Chapter 4
Testing and Validating Against Historic Spills

Abstract To validate the predictive capability of ERA Acute, a study was carried out using data from two well-studied historic oil spills, the Exxon Valdez Oil Spill (EVOS) and the Deepwater Horizon Oil Spill (DHOS) incidents. Results from the case studies with ERA Acute were compared to the impact estimates and recovery observations that have been reported in the extensive research following the two incidents. Resource data relevant for each of the two incidents were reconstructed within the analysis area. Performance boundaries were set up for evaluating the ERA Acute results, based on the ranges of the impact and recovery estimates reported in the post-spill assessments. Validation of an oil spill ERA model against post-spill assessments of historic spills is a challenging exercise due to scientific limitations of both. ERA Acute performed satisfactorily compared to the performance boundaries and the study gave useful insight into the predictive capabilities of ERA Acute. The results from the study were used to evaluate between two different impact models and to increase the individual vulnerability of cetaceans.

Keywords Model validation · ERA Acute validation · ERA Acute case studies · Impact validation · Exxon Valdez oil spill · Deepwater Horizon oil spill

4.1 Method of Validation Against Historic Spills

An ERA Acute assessment has been performed for two historic oil spill incidents. The study was performed according to the standard procedure of a regular environmental risk analysis (ERA) for exploration wells on the NCS (OLF 2007) using the ERA Acute methodology (cf. Fig. 1.6). The aim of the assessment was to compare ERA Acute results with damage estimates from post spill assessments from historic oil spill incidents, where such estimates are derived from observed and reported impacts. Required input data to perform the validation study were: (1) analysis areas and grids, (2) damage assessment from field observations, (3) pre-defined performance boundaries (4) oil drift statistics from stochastic modelling and field observation, (5) VEC datasets.

C. Stephansen et al., *Assessing Environmental Risk of Oil Spills with ERA Acute*,
SpringerBriefs in Environmental Science,
https://doi.org/10.1007/978-3-030-70176-5_4

The Deepwater Horizon oil spill and Exxon Valdez oil spill were selected as case studies for comparing results from ERA Acute against historic spills.

- The Deepwater Horizon Oil Spill (DHOS) began on 20th April 2010 in the Gulf of Mexico on the BP operated Macondo Prospect. Following the explosion and sinking of the Deepwater Horizon oil rig, a seafloor oil gusher flowed for 87 days, until it was capped on 15th July 2010. The US Government estimated the total discharge to be approximately between 701,000 to 857,000 m^3 crude oil (US Coast Guard 2011).
- The Exxon Valdez Oil Spill (EVOS) occurred 24th March 1989, when the tanker Exxon Valdez ran aground on Bligh Reef in Prince William Sound, Alaska. The vessel was traveling outside normal shipping lanes to avoid ice. Within six hours of the grounding, the Exxon Valdez spilled approximately 40,000 m^3 Prudhoe Bay crude oil (Exxon Valdez Oil Spill Trustee Council, https://www.evostc.state.ak.us/index.cfm?FA=facts.details).

The comparison studies between the results of the ERA Acute analyses of the two cases and the post-spill estimations of damages were part of the process of validating ERA Acute as a method suitable for ERA purposes. Quantitative comparison studies against historical oil spills are not commonly performed for environmental risk assessment methods but have been performed for e.g. the biological effects model in SIMAP oil spill model (French-McCay, 2004; French and Rines, 1997). Following an evaluation of data availability and quality, the EVOS and DHOS cases, limited to surface and shoreline compartments, were chosen for comparison. ERA Acute impact calculations were compared to injury estimates from post spill assessments. For the sea surface compartment, both modelled and satellite oil drift data were used in the study.

4.1.1 Analysis Areas

The analysis area for the DHOS case was set to cover the US Economic Exclusion Zone (EEZ) of the Gulf of Mexico (Fig. 4.1). The area is represented by 10,792 surface grid cells of 10 × 10 km (cells containing water) covering approximately 1,014,789 km^2 sea surface area, divided into 188,989 km^2 coastal area (<40 km from the coastline) and 825,800 km^2 offshore area (>40 km from the coastline).

The analysis area for the EVOS case was divided into the two areas, one that covers the total impact area including Cook Inlet, Kenai Peninsula, Kodiak Island and Alaska Peninsula and one that covers the Prince William Sound (Fig. 4.2). The Prince William Sound is represented by 240 surface grid cells of 10 × 10 km overing approximately 14,592 km^2 sea area. The analysis area for seabirds and marine mammals was restricted to the Prince William Sound since the VEC dataset and injury assessment estimates were most reliable in this area.

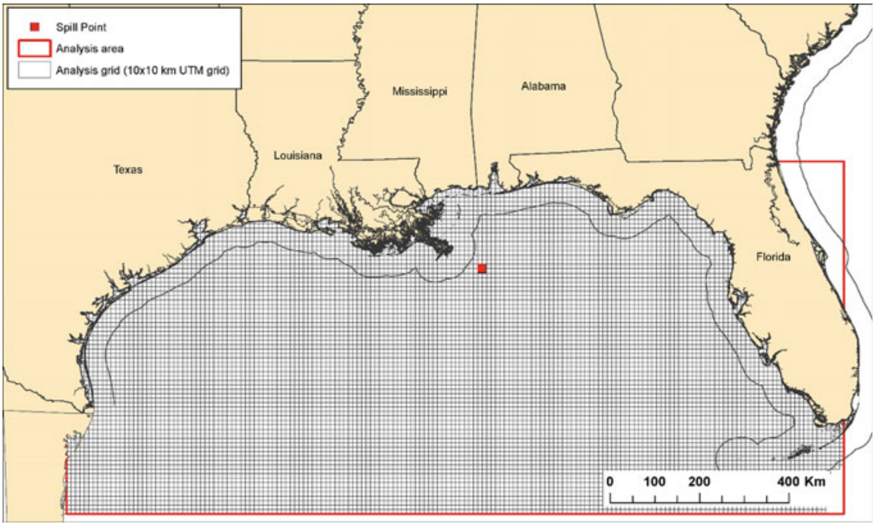


Fig. 4.1 Study area and 10 × 10 km UTM-grid used in the analysis for the Deepwater Horizon Oil Spill

