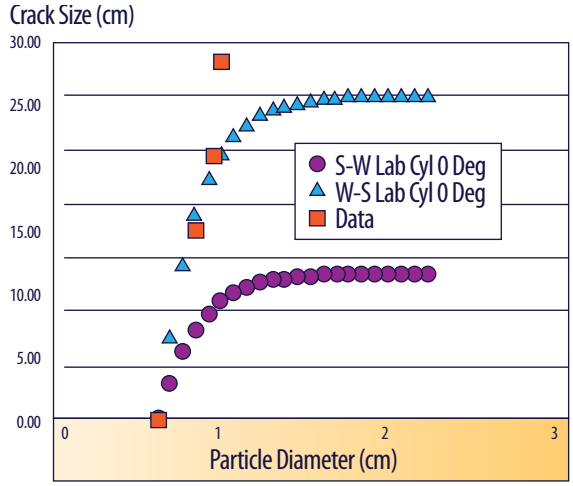


Figure 4. Predictions of Crack Length, U.S. Lab Cylinder Wall System, V = 11 km/s, 45-deg Trajectory Obliquity.

CONCLUDING COMMENTS

NASA currently uses the *MSCSurv* computer code in conjunction with *Bumper* to calculate the PNCF of the ISS due to MOD particle impact. As part of this calculation, *MSCSurv* must consider numerous assumptions regarding crew and ISS response to an impact. One of the most important of these considerations is the calculation of a hole size and a crack length following an on-orbit penetration for each of the many different ISS shield types. In this paper, we have presented a new model for calculating these quantities. Based on the results obtained thus far, we believe that the new hole- and crack-size model encoded will dramatically improve the fidelity of the PNCF predictions by *MSCSurv*.

ACKNOWLEDGMENTS

The author would like to acknowledge the support provided by the NASA Engineering and Safety Center and the NASA Johnson Space Center that made this work possible.

Dr. Williamsen holds a Ph.D. in Systems Engineering from the University of Alabama, Huntsville.

The original article was presented as an American Institute for Aeronautics Paper for the AIAA Structures, Dynamics and Materials Conference, August 2012.

SPACECRAFT MODULE HOLE-SIZE AND CRACK-LENGTH PREDICTION EQUATIONS FOLLOWING A PENETRATING DEBRIS PARTICLE IMPACT

http://www.sciencedirect.com/science/article/pii/S1877705813009107

REFERENCES

A Study of the Probability of Catastrophic MMOD Penetration of the Docking Compartment Pressurized Hull. 2005. Technical Report No. II-38332-311/SS40950. Moscow, Russia: Russian Space Agency.

Burch, G. T. 1967. *Multiplate Damage Study*. Technical Report No. AFATL-TR-67-116. Valparaiso, FL: Eglin Air Force Base, Air Force Armament Laboratory (AFATL).

Christiansen, Eric L. 2003. *Meteoroid/Debris Shielding*. Report No. TP–2003-210788. Houston, TX: NASA Johnson Space Center. http://ston.jsc.nasa.gov/collections/TRS/_techrep/TP-2003-210788.pdf.

Evans, H., K. Blacklock, and J. Williamsen. 1997. *Manned Spacecraft & Crew Survivability (MSCSurv) Version 4.0 User's Guide*. Report No. 651-001-97-006. Huntsville, AL: Sverdrup Technology, Inc.

Gehring, John W., Jr. 1970. "Theory of Impact on Thin Targets and Shields and Correlation with Experiment." Chap. 4 in *High-Velocity Impact Phenomena*, edited by Ray Kinslow and A. J. Cable, 105–156. New York: Academic Press.

Hyde, J. L., and E. L. Christiansen. 2007. *Bumper-II Micrometeoroid and Orbital Debris Threat Assessment Code: Estimation of Orbiter Uncertainty Bounds v2.0*. Report No. JSC-63999, Rev A. Houston, TX: NASA Johnson Space Center.

Maiden, C. J. and A. R. McMillan. 1964. "An Investigation of the Protection Afforded a Spacecraft by a Thin Shield." *AIAA Journal* 2 (11): 1992–1998.

Schonberg, W. P. 1995. *Pressure Wall Hole Size and Maximum Tip-to-Tip Crack Length Following Orbital Debris Penetration*. Final Report, NASA/ASEE Summer Faculty Fellowship Program. Huntsville, AL: Marshall Space Flight Center.

