Proceedings of the 2007 Winter Simulation Conference
S. G. Henderson, B. Biller, M.-H. Hsieh, J. Shortle, J. D. Tew, and R. R. Barton, eds

AGENT-MODEL VALIDATION BASED ON HISTORICAL DATA

Lance E. Champagne

Joint Data Link Information Combat Execution
5930 Devlin Dr., Bldg 585
Nellis AFB, NV 89191, U.S.A.

Raymond R. Hill

Wright State University
Dept. of Biomed., Ind., and Human Fact. Engr.
207 Russ Engineering Center
3640 Colonel Glenn Hwy.
Dayton, OH 45435-0001, U.S.A.

ABSTRACT

Combat, unlike many real-world processes, tends to be singular in nature. That is, there are not multiple occurrences from which to hypothesize a probability distribution model of the real-world system. Mission-level models may offer more flexibility on some measures due to their extended time frame. Additionally, the parameters involved in the mission-level model may be unchanged for significant stretches of the total simulation time. In these cases, time periods may be devised so that the periods hold sufficiently similar traits such that the incremental results may be assumed to come from a common distribution. This paper details a new statistical methodology for use in validating an agent-based mission-level model. The test is developed within the context of the Bay of Biscay agent-based simulation and uses the monthly data from the extended campaign as a basis of comparison to the simulation output.

1 INTRODUCTION

Agent-based simulations have become increasingly useful for studying various aspects of combat. Tighe (1999), developed an agent-based simulation using the boids flocking algorithm (Levy 1992) and ISAAC (Ilachinski 1998) for quantifying strategic effects, purported to be one of the main strengths of air power in combat. Bullock (2000) introduced the Hierarchical Interactive Theater Model (HITM). This model was intended to provide a sufficiently complex tool able to show strategic effects of air power, while retaining enough simplicity to allow identification of interactions between important factors. These results are summarized in Hill et al. (2003) Other agent-based combat simulation research includes modeling riot tactics for small military units (Woodaman 2000), small unit peacekeeping tactics in an urban environment (Brown 2000), and a German training scenario involving small units over a relatively short time period (Erlenbruch 2002).

Though each of the above provides significant results toward advancing the field of agent-based combat simulation, no attempt was made to relate simulation outcome to real-world data. This paper outlines the development of a statistical validation technique applied to an agent-based combat simulation based on the Allied offensive against the German U-Boats in the Bay of Biscay during WW II. Model results are compared to the historical data.

1.1 Historical Scenario

German U-Boats operated against Allied shipping in the North Atlantic from 1941 through the end of the war in an effort to reduce the shipments of war-time supplies to Great Britain. Many of these submarines operated from ports in occupied France, crossing the Bay of Biscay into the North Atlantic, where they hunted for Allied transport ships. Once they left the Bay of Biscay, the U-Boats could, for all practical purposes, operate outside the reach of Allied aircraft support. These U-boat efforts put the Allied war effort at great risk.

The Bay of Biscay area afforded the Allies a chance to counter the U-boat threat. In 1941 a concerted effort to hunt down and destroy U-boats transiting the Bay of Biscay began. The resulting Allied successes are well documented and have long since been a source of operational research case studies, although rarely have these studies involved agent models.

Additional historical background on the offensive search in the Bay of Biscay can be found in (McCue 1990), and an extensive record of the corresponding operational analysis may be found in (Waddington 1973) and (Morse and Kimball 1998).

1.2 Bay Of Biscay Model

A Bay of Biscay simulation was built to reproduce the results of the historical operation in both qualitative and quantitative measures. As with any model, simplifying assumptions were made. These are detailed in Champagne (2003). Despite these simplifying assumptions, the resulting model must accurately reflect the modeled events. This validation requirement is the focus of this paper.

1.3 Simulation Scenarios

Two scenarios were chosen for the initial model validation. The first was the six month period from October 1942 – March 1943 (Scenario 1), and the second was April 1943 – September 1943 (Scenario 2). Although the technologies and procedures used in these periods differed, within each period considered both aspects were fairly stable. Such consistency is important for validating simulation output to historical records.

The U-Boat fleet initially consists of 70 agents distributed randomly and uniformly throughout the Bay of Biscay. A simulation warm up period of 12 months is used to position the fleet. U-Boat fleet reinforcements was modeled according to historical numbers (McCue 1990).