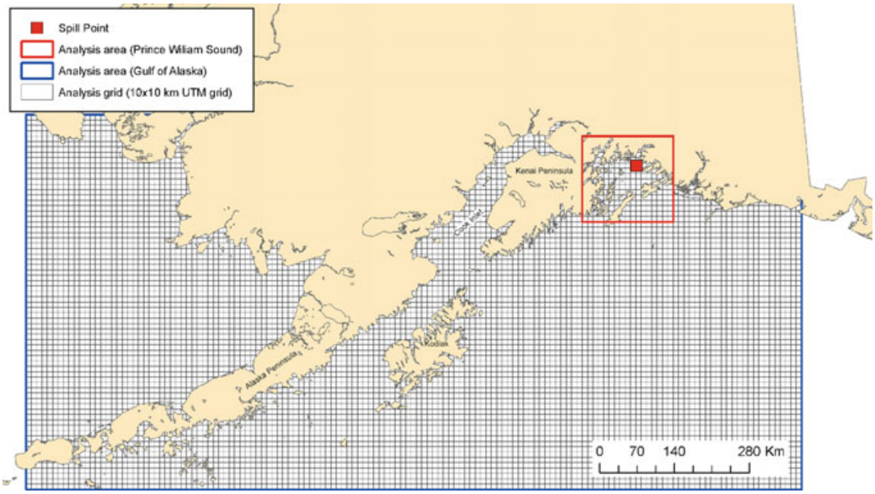


Fig. 4.2 Study areas and 10 × 10 km UTM-grid used in the analysis for the Exxon Valdez Oil Spill

4.1.2 Construction of Performance Boundaries

To evaluate the performance of ERA Acute, we defined *performance boundaries* based on field-based injury assessments and compared these with the impact and long-term damage calculated with ERA Acute. The main sources of information and

Table 4.1 The performance boundaries used to evaluate the performance of the biological impact models for seabirds, marine mammals, sea turtles and shoreline in ERA Acute for the Deepwater Horizon Oil Spill case

Valuable ecosystem component		Unit	Acute mortality and impact			
Group	Species		Threshold low	Limit low	Limit high	Threshold high
Seabirds	"All"	Individuals	8,500	56,141	900,000	1,000,000
Marine mammals	Bottlenose dolphin	Individuals	870	2,046	14,222	16,845
	Bryde's whale	Individuals	0.6	0.8	9.5	11.9
Sea turtles	Kemp's Ridley	Individuals	1,575	2,100	3,100	3,875
	Loggerhead	Individuals	1,650	2,200	3,600	4,500
Shoreline	Flora	Km	563	1,161	2,117	3,307
	Fauna	Km	704	1,451	2,646	4,134

Beyer et al. (2016); Deepwater Horizon Natural Resource Damage Assessment Trustees (2016); Haney et al. (2014a), (b), (2015), Lockyer and Morris (1990); Sackmann et al. (2015)

data were the injury assessments performed during the Natural Resource Damage Assessment s (NRDAs) process following the Deepwater Horizon and Exxon Valdez oil spills incidents, respectively and in the literature (cf. Table 4.1, Table 4.2 and Supplementary Information 1 for references).

The conceptual outline of the performance boundaries is illustrated in Table 4.3 and the values used in this study for the DHOS and EVOS cases are presented in Tables 4.1 and 4.2. The green circle is the mean impact estimated by ERA Acute from a single oil drift simulation and 500 Monte Carlo simulations. The Monte Carlo simulations are performed in three steps (cf. Fig. 5.1):

(1) assigning a probability distribution to the model parameters,
(2) drawing random values from the distribution and
(3) calculating the impact.

This is repeated 500 times per VEC dataset, resulting in either 500, 1500 or 2000 estimates of impact per VEC (cf. Sect. 4.1.4). The error bars are the 95% "credible interval" and represent the uncertainty in model parameters and natural variation in density and/or distribution of the VECs (cf. Sect. 4.1.4). The credible interval is analogous to confidence intervals and is used here to emphasize that the intervals are calculated on simulated and not measured data.

The estimates falling within the different boundaries are counted and summed up to give the percentage performance for one oil drift simulation. An example of this is illustrated for oil drift simulation No. 16 in Fig. 4.3.

Table 4.2 The performance boundaries used to evaluate the performance of the biological impact models for seabirds, marine mammals and shoreline in ERA Acute for the Exxon Valdez Oil Spill case

Valuable ecosystem component		Unit	Acute mortality and impact				Recovery	
Group	Species	Individuals	Threshold low	Limit low	Limit high	Threshold high	Low	High
Seabirds	Common murre	Individuals	1,176	3,075	15,918	23,877	10	13
	Pigeon guillemot	Individuals	135	500	1 500	2,250	Not recovering due to extrinsic factors	
Marine mammals	Harbor seal	Individuals	152	227	377	452	1	17
	Killer whale	Individuals	2	14	25	29	25	Not recovering
	Sea otter	Individuals	493	500	5,000	7,500	21	25
Shoreline	Flora	Km	20	39	185	340	3	25
	Fauna	Km	86	165	788	1,446		

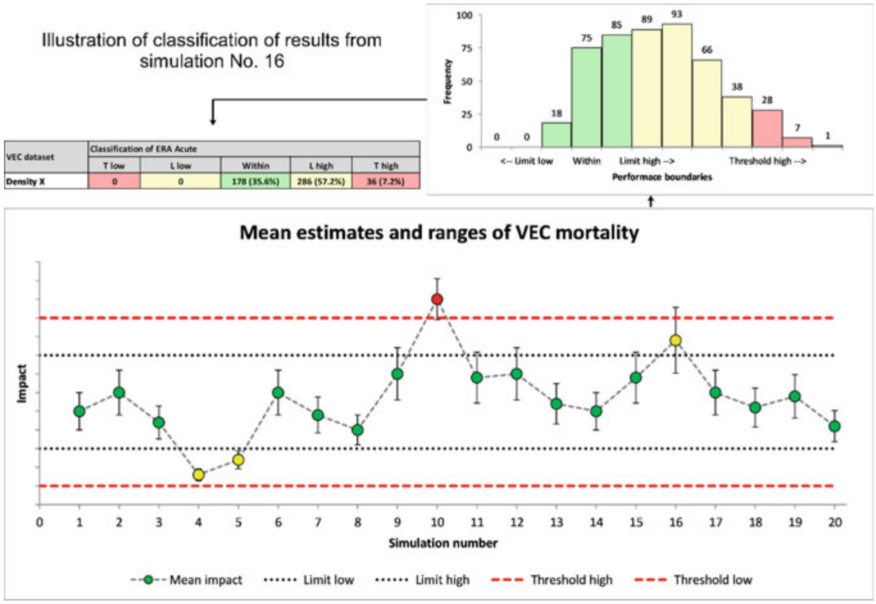
Exxon Valdez Oil Spill Trustee Council (EVOSTC) EVOSTC (2010), EVOSTC (2013), EVOSTC (2014), Piatt et al. (1990), Piatt and Ford (1996), ECI (1991), Piatt and Anderson (1996), Sanger and Cody (1994), Frost and Lowry (1994), Hoover-Miller et al. (2001), Ballachey et al. (1994), Garrott et al. (1993), Garshelis (1997), Udevitz et al. (1996), Gundlach et al. (1991)

