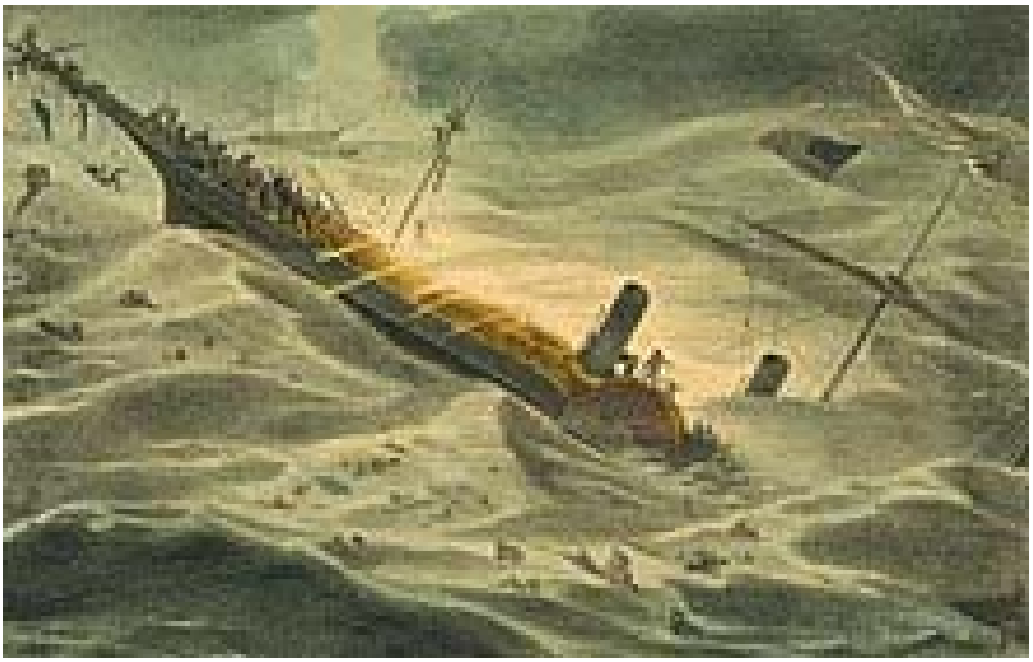
central moment

Case study on recovering sunken treasure: data and analysis used for locating the sunken SS Central America

Figure 1: An artist's depiction of the sinking of the *SS Central America* in 1857.



In 1857 a hurricane caught the steamship *SS Central America* on its way from Panama to New York and caused her to sink to the bottom of the Atlantic ocean off the southeastern coast of the United States. Along with the ship, the storm took the lives of 457 passengers and crew—making it one of the largest maritime disasters in the 19th century. Most relevant to the present story, though, is that the ship was in possession of three tons of gold freshly mined from the California gold fields. But the gold didn't stay lost at the bottom of the Atlantic ocean forever. In 1986, a recovery team led by Thomas G. Thompson of the Columbus Discovery Group of America began a search for the wreckage. And over the next couple of years it salvaged the gold. The story of the sunken wreck's discovery mixes the excitement of hidden treasure, personal drama (Tom Thompson is currently serving time in prison over the gold's recovery), and of course, data integration using Bayesian statistics. It's a perfect combination! Readers looking for a short, easy write-up of the sinking can find one here or here. In addition, the story was expanded into a fascinating book. And finally, Lawrence Stone (1992, 2010)—a participant on the wreck's recovery team--provides an academic treatment including a description of the analytics that led discovery of the wreck's undersea location. I lean heavily on Stone's (1992, 2010) treatment for the present case study.

Nowhere, however, has anyone easily made available the actual data used to locate the wreckage. And no one has published statistical code to demonstrate the Bayesian statistics that make sense of those data and that estimates the wreckage location. In this case study, I focus on these two contributions. I share the data used by the original recovery team (and scraped from Stone (1992)). And I share the R script demonstrating the analytic solution for finding the wreck's location. The R script allows anyone to reproduce my analysis—or even improve upon it.