The aircraft fleet consists of 15 aircraft agents in Scenario 1 and 35 aircraft agents in Scenario 2, collocated at a single airbase in Great Britain. These aircraft numbers were derived based on the historical values for flying hours. Aircraft search defined grids using a modified barrier search pattern constructed from the tactics discussed in (Waddington 1973).

1.4 Our Validation Approach

When using historical data, convention leads one to determine whether simulation results adequately cover the historical results available. We next demonstrate this approach but also highlight our concerns with such approaches and present our bootstrap, non-parametric approach for agent-based simulation model validation.

2 SIMULATION RESULTS

The historical values for U-boat sightings and kills for each scenario are found in Tables 1 and 2, respectively. The simulation results for Scenario 1 MOEs are found in Tables 3 and 4, respectively, for all 20 replications. Tables 5 and 6 provide results for Scenario 2, respectively.

2.1 Analysis Of The Simulations MOEs

Joint confidence intervals around the simulation means can be constructed using a t-statistic, as shown in (1).

$$Bound = \bar{x} \pm \frac{s}{\sqrt{n}} \cdot t_{\frac{\alpha}{2 \cdot k}, n-1} \qquad (1)$$

where

- $\bar{x}$ is the sample mean
- s is the sample standard deviation
- n is the sample size
- k is the number of joint confidence intervals

(1 – α) is the desired level of joint confidence

Table 1: Historical MOE values for scenario 1 (McCue, 1990)

MOE	10/42	11/42	12/42	1/43	2/43	3/43
Sightings	18	19	14	10	32	42
Kills	1	1	0	0	0	1

Table 2: Historical MOE values for scenario 2 (McCue, 1990)

MOE	4/43	5/43	6/43	7/43	8/43	9/43
Sightings	52	98	60	81	7	21
Kills	1	7	4	13	5	2

Table 3: Simulated U-boat sightings for scenario 1

Rep.	10/42	11/42	12/42	1/43	2/43	3/43
1	9	17	21	17	11	33
2	19	14	25	24	24	23
3	16	23	15	22	25	28
4	20	17	21	33	26	33
5	15	16	18	25	28	26
6	18	21	20	29	23	32
7	11	20	24	30	34	28
8	20	17	17	25	28	23
9	27	25	34	40	28	30
10	17	17	26	30	33	45
11	9	9	23	13	21	27
12	15	17	27	34	27	39
13	12	14	18	21	17	25
14	12	15	15	26	21	27
15	13	17	16	24	25	36
16	22	14	16	16	27	25
17	21	15	23	17	21	23
18	22	21	22	21	27	36
19	21	28	32	30	24	21
20	13	15	22	27	27	26

.

Using a (1 – a) = 0.8, consistent with simulation validation literature (Balci and Sargent 1984, Balci 1994, Kleijnen 1995), confidence intervals were constructed around the simulation means for each scenario assuming a t-distribution with 19 degrees of freedom. The 80% joint confidence is maintained for each scenario.

Figure 1 shows the results from scenario 1, and the results from scenario 2 are shown in Figure 2. In each case, the confidence intervals either cover or nearly cover the MOE's historical value. This means that if the actual number of sightings and kills represent the mean of the true distribution for each scenario, then the simulation does a reasonable job of emulating the scenarios and statistically captures the actual values.

Table 4: Simulated U-boat kills for scenario 1

Rep.	10/42	11/42	12/42	1/43	2/43	3/43
1	0	0	0	1	0	1
2	0	0	1	1	2	1
3	0	0	1	1	0	1
4	0	0	1	0	1	1
5	0	0	1	1	2	0
6	0	0	0	1	1	0
7	0	0	1	2	1	1
8	0	1	0	0	1	1
9	1	0	2	1	1	0
10	1	1	2	1	1	0
11	1	1	0	1	1	0
12	1	0	1	0	1	0
13	1	0	1	0	0	0
14	0	0	0	0	1	1
15	0	0	1	1	1	1
16	2	1	0	0	1	0
17	0	0	1	1	1	0
18	0	0	2	1	0	2
19	0	1	1	2	0	1
20	0	1	1	0	1	1

Table 5: Simulated U-boat sightings for scenario 2

Rep.	4/43	5/43	6/43	7/43	8/43	9/43
1	38	50	44	46	45	64
2	48	46	49	57	62	70
3	46	43	46	43	57	69
4	46	48	51	56	69	48
5	40	49	48	69	70	69
6	60	46	67	70	58	57
7	50	46	66	57	59	63
8	42	52	46	54	74	79
9	43	60	47	62	70	75
10	46	53	54	72	75	73
11	40	44	49	68	56	55
12	36	59	51	67	63	58
13	44	29	47	52	55	55
14	35	40	49	45	71	48
15	44	44	57	73	58	58
16	42	58	54	61	60	68
17	42	47	62	69	71	66
18	43	59	56	79	74	65
19	48	53	47	64	72	60
20	41	45	57	61	59	75

The problem with this practice, and thus the concern with our results so far, is the strong assumption made that the historical data, a simple sample, is in fact the mean of the true underlying distribution of outcomes. This assumption may be quite tenuous.