Table 4.3 Densities used to derive resource datasets for seabird in coastal areas (<40 km from land) and at open sea (>40 km offshore) in the DHOS analysis area

Density[a]	Coastal (ind./km^2)		Open sea (ind./km^2)		Distribution
	Mean	SD	Mean	SD	
Density 1	1.53	2.30	0.56	0.84	Log-normal
Density 2	3.60	5.40	1.6	2.40	
Density 3	6.60	9.90	1.6	2.40	
Density 4	9.40	14.10	1.6	2.40	

[a] Deepwater Horizon Natural Resource Damage Assessment Trustees (2016), Tasker et al. (1984), McFarlane and Lester (2005), Hess and Ribic (2000) cited in Haney et al. (2014b)

The red lines are referred to as "thresholds" and model results falling below or above these boundaries are lower or higher than the damage estimates from post spill assessments, typically by 25%. The black dotted lines are referred to as "limits". Model results falling within the low and high limits are regarded as valid while model results falling outside the limits are regarded as satisfactory but uncertain. The results from Simulation No. 16 would be characterized as somewhat conservative and possibly even too conservative. Since different data sources are used to derive the thresholds and limits, the limit range may vary considerably and this, together with the availability and quality of input data must be taken into consideration when interpreting the results.



VEC dataset	Classification of ERA Acute				
	T low	L low	Within	L high	T high
Density X	0	0	178 (35.6%)	286 (57.2%)	36 (7.2%)

Fig. 4.3 Illustration of impact on a surface resource estimated from 20 oil drift simulations and how the estimated impacts place themselves according to boundaries derived from reported injury from the Natural Resource Damage assessments after the Deepwater Horizon, Exxon Valdez oil spill and in the literature

A stochastic approach was used to construct oil endpoint parameters to the ERA Acute models (see Sects. 1.5.1 and 4.1.3). We performed 20 oil drift simulations using OSCAR with different start dates within the seasonal time window of the two oil spills (April and May, 2001–2010 DHOS and March and April, 2006–2010 for EVOS). The differences between the estimated mean impacts (dots) in Fig. 4.3 are a result of different wind and current conditions resulting in variation in spreading, transport and weathering of the oil slicks. It is not the result of variations in the impact calculations in ERA Acute. For the DHOS case, we also tried to, by manually preparing oil spill input for ERA Acute, replicate the actual spreading and transport of the oil spill using information from field surveys and satellite data (cf. Sect. 4.1.3).

4.1.3 Reconstruction of the Oil Spills in the Analysis Areas

4.1.3.1 Oil Spill Modelling Approach

The oil spills were modelled with OSCAR (Oil Spill Contingency And Response) v.8.0 software (SINTEF 2016). OSCAR is a three-dimensional dynamic oil trajectory and chemical fates model that computes and records the distribution of oil on the sea surface, along the shorelines, in the water column and on the seafloor.

A total of 20 single simulations were performed with start dates within the seasonal time window of the two oil spills. A single simulation was performed for the DHOS case to obtain concentration of oil in the sediment. All oil drift simulations extended for 20 days after the release had been stopped.

4.1.3.2 Satellite Data Approach

The trajectory for the DHOS was reconstructed based on field surveys and satellite datasets. For the sea surface we used the dataset "*Predictive Model Cumulative Surface Oil Extent (PDARP)*" (NOAA 2017). The dataset included daily prediction of surface oil coverage from a period of 90 days with satellite observations between 23rd of April and 11th of August 2010. The data were mapped onto the UTM grid file for the DHOS and the time averaged coverage and exposure time for each 10 × 10 km grid cell in the analysis area was calculated.

The time averaged coverage for the whole period was calculated as:

$$Time\, averaged\, coverage = \frac{1}{n \geq 1} \times \sum_{i=1}^{n} \frac{\sum_{i=1}^{k} Coverage}{k} \tag{4.1}$$

where $n \geq 1$ is the number of 10 × 10 km grid cells with observed oil during the 90-day time period, k is the number of predictions of coverage from satellites within a 10 × 10 grid cell.

The exposure time is estimated as the number of days any given 10 × 10 grid cell was oiled during the 90 days of satellite observations. The maximum value is 66 days (five cells). The thickness of the oil slick is not known. In this study it is assumed that the thickness is above the threshold thickness for the VECs of interest (i.e. >2 and 10 μm).

4.1.4 Reconstruction of Resource Data in the Analysis Areas

A challenge in field validation studies is to reconstruct the pre-spill distribution and population size of the natural resource data in the study area. Three main techniques and data sources were used to construct dataset for surface VECs: (1) Monte Carlo Simulations, (2) extrapolation from field survey transects and (3) habitat density models. Monte Carlo Simulations were also used to represent uncertainty in the model parameters and to derive 95% credible intervals (cf. Sect. 4.1.4).

A brief description and examples of each type of dataset is given below.

Surface resource datasets

Three methods were used to estimate VEC densities and construct resource data sets used in the ERA Acute modelling. The two first methods were used for the DHOS incident and the third for the EVOS incident. Different methods are used based on different availability of suitable datasets.

The first method used estimates of VEC densities in the study area (Deepwater Horizon Natural Resource Damage Assessment Trustees 2016; Haney et al. 2014a, b) combined with probability distributions and Monte Carlo Simulations. Different densities and probability distributions were used for distinct habitat types within the study area. Each density and habitat were assigned a log-normal distribution with a standard deviation equal to 1.5 times the density (mean). The distribution of organisms in the environment is often log-normal and in most plant and animal communities, the abundance of species follows a (truncated) log-normal distribution (e.g. Limpert et al. 2001). A random number was drawn from the probability distribution, representing the abundance of the VEC in that cell. The same process was repeated until all grid cells in the study area were filled, and then repeated 500 times for each density (D). The densities used are given in Fig. 4.3 and an illustration of the distribution is given in Fig. 4.4.