The background and final hours of the SS America

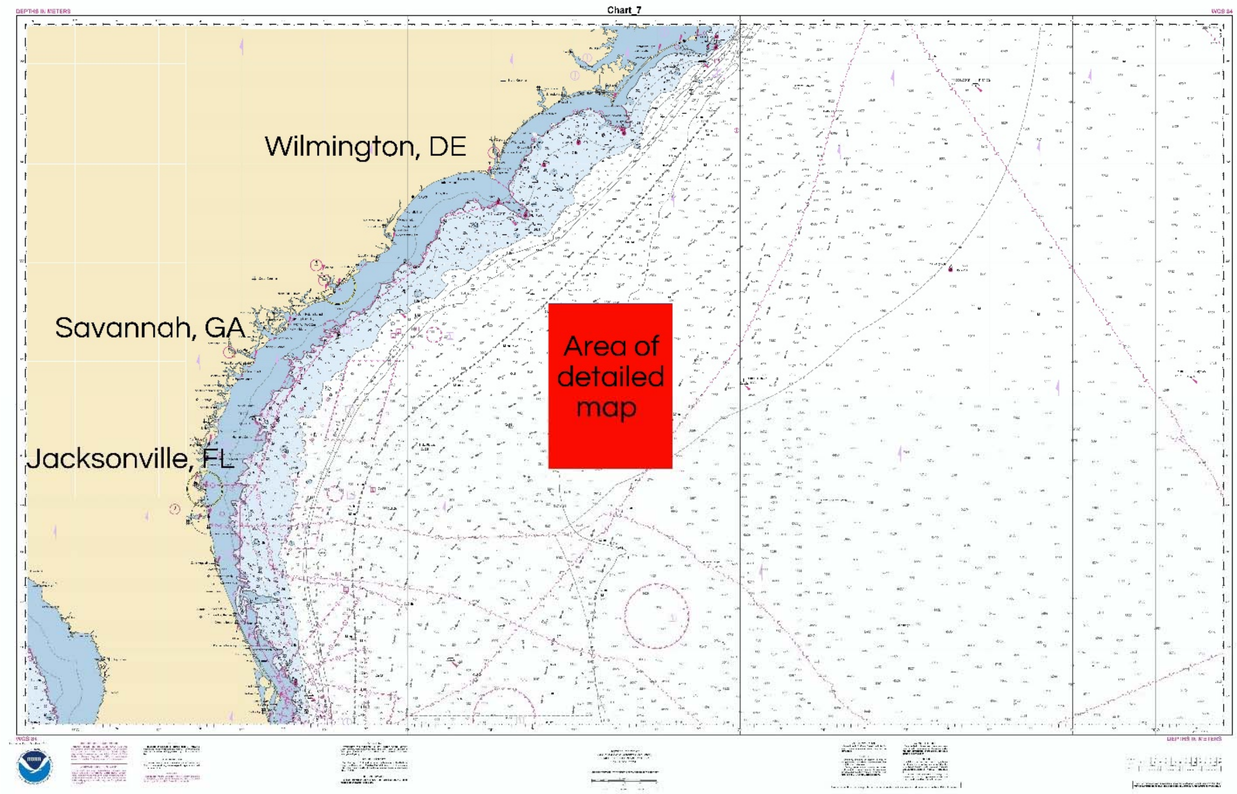
The *Central America*, a steamship captained by William Lewis Herndon, left Panama on September 3rd of 1857 and headed to New York with nearly 600 passengers and three tons of gold. Back then, travelers wanting to travel from the West to the East coast of the United States often found it more convenient to first travel south by sea to Panama, next travel overland to the eastern side of Panama, and then board another ship for the final leg of travel to the eastern United States—which is what most of the *Central America's* passengers and freight did. However, on that last leg of the trip, somewhere north of Cuba and around 150 miles East of Savanna Georgia, the *Central America* got caught in a Category 2 hurricane. Figure 2 shows a map of the general area.

Over the next couple of days, the crew and ship suffered greatly as the hurricane grew in intensity. First, the sail rigging was shredded by the heavy winds. (Although it was a steamship, the *Central America* also had three sailing masts). Next, the steamship's boiler shut down. Finally, a seal around the steamship's paddle wheels began to leak. Passengers and crew did their best to prevent the ship's ultimate sinking, but all efforts failed. The ship sank at around 8pm the night of Saturday September 12, 1857. One publication at the time described the dramatic event as follows,

> Suddenly the ship, as if in an agony of death itself, made a plunge, her interior gave one gigantic death-rattle, the next instant disappeared, leaving . . . human beings floating in darkness upon the fathomless ocean.

Other nearby ships saw the *Central America* in distress. One was able to draw close enough to exchange shouts. Another was even able to take-on some of the *Central America's* passengers. But no ship could provide complete help in the middle of a hurricane. The seas were too rough. But the nearby ships took note of the *Central America*'s location and two ships later retrieved some of the floating survivors. Out of the 578 passengers on board the *Central America*, only 121 survived. The other 457 who drowned defined the *Central America*'s sinking as one of the worst maritime disasters at the time. The story of the *Central America* dominated the nineteenth century imagination much as the *Titanic* did in the twentieth century.

Figure 2: The general location of the sinking of the *SS Central America*. The area in red matches the more detailed maps in Figures 3, 4, and 5.



The story of the ship's sinking—the loss of life and the three tons of lost gold—made news headlines at the time and has long intrigued investigators. It wasn't until the 1980's, however, that technology to reliably locate the wreck under more than a mile of ocean finally emerged. Tom Thompson, working with Columbus-America Discovery Group, led the recovery project. His team began their search by combing through historic accounts to trace the ship's location in the hours before its sinking. The team found the following three first-hand reports of the ship's last known location.

1. From Captain Herndon prior to sinking. On the Saturday evening before the *Central America's* sinking, another ship, the *SS El Dorado* approached the *Central America* upon seeing her in distress. At around 6pm the *El Dorado* even got within 50 feet of the *Central America* but the storm soon forced the two ships back apart. Before getting pulled away, however, Captain Herndon of the *Central America* verbally shouted the *Central America's* location (31°25'N, 77°10'W) to the *El Dorado*.