Table 6: Simulated U-boat kills for scenario 2

Rep.	4/43	5/43	6/43	7/43	8/43	9/43
1	0	6	7	3	6	6
2	1	3	4	8	5	5
3	6	5	5	5	4	3
4	2	9	4	3	9	3
5	2	2	5	4	6	9
6	4	5	8	8	8	5
7	6	2	12	9	4	6
8	3	2	8	8	9	13
9	4	5	1	5	6	7
10	5	4	4	6	13	5
11	7	7	3	9	6	2
12	6	3	2	12	9	5
13	5	4	3	5	4	4
14	2	4	7	2	8	4
15	5	7	3	7	6	3
16	6	6	6	3	5	11
17	3	3	8	6	5	4
18	2	6	5	6	5	6
19	5	3	6	4	9	7
20	3	7	4	6	5	7

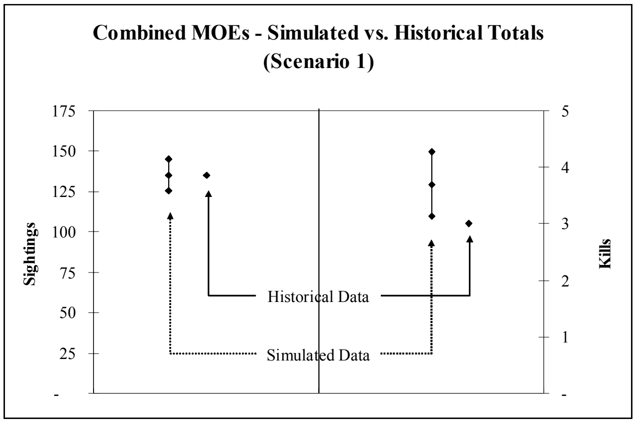

Figure 1: October 1942 – March 1943 MOEs

3 MOTIVATION FOR NEW TEST

Rather than use the historical data in the aggregate, examining Bay of Biscay historic outcomes by month, instead of aggregated, provides a convenient method for examining the variability of the real-world system. Mean monthly values for each MOE of interest, both real-world and simulated, can be calculated and compared. Although the resulting analysis provides additional insight, it still lacks quantifiable confidence to conclusions about the validity of the simulation.

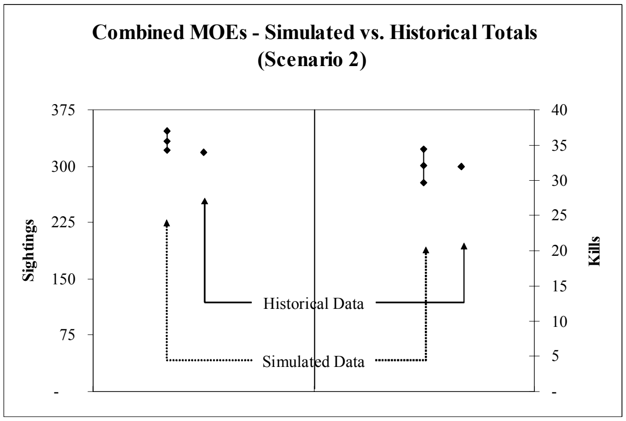

Figure 2: April 1943 – September 1943 MOEs

Figures 3 through 6 depict the historic versus simulated mean monthly MOE values via joint confidence intervals for each MOE, U-Boat sightings and kills, in both scenarios, respectively. Each figure shows 21 individual confidence intervals – the left-most being the historic value with the remaining 20 coming from each of 20 simulation iterations. Joint confidence intervals were constructed to allow an overall 80% joint confidence level (k = 2) for each comparison.