The second method used datasets constructed by ecologists based on long term census (surveys) in the oil spill area (and season) and further processed using oceanographic and biological covariates to extrapolate abundance to areas not surveyed. These datasets are similar to the standardized VEC dataset used in ERAs on the Norwegian Continental Shelf today. The dataset for common bottlenose dolphin (*Tursiops truncatus*) is illustrated in Fig. 4.5. The dataset is derived by Marine Geospatial Ecology Laboratory/Duke University, based on habitat-based cetacean density models for the U.S. Atlantic and Gulf of Mexico (2015 Version). It is the first cetacean density map for these regions to be published in the peer-reviewed literature (Roberts et al. 2016). The abundance in each grid cell is given as the 5-percentile (P_5), mean (P_{50}) and the 95-percentile (P_{95}).

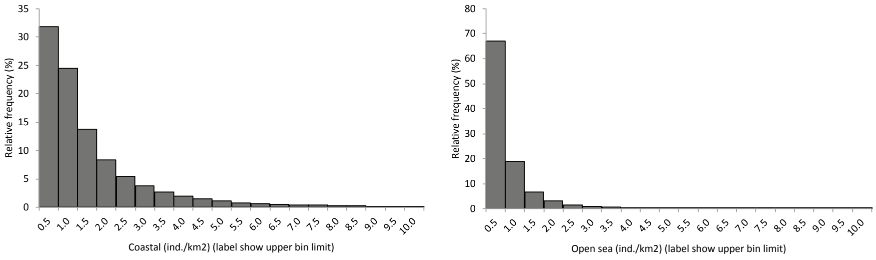


Fig. 4.4 Distribution of density in grid cell in the DHOS analysis area in coastal (left) and at open sea (right) using Density 1 in Table 4.3. The x-axis is cut-off at 10 individuals per km^2

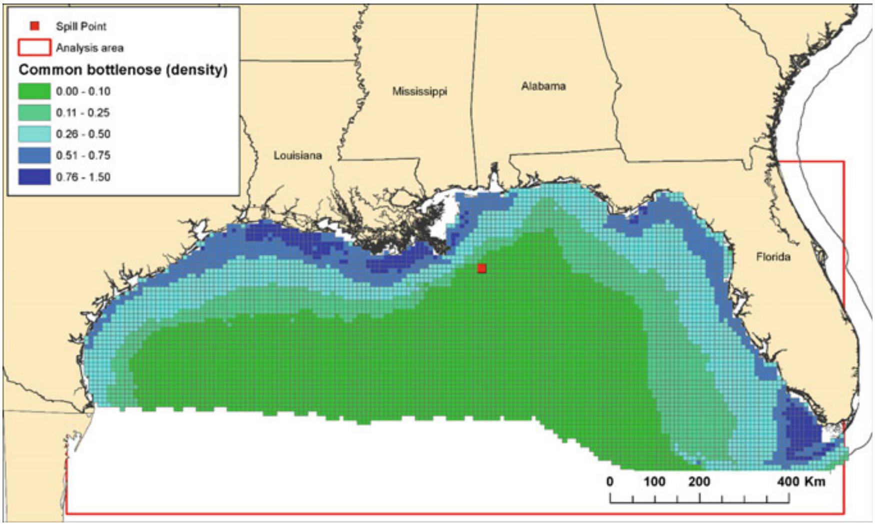


Fig. 4.5 Surface VEC datasets for the common bottlenose dolphin constructed from the habitat-based cetacean density models. *Source* Roberts et al. 2016

The third method uses the North Pacific Pelagic Seabird Database (NPPSD) (Drew et al. 2015) to construct resource dataset for surface VECs. The database includes more than 350,000 survey transects that were designed and conducted primarily to census seabirds but also includes data of several marine mammals. Transect areas and number of individuals during a transect were used to derive the distribution and density in each grid cell in the study area. The density was multiplied with the total area with suitable habitat in the grid cell (defined as cells containing seawater) and normalized against the estimated pre-spill population size of the VEC in the study area. The dataset for sea otter (*Enhydra lutris*) in Prince William Sound is illustrated in Fig. 4.6.

Vulnerability factors: A triangular probability distribution was selected to represent the uncertainty in the individual behavior factors, p_{beh} and physiological factors p_{phy} (Table 4.4). Seabirds in the Gulf of Mexico in May, June and August are dominated by surface feeding seabirds. The minimum, mode (the most likely value) and maximum values for birds in coastal habitats and open sea habitat were set equal to the estimates for coastal surface feeding seabirds (Wildlife Group 4) and pelagic surface foraging seabirds, respectively (Wildlife Group 2). For the other VECs, species or wildlife group specific values were used.

Shoreline resource datasets

ESI shoreline ranking data for the US coast of Gulf of Mexico and Alaska were downloaded from NOAA (https://response.restoration.noaa.gov/maps-and-spatial-data/download-esi-maps-and-gis-data.html). Post processing of these data included summary of shoreline length per ESI ranking in each 10 × 10 km UTM grid cell. For

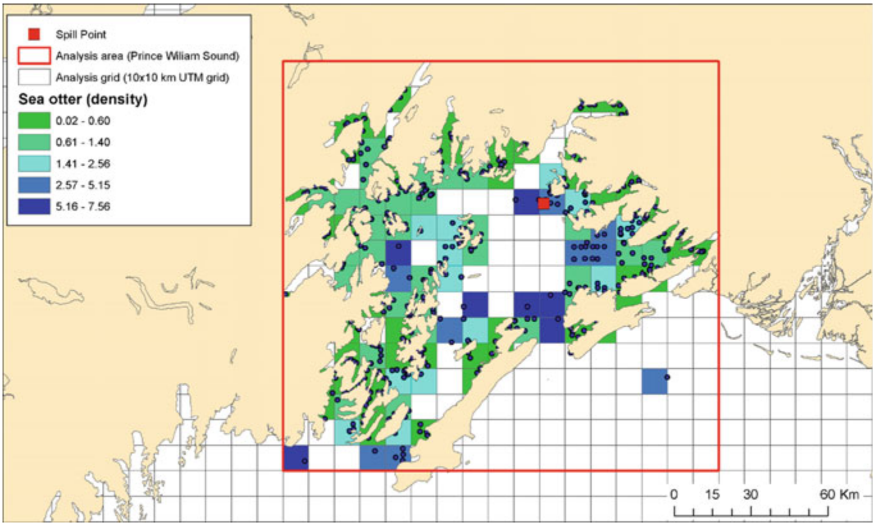


Fig. 4.6 Surface VEC datasets for sea otter constructed from the North Pacific Pelagic Seabird Database (NPPSD) in the EVOS analysis area. *Source* Drew et al. 2015

Table 4.4 Behavioral (p_{beh}) and physiological (p_{phy}) factors representing the likelihood of being oiled and the likelihood of lethal effect given exposure, and the parameters for the triangular distribution used to represent uncertainty in the Monte Carlo Simulations