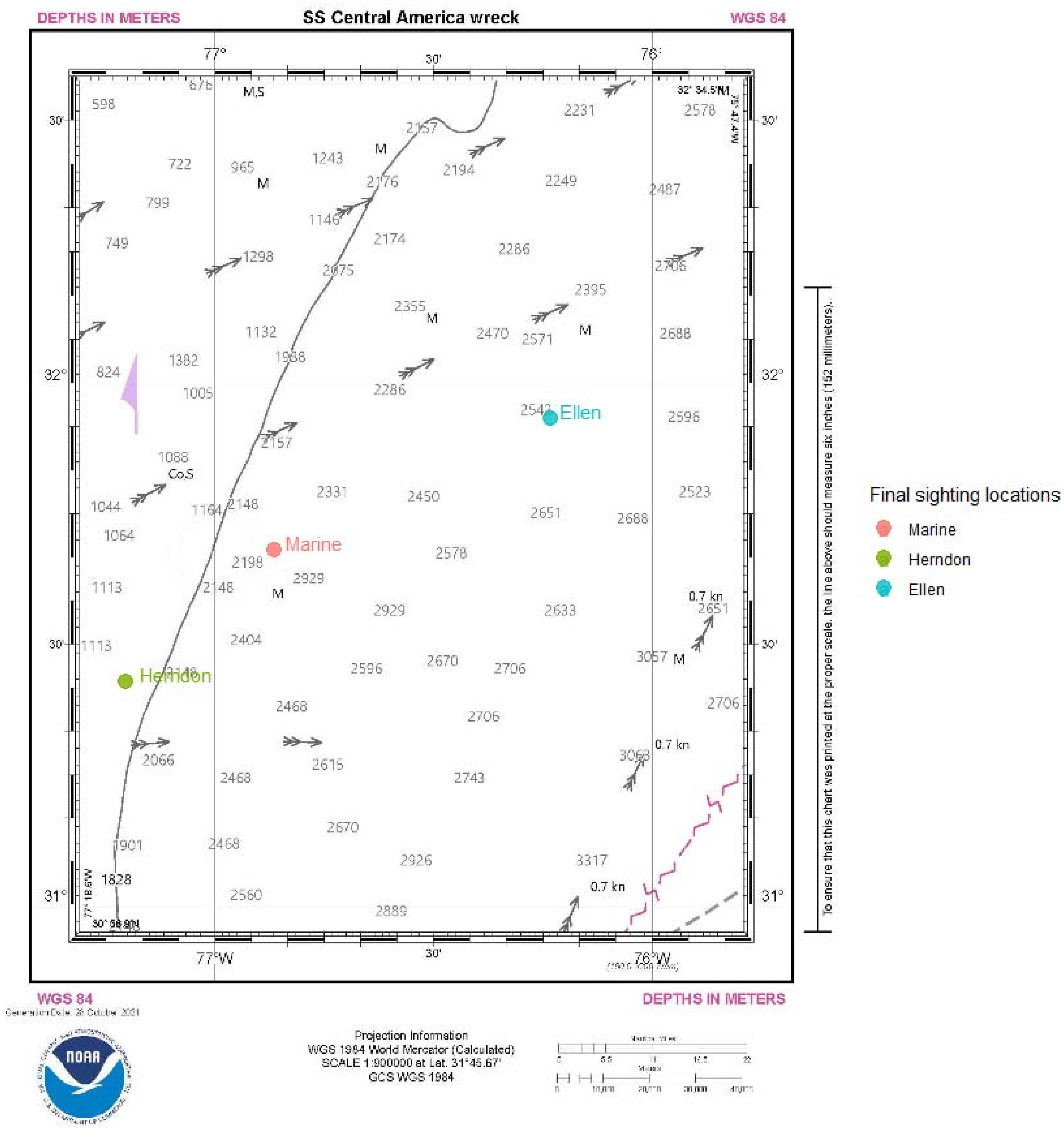
2. From the *Ellen*. The *SS Ellen* was the first ship to come upon any of the surviving passengers of the *Central America* at 8am on Sunday morning September 13 following the previous night's sinking. The *Ellen* took a fix of its location at the time and marked the spot as 31°55'N, 76°13'W in the ship's log.

3. From the *Marine* prior to sinking. The *SS Marine* noticed the struggling ship from a distance on Saturday and even took-on some of the ailing ship's passengers. Captain Burt of the *Marine*

reckoned the *Central America's* location to be 31° 40'N, 76° 50'W relative to his own known nearby position at 12:45pm.

To be complete, Tom Thompson's researchers noted a fourth estimate. Thomas Badger, another sea captain, was a surviving passenger from the *Central America* and was picked up among those rescued by the *Ellen* on Sunday morning. Captain Badger provided an estimate of the *Central America's* location when it sank. However, Badger's location estimate was so close to that of the *Ellen's* that Thompson's recovery team found it redundant. So, rather than include it as an independent data source, the recovery team simply used Badger's information to corroborate the *Ellen's*. Figure 3 shows a detailed nautical map and the three location sightings of the *Central America* and rough locations for the sunken ship's location. This map is a zoomed-in version of the area marked in red marked in Figure 2.

Figure 3: The three final sighting locations for the *SS Central America*. This map area is a more detailed view of the area in red in Figure 2.