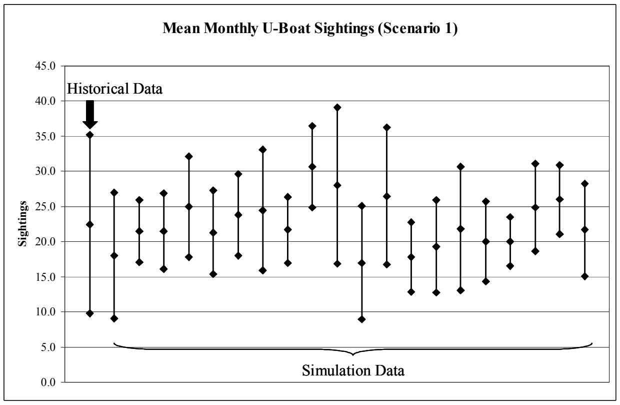

Figure 3: Comparisons of mean monthly U-boat sightings, historic vs. simulated scenario 1

Figures 3-6 indicate 100% confidence level overlap in each figure; face-level validation is supported. The analytical dilemma is that the overall confidence level is not 80%. To achieve meaningful 80% coverage would yield extremely wide confidence intervals given no practical insight.

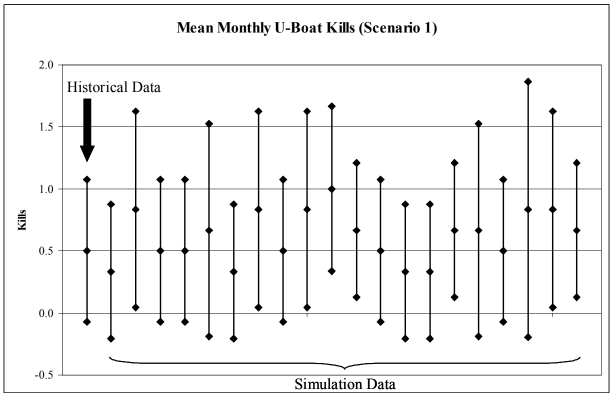

Figure 4: Comparisons of mean monthly U-boat kills, historic vs. simulated scenario 1

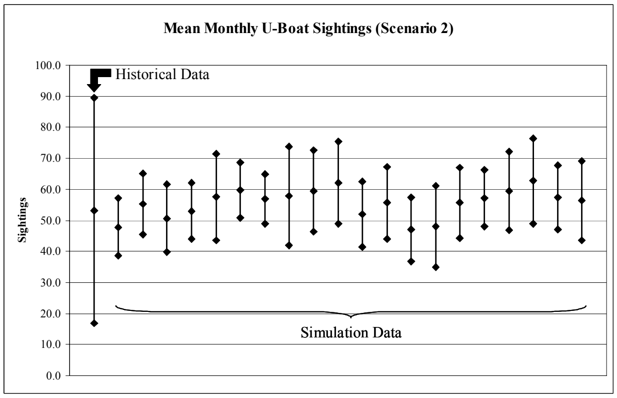

Figure 5: Comparisons of mean monthly U-boat sightings, historic vs. simulated scenario 2

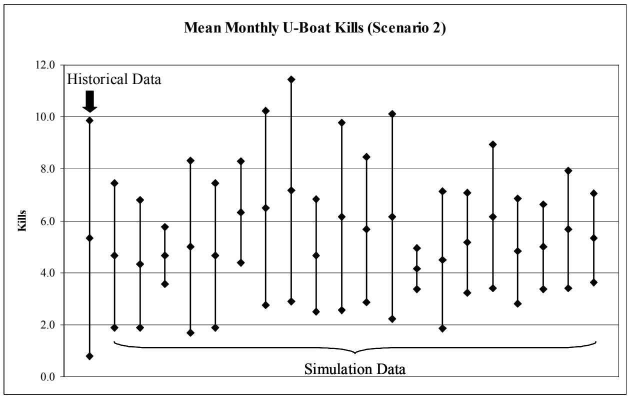

Figure 6: Comparisons of mean monthly U-boat kills, historic vs. simulated scenario 2

We next present and demonstrate a methodology that allows for statistically significant comparisons, despite having a single real-world sample.

3.1 Methodology For Comparison Of Historic Versus Simulated Data

Any test allowing a meaningful comparison between the historic outcome and the simulated data, while still providing insight into the underlying stochastic real-world system, requires two characteristics. First, the method must provide a means of deriving multiple samples from the stochastic process underlying the real-world system. Second, the method must provide a meaningful, quantifiable level of confidence in the result. Figure 7 illustrates an approach that meets both requirements.

Once the simulation results from *n* iterations are generated, the historic data is used to generate *n* bootstrap samples. A sign test is used to test the hypothesis that the two samples are statistically identical. The bootstrap and sign test is then replicated for multiple experiments.