VEC group	Case	VEC	Behavioral factor p_{beh}			Physiological factor p_{phy}		
			Min	Mode	Max	Min	Mode	max
Seabirds	DHOS	Seabirds—coastal	0.31	0.33	0.89	0.800	0.900	1.000
		Seabirds—open sea	0.31	0.45	0.89	0.800	0.900	1.000
	EVOS	Common murre	0.79	0.80	0.89	0.800	0.900	1.000
	EVOS	Pigeon guillemot	0.67	0.68	0.76	0.800	0.900	1.000
Marine mammals	EVOS	Harbor seal	0.83	0.90	0.96	0.004	0.028	0.058
	EVOS	Sea otter	0.79	0.88	0.97	0.500	0.720	0.930
	DHOS	Common bottlenose dolphin[a]	0.800	0.90	1.00	0.040	0.080	0.120
	DHOS	Bryde's whale[a]	0.70	0.79	0.88	0.025	0.050	0.075

[a] The values of the factors were calibrated during this study. See Sect. 4.2.1.2 for details

grid cells with extensive coverage of marsh/wetland (>100 km), a limit was set to 100 km per grid cell in order to capture the essential outer coastline reachable by oil in the oil spill model. The major shoreline habitat types in the Gulf of Mexico datasets are salt-, brackish- and freshwater marshes and swamps (ESI 10ABE) and the major shoreline types along the coast of Gulf of Alaska is gravel beaches, riprap (cobbles

Table 4.5 Overview of shoreline habitats used in this study. The shoreline habitat is more susceptible to damage by oiling with increasing ESI numbers

ESI	DHOS		EVOS		Description of ESI category
	km	%	km	%	
ESI1	133	0.8	1 888	11	Exposed rocky shores, exposed, solid man-made structures, exposed rocky cliffs with boulder talus base
ESI2	14	0.1	1 900	11	Exposed wave-cut platforms in bedrock, mud, or clay, exposed wave-cut platforms in bedrock, mud, or clay Exposed scarps and steep slopes in clay
ESI3	956	5.8	266	2	Fine to medium-grained sand beaches, scarps and steep slopes in sand, tundra cliffs
ESI4	12	0.1	137	1	Coarse-grained sand beaches
ESI5	161	1.0	3 110	18	Mixed sand and gravel beaches
ESI6	406	2.5	3 609	21	Gravel beaches, riprap (cobbles and boulders)
ESI7	555	3.4	259	2	Exposed tidal flats
ESI8	507	3.1	3 442	20	Sheltered scarps in bedrock, mud, or clay, sheltered riprap, sheltered rocky rubble shores, peat shorelines
ESI9	158	1.0	319	2	Sheltered tidal flats, vegetated low banks, hypersaline tidal flats
ESI10 ABE	11 851	72.5	1 985	12	Salt- and brackish-water marshes, freshwater marshes, swamps
ESI10 CD	1 599	9.8	0	0	Scrub-shrub wetlands, mangroves, Inundated low-lying tundra
Total	16 353	100	16 915	100	

and boulders) (ESI 6) (Table 4.5). Wetlands and marshes etc. are most widespread in Louisiana around the Mississippi River Delta while gravel beaches are scattered throughout the analysis area in PWS (Table 4.7).

4.2 Results of the Validation

4.2.1 Oil Drift

There were large differences in swept areas estimated from the oil drift model and from the oil drift constructed from satellite data in the Gulf of Mexico (Table 4.6). The modelled oil slicks ($n = 20$) on the surface cumulatively covered an area between 42,615 and 165,105 km^2. The mean area of oil thicker than 2 μm (oil slicks assumed to be harmful for seabirds) was 120,492 km^2 (range 56,944–165,105 km^2) and the mean area of oil thicker than 10 μm (oil slicks assumed to be harmful for marine mammals) was 90,302 km^2 (range 42,615–125,864km^2).

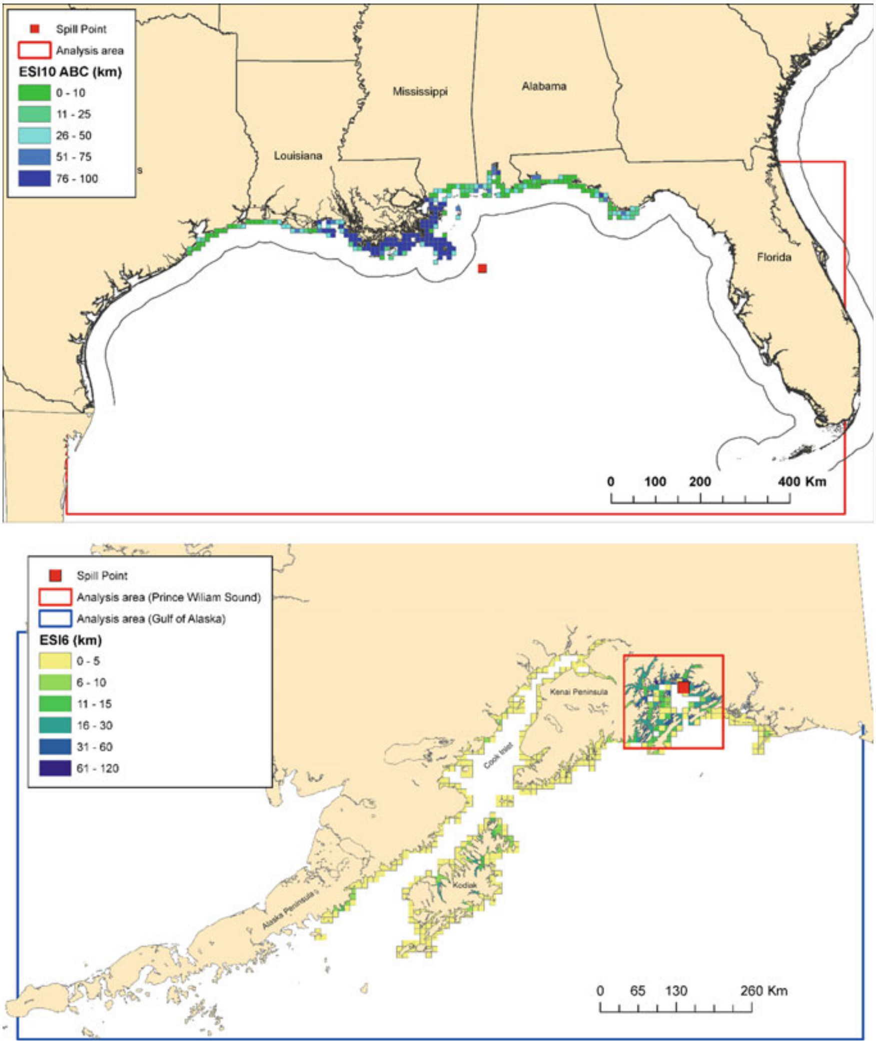


Fig. 4.7 Maps of shoreline habitat classified ESI 10ABE in the Gulf of Mexico VEC dataset (top) and ESI6 along the coast of the Gulf of Alaska (bottom)

This is within the range of the estimation of at least 112,115 km^2 by the Deepwater Horizon Natural Resource Damage Assessment Trustees (2016) but on the high side compared to the cumulatively swept area of oil slicks constructed by the satellite data. The largest differences between the modelled and observed data is the exposure time, with an overall mean of 1.2 and 1.0 days in the modeled data versus 11.8 days in the satellite data, respectively.