Expanding the location sightings into sinking scenarios

The second step in Tom Thompson's team's effort to locate the *Central America's* final resting place involved yet more research. None of the three independent estimates of the *Central America's* sighting describe its actual sinking location. Two of the location estimates were made some hours prior to the ship's sinking. And one location estimate, that from the Ellen, was made the morning after the sinking. In the intervening time, other factors influenced the ship's ultimate location on the seafloor including…

- Ocean currents. Without sail and without a working boiler, the *Central America* must have been pushed by prevailing ocean currents. As per the small arrows shown inside the nautical map in Figure 3, the normal ocean current at the surface of the ocean would have moved the *Central America* in an East-by-northeast direction from the locations noted by Captain Herndon and the *Marine* prior to the ship's sinking. And after the ship's sinking, it's also likely those same currents pushed the surviving passengers East-by-northeast to where the Ellen picked them up the following morning.
- Wind. The direct effect of the wind blowing against the hull and other parts of a ship is called leeway. Stone (2010) tells us that Thompson's team believed the direction of leeway at the time of the *Central America's* sinking mirrored that of the current. Thompson's team also researched the profile of the *Central America* to estimate just how much of a role leeway likely played in moving the ship vis-à-vis the sighting locations. .

Starting with each final sighting location, Thompson's team expanded the three location estimates into three probabilistic ranges or sinking "scenarios" for the *Central America*. Indeed, all three of the latitude and longitude estimates from the final sighting locations were subject to error. There was also uncertainty in the speed and precise direction of the prevailing wind and current. And finally, there was some uncertainty in the time at which the final sighting locations were made (or when the *Central America* sank) and, therefore, how long the ship was subject to the prevailing currents and winds. To reflect all this uncertainty, Thompson's team projected the location of the sunken *Central America* over a wide range. The team divided a map like that in Figure 3 into 4-mile wide grids. Using simulation methods, they then assigned values to each square of the grid to represent the relative probability that each grid area contained the final wreckage of the *Central America*.

I refer the interested reader to Stone (1992) for figures showing the team's three maps, the grid system, and the points assigned to each grid location for each sinking scenario. For convenience, however, I've scraped the team's grid data into the following three files for easy download:

- Marine scenario data.txt
- Herndon scenario data.txt
- Ellen scenario data.txt

Each file above contains three columns—the latitude and longitude for the 4-mile wide grids and the probability values for each grid assigned by Thompson's team as described in Stone (1992). In Thompson's team's work, they scaled the grid values such that they summed to 1,000 for each sinking scenario. Due to rounding, however, the scraped points in the files above sum to a total slightly less than 1,000.

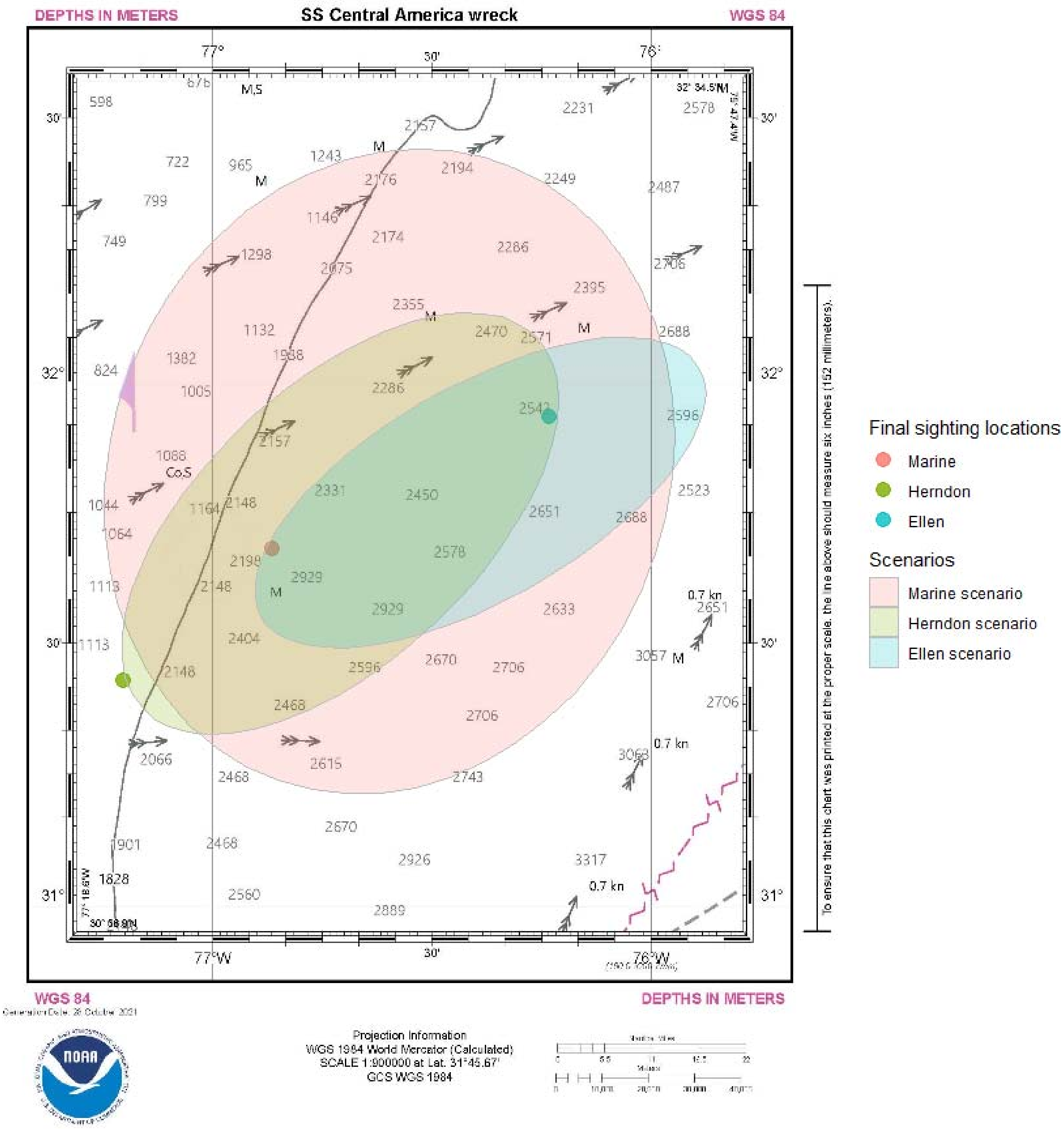
Figure 4 updates the detailed map with the three sinking scenarios assuming bivariate normal distributions fit to the three sinking scenarios. Indeed, while Thompson's team described the three sinking scenarios using a grid system, I fit a bivariate normal distribution to those data. Stone (2010), in fact, describes how the team used normal distributions for modeling the effects of the leeway, the current, and the various sources of error surrounding the final sighting locations. And the probabilities assigned to the map grids developed by Thompson's team strongly resemble normally distributed bivariate densities. Therefore, it's quite natural (and later sets us up for more generalizable Bayesian solution) if I model the three sinking scenarios using normal distributions and then use those distribution parameters rather than the discrete values from the map grids. Table 1 describes these parameters--the means, covariances, and also the correlations for all three normally distributed sinking scenarios.