3.1.1 Bootstrap

Several statistical resampling techniques have been developed to provide estimators of population parameters that are difficult or impossible to treat theoretically (Conover 1999) or when obtaining multiple samples from a system is prohibitively expensive (Cheng 2001). Resampling is based on the idea that when one random sample is available and obtaining another sample is not feasible, then the best estimate for the distribution under study is the random sample in-hand.

Efron (1979) first proposed the bootstrap method of resampling. Since it was first proposed, the method has found wide acceptance and applicability. Efron and Tibshirani (1986) review the bootstrap method and its applications.

The Method. Consider the statistic q calculated from the random sample X = {X_1, X_2, …, X_n}. A bootstrap sample X^* = {X_1^*, X_2^*, …, X_n^*} is generated by taking a random sample from X, where

$P(X_j^*(j=1,2,...,n) = X_i(i=1,2,...,n)) = \frac{1}{n}$, for which q^*, an estimator for q, is computed from the bootstrap sample. If some number, B, Monte Carlo replications are taken, then the distribution of q can be estimated by the sample mean and standard deviation of q^*.

Sample Size, B. The number of bootstrap samples needed to accurately estimate the properties of the sample statistic vary. Efron and Tibshirani (1986) note that for most situations, B = 50 to 200 is "quite adequate," though 250 or more are often needed for accurate computation of confidence intervals. Conover (1999) adds that "as few as 25 replications can be very informative".

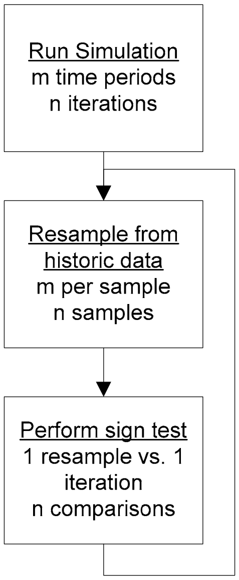

Figure 7: Methodology for comparisons of a single-sampled real-world process to simulated results

Proposed Use. The bootstrap used differs slightly for the proposed methodology. Instead of a single collection of bootstrap samples of the historic data, m groups of b bootstrap samples were generated for comparison with the simulation, where b = the number of simulation iterations and m = number of sign test trials desired.

3.1.2 Sign Test

The sign test is used to test whether one random variable in a pair (X, Y) tends to be larger than the other random variable in the pair. It is a variant of the binomial test in which the probability of outcome is assumed to be equally likely, p = 1 – p = 0.5 (Conover 1999).

Data for the sign test consists of *n'* pairs of observations (X_1, Y_1), (X_2, Y_2), …, ($X_{n'}$, $Y_{n'}$), each observation being a bivariate random sample. Within each (X_i, Y_i) observation, a comparison is made, and the pair is classified as "+" if $X_i < Y_i$, "–" if $X_i > Y_i$, or "0" if $X_i = Y_i$. The test statistic, *T*, is the number of "+" pairs. The null distribution of *T* is the binomial distribution with *p* = ½ and *n* = number of non-tied pairs (tied pairs are disregarded).

The sign test assumes that the bivariate pairs are mutually independent, and the probability of outcome is constant for all trials. It further assumes that the measurement scale within each pair is at least ordinal, that is each (X_i, Y_i) pair may be determined to be "+", "–", or "0". Finally, the sign test assumes there is internal consistency between the observed pairs.

For model validation purposes, the two-tailed test is desired. That is,

$$H_0: P(+) = P(-)$$
$$H_1: P(+) \neq P(-).$$

The critical a-values are determined for each test once *n* has been determined. Because the binomial distribution is discrete, the critical α-values cannot be arbitrarily set. Instead, the critical α-level is selected such that the total $(1 - \alpha)$ level is as close to 0.9 as possible, without being less than 0.9, given a particular *n*. That is, H_0 is rejected if the p-value for the test is less than 0.05.

3.2 Bay Of Biscay Agent-Based Simulation Results

The presentation of results follows the same order as in the previous analyses.

Each MOE was subjected to identical experiments. Each experiment consists of twenty sign tests ($m = 20$), with each sign test incorporating twenty (one per simulation iteration) bootstrap samples ($b = 20$). For each MOE, one sign test is presented in detail, and the remaining tests are summarized prior to validation discussions.

3.2.1 Scenario 1 MOEs

Previous analyses of Scenario 1 MOEs provided a somewhat mixed picture of the simulation's fidelity with respect to the historic data.