A comparison of the cumulative oil slick constructed from the satellite data with a deterministic simulation (Simulation No. 19 with start date 4th of March 2010) is

Table 4.6 Selected exposure statistics (mean, standard deviation) for the 20 probabilistic runs for the Deepwater Horizon Oil Spill and satellite data

Exposure statistics	Modelled data (n = 20)		Satellite data (n = 1)
	T = 2 μm	T = 10 μm	
Number of 10 × 10 km cells	5,309 (1,703)	5,155 (1,730)	2,013
Swept area (km^2)	120,492 (38,209)	90,302 (28,634)	58,578
Exposure time (days)	1.2 (0.2)	1.0 (0.2)	12.0

illustrated in Figs. 4.8 and 4.9. The modelled data (without use of oil spill response measurements) cover a larger area but with more variable coverage of oil in the cells and considerably shorter exposure time.

An illustration of accumulated oiling along the shoreline (from the same simulation) and data derived using Shoreline Cleanup Assessment Technique (SCAT) as part of the NRDA process (Nixon et al. 2016; NOAA 2017) is shown in Fig. 4.10. The modelled data show highest amount of beached oil in areas classified as "heavier oiling" (cf. Nixon et al. 2016) but also predict beaching in areas with no observed oil in the NRDA.

Differences between modelled and observed oil drift trajectory is expected due to uncertainty in the model and parameters, input data such as oil type and wind and current conditions. The extensive response to the oil spill is also likely to account for some of the differences in extent and area between the modelled and observed spreading of oil.

4.2.2 Acute Mortality in the Surface Compartment

A summary of estimated acute mortality for VECs in the surface compartment is presented in Table 4.7. The table lists the mean with two percentiles, the performance boundaries used to evaluate the results and the percentage of the simulations within each boundary. The performance varies between the different animal groups (seabirds and marine mammals) and between modelled (M) and field (F) oil drift data.

4.2.2.1 Seabirds

The estimated ERA Acute mortality with modelled oil drift data showed that on average 70% (range 61–100%) of the simulations resulted in mortality in the "within" category, 5% (range 0–16%) below the "limit low" and 24% (range 0–36%) above "limit high". No simulations yielded mortality below the "threshold low" and 1% above the "threshold high". The estimated ERA Acute mortality for seabirds with

Table 4.7 Estimated mortality with ERA Acute, the performance boundaries and classification of the estimated impacts for VECs in the surface compartments according to the performance boundaries. The number are given as percentage of the simulations falling within the different performance boundaries rounded to nearest whole number. M is modelled oil drift data and F is field (satellites) oil drift data. The number for cetaceans is based on the calibrated values for p_{beh} and p_{hy} (cf. Sect. 4.2.1.2)

Group	Species	Case	Oil drift data	Estimated mortality with ERA acute			Performance boundaries				Classification of ERA acute				
				Mean	$P_{2.5}$	$P_{97.5}$	T low	L low	Within	L high	T low (%)	L low (%)	Within (%)	L high (%)	T high (%)
Sea birds	Seabirds, four densities	DHOS	M	148,576	31,497	314,540	8,500	56,141	900,000	1,000,000	0	16	84	0	0
			F	273,753	82,279	484,625					0	0	100	0	0
	Common murre	EVOS	M	12,818	4,821	20,826	1,176	3,075	15,918	23,877	0	0	65	35	0
	Pigeon guillemot	EVOS	M	1,411	859	2,253	135	500	1,500	2,250	0	0	61	36	3
Marine mammals	Common bottlenose dolphin	DHOS	M	1,175	447	2,673	870	2,046	14,222	16,845	31	62	6	0	0
			F	9,796	6,256	14,092					0	0	98	2	0
	Bryde's whale	DHOS	M	0.3	0.0	1.1	0.6	0.8	9.5	11.9	80	16	3	0	0
			F	7.0	1.8	16.0					0	0	69	8	23
	Harbor seal	EVOS	M	195	28	685	152	227	377	452	35	37	17	6	5
	Sea otter	EVOS	M	2,374	1,205	3,997	493	500	5,000	7,500	0	0	100	0	0