Table 1: The means, variances, and correlations of the three sinking scenarios for the *SS Central America*. Latitude and Longitude figures are in decimal notation.

	Means	Covariance matrices			Correlation matrices		
Marine scenario	Lat: 76.581		Long.	Lat.		Long.	Lat.
	Long: 31.813	Long.	0.029	-0.003	Long.	1.000	-0.094
		Lat.	-0.003	0.026	Lat.	-0.094	1.000
Herndon scenario	Lat: 76.692		Long.	Lat.		Long.	Lat.
	Long: 31.712	Long.	0.017	-0.008	Long.	1.000	-0.579
		Lat.	-0.008	0.011	Lat.	-0.579	1.000
Ellen scenario			Long.	Lat.		Long.	Lat.
	Long: 76.375	Long.	0.018	-0.006	Long.	1.000	-0.601
	Lat: 31.777	Lat.	-0.006	0.006	Lat.	-0.601	1.000

I note that the ranges for the three sinking scenarios in Figure 4 make sense relative to the final location sightings. For example, the center of mass for the Ellen scenario is to the West of where the Ellen picked-up the surviving passengers on Sunday morning. After all, the prevailing wind and current was moving in a West-to-East direction and would have most likely pushed the floating survivors East following their ship's sinking. And over against the Ellen scenario, the center of mass for the Herndon and *Marine* scenarios is to the East of their location sightings because those location sightings were made on Saturday afternoon before to the ship's sinking. After their sightings the crippled *Central America* was subject to that same West-to-East current. I observe from Figure 4 that the *Marine* scenario is the least certain and, therefore, covers the widest range since the *Marine's* estimated location of the *Central America* was made through reckoning, over some distance, and the estimate was made Saturday around noon and furthest away from the time of the *Central America's* sinking.

Figure 4: The three location sightings of the *Central America* along with the three sinking scenarios. This map area is a more detailed view of the area in red in Figure 2.


DEPTHS IN METERS
SS Central America wreck
WGS 84
77°
30'
76°
32°
31°
77°W
76°W
Final sighting locations
Marine
Herndon
Ellen
Scenarios
Marine scenario
Herndon scenario
Ellen scenario
0.7 kn
To ensure that this chart was printed at the proper scale, the line above should measure six inches (152 millimeters).
WGS 84
DEPTHS IN METERS
Projection Information
WGS 1984 World Mercator (Calculated)
SCALE 1:900000 at Lat. 31°45.67'
GCS WGS 1984

Deciding where to search for the wreckage

The third step in locating the *Central America's* wreck location involved distilling the information from all three sinking scenarios into a recommended search area. Obviously, embarking on a search for the remnants of a 150 year-old wreck, located only imprecisely, under over one mile of ocean water is an expensive proposition. The team naturally wanted to make their search as productive as possible. They wanted to prioritize searching those areas where the ship's location was most probable. But they also didn't want to come away empty handed and, so, planned on searching low probability but still plausible areas as well. However, determining the most efficient search area isn't directly obvious given three competing scenarios. There are a numerous somewhat simple strategies. For example, one such plausible strategy would be to perhaps search just within the area in which all three scenarios intersect. Perhaps the overlapping scenario area is the most likely to contain the wreck. Or at the other extreme, the team could have perhaps searched over the union of all three scenarios, i.e., the area covered by any one scenario. I note that this last idea, however, would involve searching over 3,000 square miles and might require multiple years to properly search.