Table 7 shows the bootstrap samples for Scenario 1 U-Boat sightings generated for comparison with the simulation results.

Table 8 summarizes the sign test classifications for the paired data (X_i, Y_i) for Scenario 1 U-Boat sightings, where X_i is the i^{th} bootstrap U-Boat sightings total and Y_i is the U-Boat sightings total from the i^{th} simulation iteration. The sign test statistic *T* and number of non-tied pairs *n* are displayed as well.

For n = 19, $P(t \leq 5) = 0.0358$ and $P(t \geq 13) = 0.0358$ defining an overall $(1 - \alpha) = 0.9284$. Since $5 < T = 9 < 13$, there is insufficient evidence to reject H_0. There is no compelling evidence to suggest the simulation does not faithfully represent the real-world system with respect to Scenario 1 U-Boat sightings.

Table 9 shows the bootstrap samples of Scenario 1 U-Boat kills generated for a single replication of the bootstrap/sign test experiment. Of the 20 sign test trials, the p-values ranged in value from 0.011 to 0.5. Under the rejection criteria, the null hypothesis was rejected in 5 of the 20 trials.

Table 7: Bootstrap U-boat sightings – scenario 1

Trial	10/42	11/42	12/42	1/43	2/43	3/43
1	14	18	10	42	42	42
2	18	14	42	18	19	18
3	18	18	19	18	19	14
4	10	14	14	14	42	14
5	14	19	42	32	42	19
6	42	18	32	32	42	14
7	19	32	14	32	18	19
8	18	14	14	10	14	42
9	18	19	18	42	18	19
10	32	32	32	32	18	18
11	32	10	19	14	10	32
12	10	19	42	32	10	32
13	32	19	19	42	18	18
14	32	32	42	42	42	10
15	10	32	14	18	18	32
16	32	32	10	18	42	14
17	19	19	14	19	19	32
18	32	19	42	18	32	14
19	10	19	19	32	32	32
20	32	42	10	32	42	14

Table 8: Sign test calculations – U-boat sightings, scenario 1

Observation	1	2	3	4	5	6	7	8	9	10
Sign	–	0	+	+	–	–	+	+	+	+
Observation	11	12	13	14	15	16	17	18	19	20
Sign	–	+	–	–	+	–	–	–	+	–
T	9									
n	19									

Both sign test experiments tend to indicate that the simulation is representative of historical combat operations for Scenario 1. In the case of Scenario 1 U-Boat sightings, the bootstrap/sign test rejected the null hypothesis in 15% of the trials. With respect to Scenario 1 U-Boat kills, the bootstrap/sign test method rejected the null hypothesis in 25% of the trials. Rather than make a validation conclusion based on a single statistical pass/fail, as in the first analysis method, the bootstrap/sign test methodology provides a broader context to the simulation results. These conclusions provide stronger rationale than either of the previous tests for accepting the model as valid with respect to the MOEs.

3.2.2 Scenario 2 MOEs

Previous analyses of Scenario 2 MOEs also provided a somewhat mixed picture of the simulation's fidelity with respect to the historic data.

Table 10 shows the bootstrap samples for Scenario 2 U-Boat sightings generated for a single replication of the bootstrap/sign test experiment.

Of the 20 sign test trials, the p-values ranged in value from 0.058 to 0.412. Under the rejection criteria, the null hypothesis was not rejected in any of the 20 trials.

Table 11 shows the bootstrap samples of Scenario 2 U-Boat kills. Of the 20 sign test trials, the p-values ranged in value from 0.058 to 0.5. Under the rejection criteria, the null hypothesis was not rejected in any of the 20 trials.

Both sign test experiments indicate the simulation is representative of historical combat operations for Scenario 2; null hypothesis was not rejected in 20 trials for either MOE. Though the original validation test showed a statistical difference in the number of U-Boat sightings, the results of the sign test indicate the simulation was a better model than the original test indicated. The monthly mean test demonstrated 100% overlap between the historic and simulation confidence intervals. The conclusions drawn from the bootstrap/sign test methodology provide stronger indication than either of the previous tests for accepting the model as valid with respect to the MOEs.