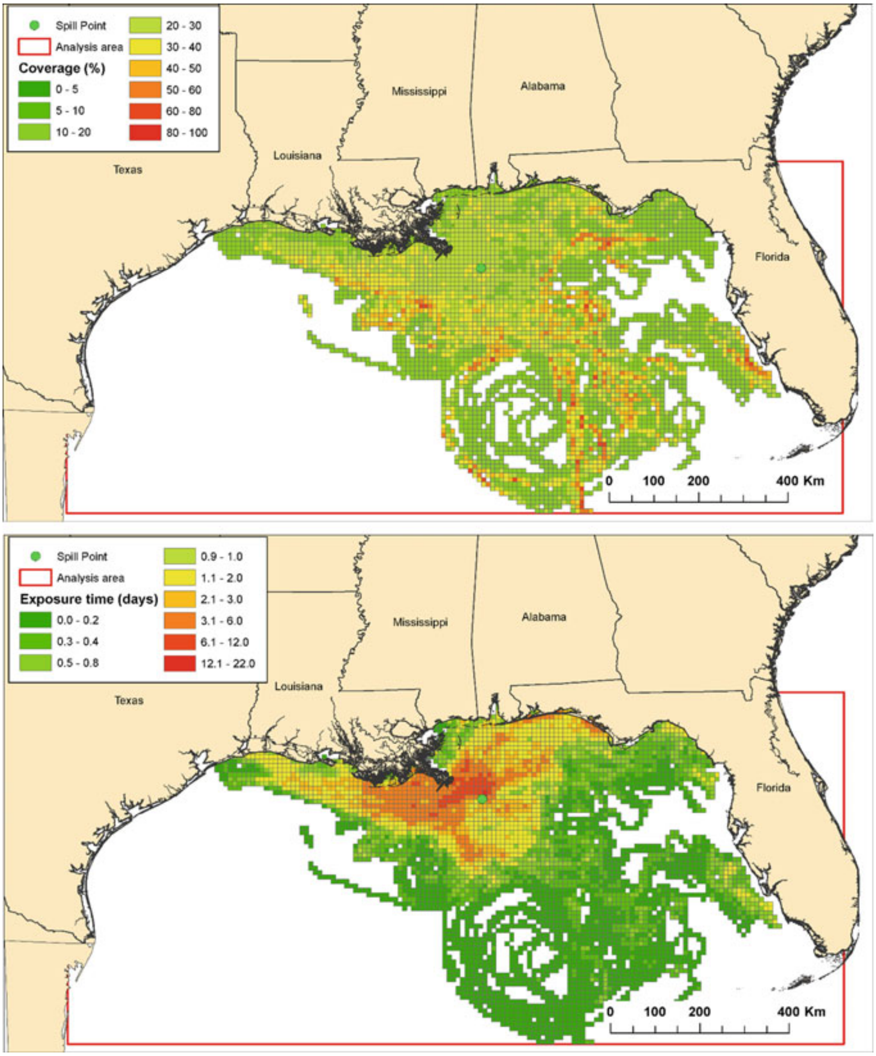


Fig. 4.8 The cumulative oil slick of the Deepwater Horizon Oil Spill derived from simulation No. 19. The color codes show the coverage of oil above thicker than 2 μm (top) and exposure time (bottom). Note that the classification of exposure time differs from the legend in Fig. 4.7

modelled oil drift data results in comparable values, and somewhat high estimates in comparison with the performance boundaries for the EVOS (Table 4.7 and Fig. 4.11).

The injury estimates for seabirds for DHOS vary considerably in the literature. The NRDA process estimated a mortality of 56,141–102,399 individuals (cf. Table 4.4.7–3 in Deepwater Horizon Natural Resource Damage Assessment Trustees 2016) while Haney et al. (2014a, b), using a carcass sampling model and an exposure probability model, estimated 600,000–800,000 individuals as a most likely value. The estimated

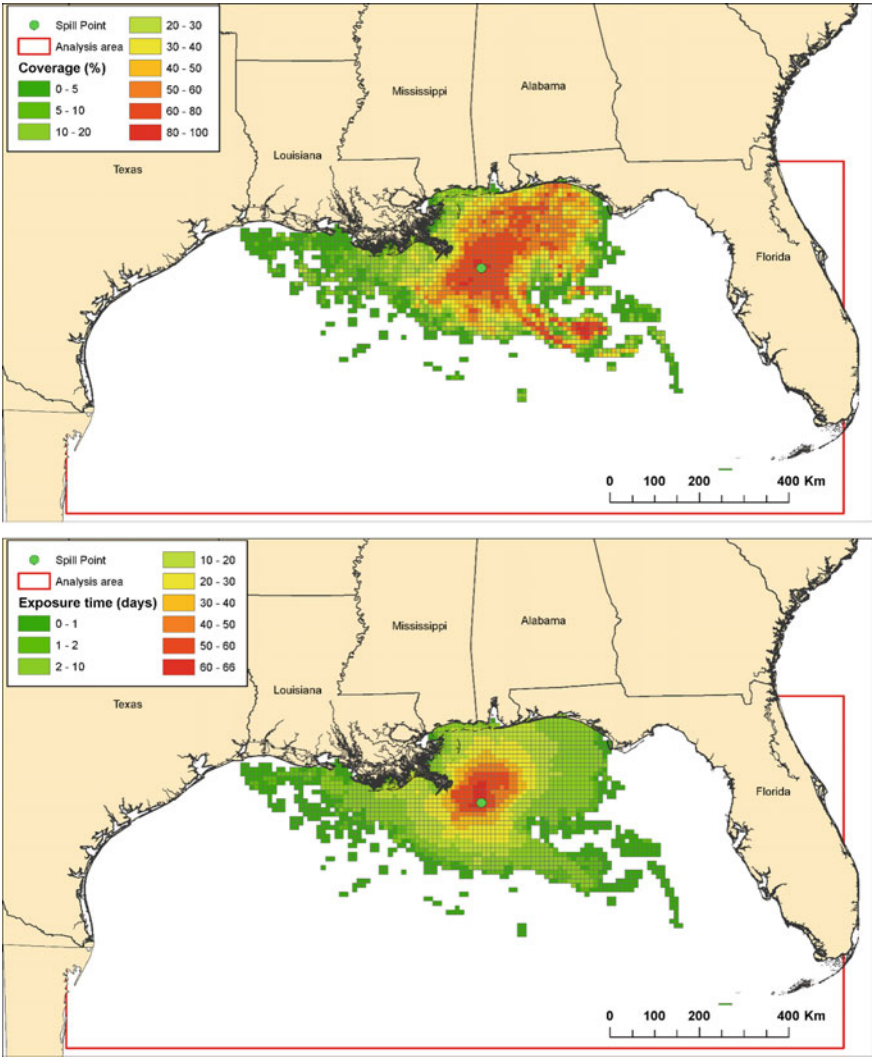


Fig. 4.9 The cumulative oil slick of the Deepwater Horizon Oil Spill derived from simulation field data (satellite). The color codes show the coverage of oil assumed to be thicker than 10 μm (top) and the exposure time (bottom). Note that the classification of exposure time differs from the legend in Fig. 4.8. *Source* Predictive Model Cumulative Surface Oil Extent (PDARP) (NOAA 2017)

ERA Acute mortality with modelled oil drift data in this study was 148,576 with a 95% credible interval (CI) of 31,497–314,540. This is a higher estimate range than the official NRDA estimates but lower than the mortality estimated by Haney et al. (2014a, b).