Thompson's team found the solution by creating a composite probability map. Put simply, they created a fourth map built from a weighted average of the probabilities and grid locations drawn from the three independent sinking scenarios and maps. They used the following proportions or weights for combining the scenarios:

- Herndon – 23%
- Ellen – 72%
- Marine – 5%

According to Stone (2010), the above proportions came after internal discussion between Thompson, Stone, and another project director. This group was most confident in and put the highest weight on the Ellen scenario because the Ellen location sighting was made during a period of calm following the storm. The location sighting was also formally recorded in the ship's log. In addition, Captain Badger's location estimate corroborated the *Ellen's*. At the other extreme, the team gave the lowest weight to the *Marine* scenario because it's location sighting was taken earlier in the day on Saturday and only reckoned from a distance.

Thompson's resulting composite map now contained relative probabilities for each grid area that reflected all three scenarios. The team could properly prioritize its search.

In contrast to Thompson's team, I follow a more formal Bayesian approach for combining the three sinking scenarios. Bayesian statistics provides a mathematically optimal formula for combining different inputs when one can describe the inputs using formal probability distributions. And when all of the inputs are bivariate normally distributed, such as each of the three sinking scenarios, the formula is remarkably easy. The result is simply another bivariate normally distributed search area optimally derived from the ingredient scenarios' means and covariances. In brief, the mean of the optimal search range is a weighted mean computed from the three sinking scenarios, where the mixing weights are covariances from each scenario. Scenarios with a large covariance or a wide perimeter, such as the *Marine* scenario, automatically receive less weight. And scenarios with a low covariance or narrow perimeter, such as from the *Ellen*, receive more weight. The resulting final search range is called the "posterior estimate" and balances the information from the input probability distributions or, in this case, three sinking scenarios. [1]

Appendix 1 describes the Bayesian math for combining the three sinking scenarios. And the easily downloaded R script contains a function that implements this math and computes the posterior estimate or optimal search range. The script also reads the data from the three sinking scenarios and produces all of the maps used in this case study. Figure 5 now displays the final Bayesian search range—along with the actual location of the discovered wreck that Thompson's team later found. The final posterior estimate shown in Figure 5 doesn't precisely match the subjective composite probability map developed by Thompson's team, but it's very close. Thompson's team created their composite probability map (see Stone 1992, 2010) after summing the subjective probabilities for each grid across the three sinking scenarios. And their search area may not precisely result in a normal distribution. By contrast, the final search range in Figure 5 reflects the bivariate normally distributed inputs after the application of Bayes rule and is itself a bivariate normal distribution. The final search range here covers an area of roughly 1,400 square miles.