Table 9: Bootstrap U-boat kills – scenario 1

Trial	10/42	11/42	12/42	1/43	2/43	3/43
1	0	0	1	0	1	1
2	1	1	1	1	0	1
3	1	0	0	1	0	0
4	1	0	1	1	0	1
5	0	1	1	1	0	0
6	0	0	1	1	0	1
7	0	1	1	1	0	1
8	0	1	1	0	1	1
9	1	1	1	1	1	1
10	0	1	0	1	0	0
11	0	0	1	1	1	1
12	1	0	1	1	1	1
13	0	0	0	1	1	1
14	0	1	0	1	1	1
15	1	0	1	1	0	0
16	0	0	1	0	0	1
17	1	1	0	1	1	1
18	0	1	1	1	0	0
19	1	0	1	0	0	1
20	0	0	1	1	0	1

Table 10: Bootstrap U-boat sightings – scenario 2

Trial	4/43	5/43	6/43	7/43	8/43	9/43
1	81	7	52	60	98	52
2	98	98	21	98	81	98
3	98	81	81	21	60	7
4	98	7	52	52	60	52
5	81	52	52	52	60	60
6	81	81	98	52	7	52
7	60	98	98	21	7	21
8	7	52	98	81	21	98
9	52	52	52	52	21	98
10	60	98	60	52	81	60
11	81	81	21	21	52	98
12	98	60	21	52	52	21
13	60	7	81	52	21	52
14	7	52	60	52	21	52
15	52	81	98	21	81	81
16	7	81	21	60	81	52
17	98	52	7	21	21	21
18	60	98	98	21	7	60
19	52	60	21	81	81	98
20	7	81	98	21	81	21

Table 11: Bootstrap U-boat kills – scenario 2

Trial	4/43	5/43	6/43	7/43	8/43	9/43
1	4	4	1	2	1	13
2	4	13	1	13	5	2
3	4	4	1	5	7	2
4	1	2	7	5	2	13
5	2	7	1	1	4	1
6	7	1	5	1	2	5
7	2	4	1	5	1	13
8	1	5	1	5	7	4
9	13	5	5	7	5	7
10	13	13	5	1	5	5
11	4	1	1	2	1	2
12	1	7	1	1	1	2
13	13	5	13	1	2	1
14	13	4	2	5	2	1
15	2	7	13	4	13	13
16	4	1	5	13	13	1
17	13	2	13	13	1	1
18	4	7	13	5	1	7
19	4	4	5	7	2	7
20	5	7	7	7	7	13

4 VALIDATION CONCLUSIONS

In the first validation analysis, the traditional t-test showed half of the six tests with statistical difference between the simulation and historic data, although the practical differences were essentially negligible. These tests assumed the historic outcome represented the mean of all such outcomes – a possibly risky assumption.

In the second validation analysis, the simulation appeared to perform exceedingly well against the real-world data. However, due to the joint confidence dilemma discussed previously, little insight could be made with practical statistical confidence.

The proposed bootstrap/sign test validation methodology provides more information either traditional method. The sortie hour tests produced null hypothesis rejection rate of 15% for Scenario 1 and 5% for Scenario 2. The remaining MOEs for Scenario 1 produced a null hypothesis rejection rate of 15% for U-Boat sightings and 25% for U-Boat kills. Scenario 2 produced a null hypothesis rejection rate of 0% for both MOEs.

Ultimately, the validation determination rests with the decision maker, who takes risk, practical differences, and other associated costs into account. Our experiences and test suggest the BoB model is sufficiently valid, and its success as an experimental platform has been demonstrated and well documented in Champagne and Hill (2003), Carl (2003), Champagne (2003b), Hill et al. (2004), and Hill et al. (2006).

ACKNOWLEDGMENTS

This research was funded by the Defense Modeling and Simulation Office (DMSO) and the Air Force Research Laboratory, Human Effectiveness Directorate (AFRL/HES). A special thanks goes to Col Eileen Bjorkman for her initial vision and support of this project.

DISTRIBUTION AND DISCLAIMER

DISTRIBUTION A. Approved for public release; distribution unlimited. The views expressed in this article are those of the authors and do not reflect the official policy of the United States Air Force, Department of Defense, or the US Government.

REFERENCES

Balci, O. and R. G. Sargent. 1984. Validation of simulation models via simultaneous confidence intervals. *American Journal of Mathematical and Management Sciences* 4: 375-406.

Balci, O. 1994. Validation, verification, and testing throughout the life cycle of a simulation study. *Annals of Operation Research* 54: 121-174.

Brown, L. 2000. Agent based simulation as an exploratory tool in the study of the human dimension of combat. Monterey, CA: Naval Postgraduate School.

Bullock, R. K. 2000. Hierarchical Interactive Theater Model (HITM). WPAFB, OH: Air Force Institute of Technology. AFIT/GOA/ENS/00M-05.