Schonberg, William P., Alan Bean, and Kent Darzi. 1991. *Hypervelocity Impact Physics*. Report No. NASA CR-4343. Huntsville, AL: Marshall Space Flight Center.

Squire, M. et al. 2011. *Lightweight Installable Micrometeoroid and Orbital Debris (MMOD) Shield Concepts for International Space Station (ISS) Modules*. Report No. NESC-RP-09-00593. Washington, DC: NASA Engineering and Safety Center.

Williamsen, J. E. 1994. *Vulnerability of Manned Spacecraft to Crew Loss from Orbital Debris Penetration*. NASA Technical Memorandum 108452. Huntsville, AL: NASA Marshall Space Flight Center.

Williamsen, J. E., and W. P. Schonberg. 2012. "An Improved Prediction Model for Spacecraft Damage Following Orbital Debris Impact." Paper presented at the *53rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, Honolulu, HI, April 23–26.

Williamsen, Joel, Donald Grosch, and William Schonberg. 1996. "Empirical Prediction Models for Hole and Crack Size in Space Station Shielding from 6 to 12 km/s," In Proceedings of SPIE, Volume 2813, *Characteristics and Consequences of Orbital Debris and Natrual Space Impactors*, eds. Timothy D. Maclay and Firooz A. Alladadi, 169–180. doi:10.1117/12.256060.

Past Issues

Acquisition

- Defining Acquisition Trade Space Through "DERIVE"
- Supporting Acquisition Decisions in Air Mobility
- Assessing System Reliability with Limited Flight Testing
- Promise, Reality, and Limitations of Software Defined Radios
- Implications of Contractor Working Capital on Contract Pricing and Financing
- The Mechanisms and Value of Competition
- Initiation and Early Management of Acquisition Programs

Security in Africa

- Trends in Africa Provide Reasons for Optimism
- China's Soft Power Strategy in Africa
- Sudan on a Precipice
- A New Threat: Radicalized Somali-American Youth
- Chinese Arms Sales to Africa
- Outsourcing Imagination: The Potential of Informal Engagement Networks in Africa
- Defense Environmental Cooperation with South Africa

Challenges in Cyberspace

- Cyberspace – The Fifth and Dominant Operational Domain
- Transitioning to Secure Web-Based Standards and Protocols
- Information Assurance Assessments for Fielded Systems During Combatant Command Exercises
- Supplier-Supply Chain Risk Management
- Internet-Derived Targeting: Trends and Technology Forecasting
- Training and Educating the DoD Cybersecurity Workforce

Today's Security Challenges

- A Framework for Irregular Warfare Capabilities
- Bridging the Interagency Gap for Stability Operations
- Developing an Adaptability Training Strategy
- Force Sizing for Stability Operations
- Planning Forces for Steady State Foreign Internal Defense and Counterinsurgency
- Test and Evaluation for Rapid-Fielding Programs
- Understanding Security Threats in East Africa
- Supporting Warfighting Commands
- Detecting Improvised Explosive Devices
- Building Partner Capacity
- Combating the Trans-South Atlantic Drug Trade
- Countering Transnational Criminal Insurgents
- Using Economic and Financial Leverage
- Understanding the Conflict in Sudan

Resource Analyses

- Evaluating the Costs and Benefits of Competition for Joint Strike Fighter Engines
- Analysis and Forecasts of TRICARE Costs
- Cost Savings from the Post-Cold War Consolidation of the Defense Industrial Base: A Case Study of the Shipyards
- The Effects of Reserve Component Mobilization on Employers
- Does DoD Profit Policy Sufficiently Motivate Defense Contractors?
- Auctions in Military Compensation

Focusing on the Asia-Pacific Region

- Making American Security Partners Better Resource Managers
- Collaborating with Singapore to Better Understand the U.S. and Asian Defense Environments
- Intellectual Outreach to the Muslim World: The Council for Asian Terrorism Research
- Inside North Korea
- Red Teaming for Terminal Fury
- Promoting Interagency Cooperation in Shaping U.S.-China Relations
- Extending Trilateral Cooperation in Dealing with Disaster
- Developing Human Capital in China – Implications for the United States

Homeland Security

- Port Vulnerability
- Assessing the EMP Threat
- Homeland Defense Scenarios
- Transport and Dispersion Models
- IT Security

IDA | RESEARCH NOTES
© Institute for Defense Analyses
4850 Mark Center Drive • Alexandria, VA 22311-1882
www.ida.org