The final resting place for the *SS Central America*—and lessons learned

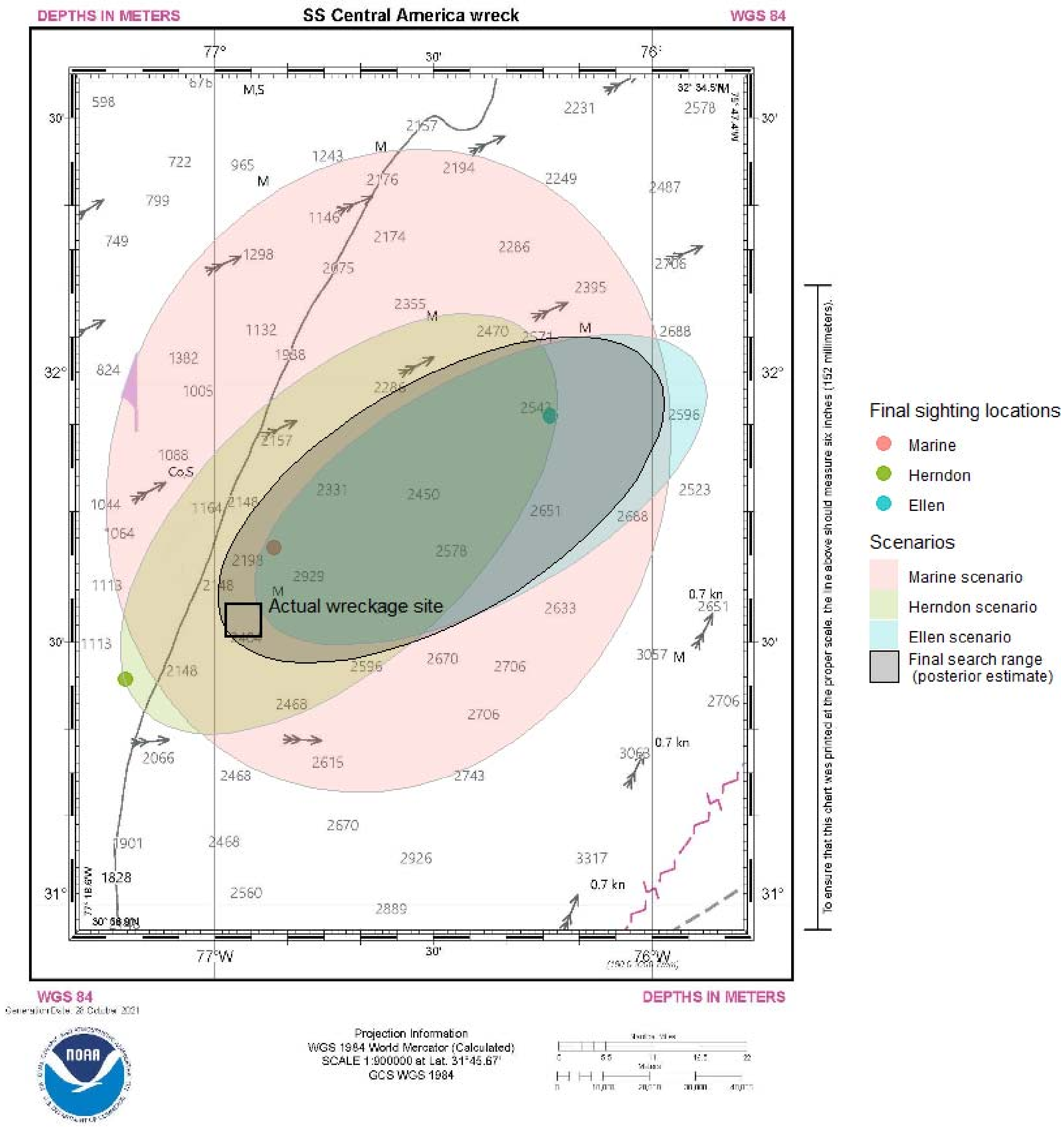
Following their construction of a search area, Thompson's team set out to look for the wreck. Starting in the summer of 1986, the team's recovery vessel traveled in a largely diagonal pattern across the darkly shaded final search area shown in Figure 5. The ship carried a side-scan sonar device that looked for anomalies on the sea floor that could indicate the presence of the sunken *Central America*. When the side-scanning sonar found a possible location, Thompson's team later returned with a remote operated vehicle (ROV) for a more detailed undersea visual search. After returning in the summer of 1988, the team finally located the wreck of the *Central America* as confirmed by recovering the ship's bell. They then recovered a gold bar and several gold pieces that reassured the team's investment group that this was indeed the right spot and that the stories of the *Central America's* valuable cargo were indeed true. As derived from the images found in Stone (1992), the wreck's location lies within a square marked by the following coordinates:

[1] Some sources use an entirely different interpretation of the phrase Bayesian search theory. In some applications, researchers use a Bayesian approach to update a lost object's probable location following the search of each successive grid location. I didn't follow this approach in my analysis here. And it appears that neither did Thompson's research team at Columbus America Discovery Group. But for completeness, I'll describe the general idea. It starts by recognizing that the finding of a negative result after searching a particular grid area usually doesn't provide absolute certainty that the search object isn't located within that grid area. After all, the search equipment may not be terribly precise. Or perhaps there's operator or analyst error in interpreting the search result. Bayesian object search provides a method for reflecting that uncertainty and setting rules to repeat the search in a particular grid area. For example, after unsuccessfully searching a particular grid with a very high initial probability of containing the searched object, it could perhaps be more efficient to repeat a search of that same high probability area before moving the search to a much lower probability grid area--especially if the search instrument could yield false negative result.

Longitude	Latitude
76.941	31.562
76.865	31.562
76.941	31.499
76.865	31.499

Thompson’s team successfully found the wreck by following a rigorous approach.

Figure 5: The Bayesian suggested search location (posterior) along with the actual wreckage site for the sunken *SS Central America's* final location. This map area is a more detailed view of the area in red in Figure 2.



There are a number of noteworthy analytic lessons from the search for the *Central America*. First, as noted by Stone (2010), it proves the value of a methodical search using appropriate statistical methods. Previous attempts by other salvage operators to locate the *Central America* did not proceed as methodically and failed to find the sunken *Central America*. Instead, Thompson's team built three independent, logical sinking scenarios using multiple historical reports. They then assigned probabilistic values and weights to each scenario. And Bayesian statistics is the ideal method for combining such probabilistic information.

But interestingly, as shown in Figure 5, the wreck's actual location wasn't found near the center of the Bayesian search range. The wreck wasn't inside the grid area with the highest posterior probability. Rather, the ship's final resting place was near the perimeter of the suggested search area. Indeed, the boundaries in Figure 5 for the Bayesian search range as well as for all three sinking scenarios represent the extreme percentile of their respective bivariate normal distributions. Had the team set a less conservative, i.e., lower percentile threshold for searching over the grid areas, the Bayesian posterior estimate would have failed to include the true wreckage location. And Thompson's team might have never found the sunken ship. All of the library research and expensive side-scanning of the ocean floor would have been in vain.