Carl, R. G. 2003. Search theory and U-boats in the Bay of Biscay. *OR/MS Tomorrow*. Spring 2003.

Champagne, L. 2003a. Development approaches coupled with verification and validation methodologies for agent-based mission-level analytical combat simulations. WPAFB, OH: Air Force Institute of Technology. AFIT/DS/ENS/03-02.

Champagne, L. 2003b. Bay of Biscay: Extensions into modern military issues. *Proceedings of the 2003 Winter Simulation Conference*, ed. S. Chick, P.J. Sanchez, D. Ferrin, and D. J. Morrice, 1004-1012. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, New Orleans, LA.

Champagne, L. and R. Hill. 2003. Multi-agent simulation analysis: Bay of Biscay case study. *Proceedings of SimTecT 2003*. Adelaide, Australia. May 26-29.

Cheng, R. 2001. Analysis of simulation experiments by bootstrap resampling. *Proceedings of the 2001 Winter Simulation Conference*. B. A. Peters, J. S. Smith, D. J. Medeiros, and M. W. Rohrer, 179-186. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.

Conover, W. J. 1999. *Practical nonparametric statistics, third edition*. John Wiley and Sons, Inc.: New York, New York.

Efron, B. 1979. Bootstrap methods: another look at the jacknife. *Annals of Statistics* 7: 1-26.

Efron, B. and R. Tibshirani. 1986. Bootstrap methods for standard errors, confidence intervals and other measures of statistical accuracy. *Statistical Science* 1: 54-77.

Erlenbruch, T. 2002. Agent-based simulation of German peacekeeping operations for units up to platoon level. Monterey, CA: Naval Postgraduate School.

Hill, R. R., G. A. McIntyre, T. R. Tighe, R. K. Bullock. 2003. Some Experiments with Agent-Based Combat Models. Military Operations Research, Vol. 8, No. 3, September 2003, pp. 17-28.

Hill, R. R., L. E. Champagne, and J. C. Price. 2004. Using Agent-Based Simulation and Game Theory to Examine the WWII Bay of Biscay U-Boat Campaign. Journal of Defense Modeling and Simulation, Vol. 1, No. 2, pp. 99-109.

Hill, R. R., R. G. Carl, L. E. Champagne. 2006. Using Agent Simulation Methods to Examine and Investigate Search Theory Against a Historical Case Study. Journal of Simulation, Vol. 1, No. 1, 29-38.

Ilachinski, A. 1998. Irreducible semi-autonomous adaptive combat (ISAAC). *Maneuver Warfare Science 1998*. ed. F.G. Hoffman and Gary Horne. United States Marine Corps.

Kleijnen, J. P. C. 1995. Statistical validation of simulation models. *European Journal of Operational Research* 87: 21-34.

Levy, S. 1992. *Artificial life: A report from the frontier where computers meet biology*. New York: Vintage Books, a division of Random House, Inc.

McCue, B. 1990. *U-boats in the Bay of Biscay: An essay in operations research*. Washington DC: National Defense University Press.

Morse, P. M. and G. E. Kimball. 1998. *Methods of Operations Research*. Alexandria, Virginia: Military Operation Research Society. Reprinted in its entirety from © 1951 first edition printed by MIT Press and John Wiley & Sons, Inc.

Tighe, T. 1999. Strategic effects of airpower and complex adaptive agents: an initial investigation. WPAFB, OH: Air Force Institute of Technology. AFIT/GOA/ENS/99M-09.

Waddington, C. H. 1973. *O.R. in World War 2: Operational research against the U-boat*. London, England: Paul Elek (Scientific Books) Ltd.

Woodaman, R. 2000. Agent-based simulation of military operations other than war small unit. Monterey, CA: Naval Postgraduate School.

AUTHOR BIOGRAPHIES

LANCE E. CHAMPAGNE is a Lt Col in the United States Air Force and Director of Operational Analysis in the Joint Data Link Information Combat Execution (JDICE) program. He has a Ph.D. in Operations Research from the Air Force Institute of Technology. His research interests include agent-based modeling and verification and validation methodology. His email address is <lance.champagne@nellis.af.mil>.

RAYMOND R. HILL is a Professor of Industrial and Human Factors Engineering with Wright State University. He has a Ph.D. from the Ohio State University and has research interests in heuristic analysis, applied optimization and simulation modeling. His email address is <ray.hill@wright.edu>.