This result—that the location of the *Central America* was near the boundary of the optimal search range--provides a second valuable lesson. It reminds that ocean currents, winds, and celestial location sighting, can be more variable than one might expect. And developing probabilistic sinking scenarios may need to reflect greater uncertainty. Indeed, the *Central America* appears not to have drifted nearly as far before sinking as the Herndon and *Marine* scenarios suggested. That is, the wreck's location is very much on the West end of the *Marine* and Herndon scenarios and relatively close to their final sighting locations. This suggests that either there was less current and leeway than assumed, the *Central America* sank earlier in the evening than believed, or perhaps the Herndon and *Marine* location sightings were simply not as accurate as assumed. And then similarly, the ship's floating survivors appear to have drifted much further than assumed in the *Ellen* scenario suggesting that either the current and wind were stronger than expected, that the *Central America* sank earlier than expected, or that there was greater error in the *Ellen's* location sighting as well. Indeed, the actual wreckage site is actually outside the western boundary of the *Ellen* scenario per Figure 5. Yet the *Ellen* scenario was the one given the greatest subjective weight by Thompson. In sum, the second lesson is that all three sinking scenarios could have used an assumption of greater uncertainty or greater variance in their normally distributed bivariate densities.

A third valuable lesson from the search comes from the great help provided by the three importance weights that Thompson's team gave the three sinking scenarios. As described in the Appendix, the conventional Bayesian approach actually doesn't require explicit additional weights for the input scenarios (beyond the covariances of their ingredient probability distributions) before one combines them. Rather, according to Bayes' rule, the weights for combining multiple bivariate normal distributions are simply implied by their respective covariances. Inputs or scenarios with large variances and covariances are automatically given less weight per Bayes' rule. And inputs or scenarios with small variances and covariances automatically receive more weight. All uncertainty is assumed to be captured by their covariance matrices. In my present analysis, however, I expanded upon the traditional Bayesian formula. As described in the Appendix, I inserted the three weights from Thompson's team into the standard Bayesian analysis. I did this by including the additional weights as g-priors (sometimes called Zellner's priors after Zellner (1997)) which are multiplied with their respective scenario covariance matrices. Had I not used the weights from Thompson's team and had I just computed the standard

Bayesian posterior estimate, the final search range would've narrowly missed the wreck's actual location. Following this simpler approach, the recovery team would've failed to discover the sunken *Central America*. So, the final lesson from the story is the usefulness of additional scenario weights. Including such weights effectively inflated the covariances of the three sinking scenarios and enlarged the resulting posterior estimate and final search area. In effect, the additional weights provided the three scenarios with the added level of needed uncertainty. Their inclusion resulted in a slightly larger posterior estimate or final search range.

Locating the sunken *Central America* is an ideal application of Bayesian statistics and the integration of multiple data sources using additional importance weights. Less rigorous approaches to conducting such a search may not have been as successful. For example, if Thompson's team had searched within, say, just the areas of overlap among the three scenarios, the search would've completed much sooner but without success. Similarly, if Thompson's team had decided to merely search only within the most probably scenario—that built from the *Ellen* sighting—the team would've also missed discovering the wreck. Or at the other extreme, had the team searched the region covered by the union of all three scenarios, i.e., all the colored areas in Figure 5, the search would've taken nearly three times as long as such a search would have traversed extremely low probability areas. In sum, there's great power in combining all available information when making important decisions and combining that information in mathematically optimal ways.

Epilogue

This case study focuses on the Bayesian statistics behind the search and discovery of the sunken *Central America* and doesn't detail what followed afterwards. In brief, Thompson's team returned to the wreck location of the *Central America* in 1989 and began recovering its sunken treasure in earnest, retrieving one ton of gold bars and gold coins. But the story didn't end happily for Thompson or his investors. After retrieving that first batch of gold, numerous insurance companies and another salvage operation stepped forward and laid claim to the recovered treasure. And the investors who bankrolled the exploration have yet to be fully compensated. Tom Thompson and the Columbus America Discovery Group have become embroiled in legal proceedings that are still ongoing. It seems that finding resolution in the courtroom has taken longer than did finding the sunken *Central America* at the bottom of the Atlantic.

Part of the courtroom conflict, though, appears due to Thompson himself. In fact, as of this writing, Tom Thompson is serving a prison term for contempt of court related to the ship's recovered treasure. In 2012 Thompson was supposed to appear in court and tell the judge what happened to 500 gold coins that were never sold-off and remain unaccounted for. Thompson, however, skipped his court appearance and went on the run. He was soon apprehended by US marshals and forced to appear back in court. He told the federal judge in the proceedings that he suffered memory loss and didn't exactly know the location of the 500 disputed gold coins. But Thompson also told the court that he believed the gold coins were his rightful compensation for the years he spent locating the wreck. The judge in the case refused to believe Thompson's claims of a poor memory and ruled, "As long as you are content to be a master of misdirection and deceit to the court, I am content to let you sit." And so, there Thompson continues to sit. He resides in a minimum security prison as of this writing.

References

Lawrence D. Stone, (1992) "Search for the SS Central America: Mathematical Treasure Hunting." Interfaces 22(1):32-54. http://dx.doi.org/10.1287/inte.22.1.32

Lawrence D. Stone, (1992) "Revisiting the SS Central America search." 13th Conference on Information Fusion (FUSION), August. https://ieeexplore.ieee.org/document/5712111

Zellner, Arnold (1997). Bayesian Analysis in Econometrics and Statistics: The Zellner View and Papers. Edward Elgar. ISBN 978-1-85898-220-5.