The estimated ERA Acute mortality with oil drift statistics derived from satellite data generated on average 1.8 times higher acute mortality than calculations based

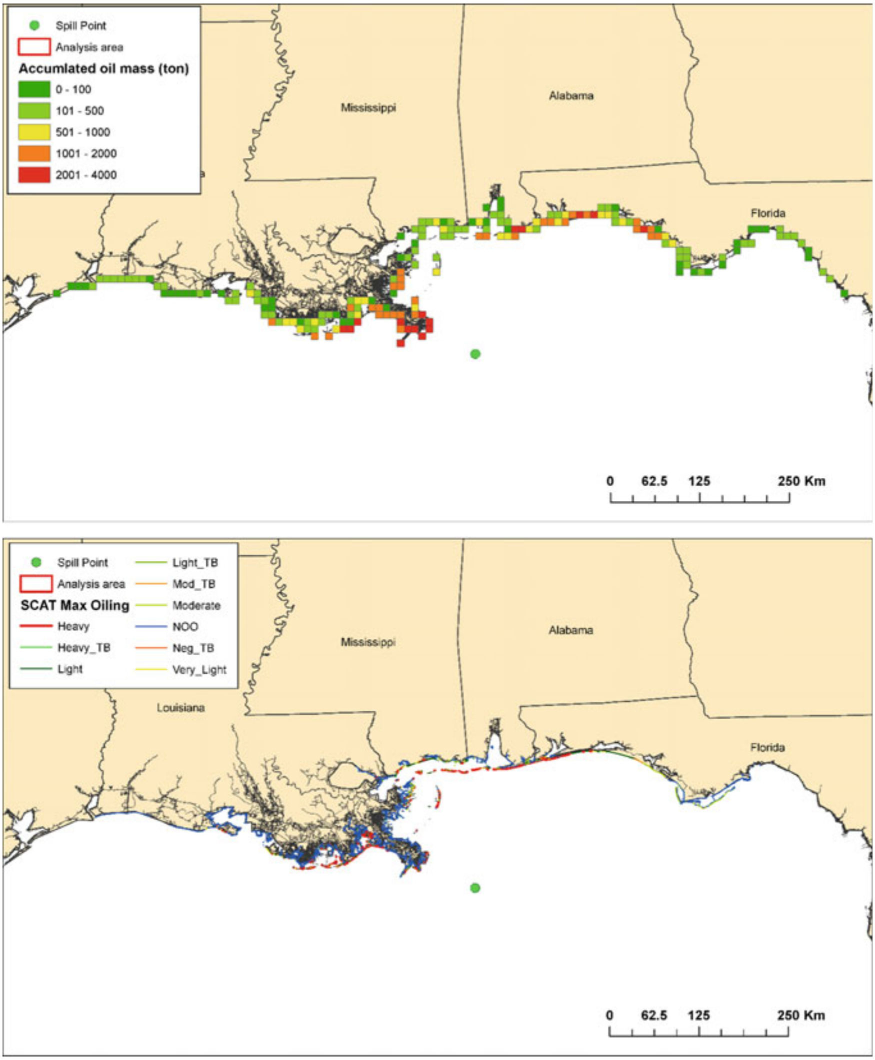


Fig. 4.10 Map showing accumulated oiling along the shoreline for the Deepwater Horizon Oil Spill derived from simulation No. 19 (top) and an illustration of shorelines classified by final oil exposure categories for beaches, coastal wetland and other shoreline habitats (bottom). See Nixon et al. (2016) and NOAA (2017) for original and detailed maps. NOO = No Observed Oil. Sources: ERMA Layer: 11-Nov-10 Mobile SCAT Maximum Oiling and ERMA Layer: 23-Jan-11 Houma SCAT Maximum Oiling

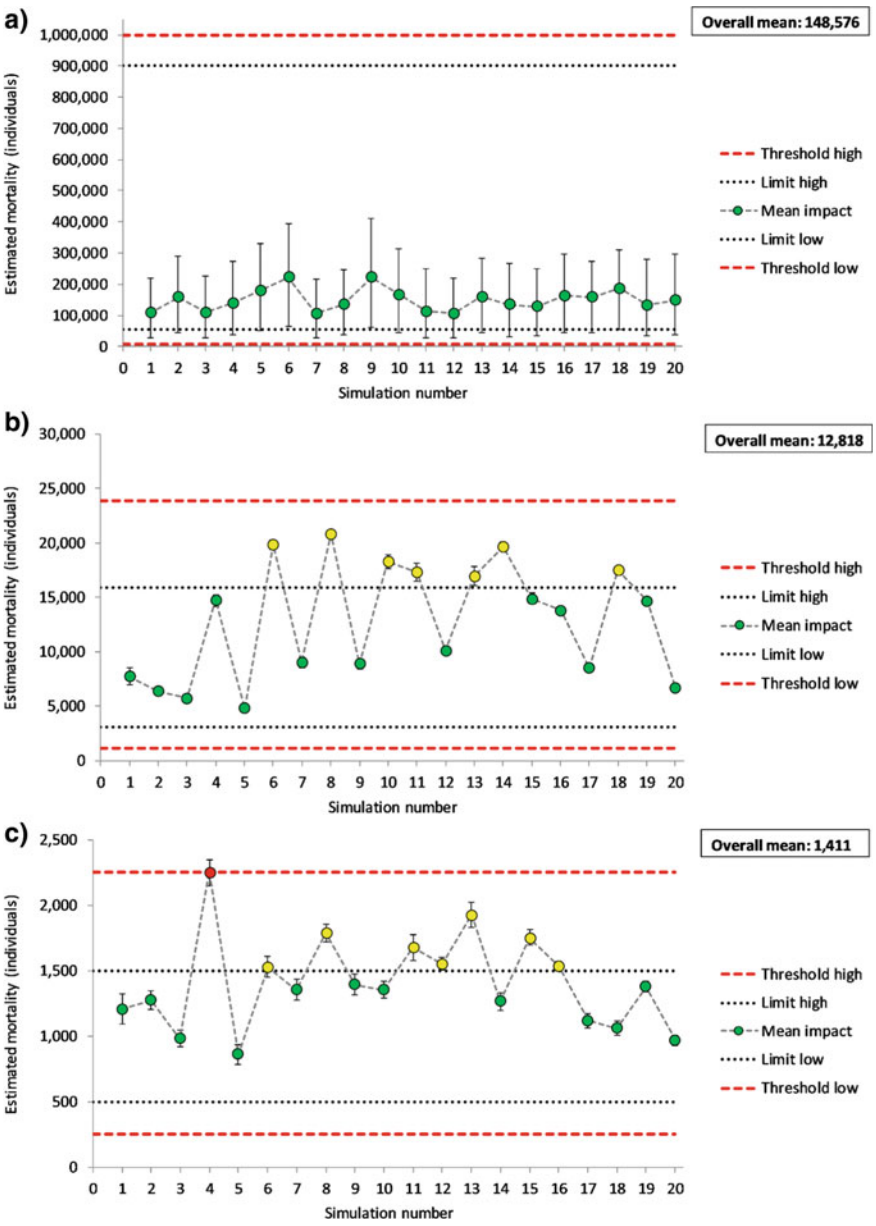


Fig. 4.11 Mortality for seabirds estimated from the 20 oil drift simulations performed in OSCAR for **a** Seabirds in the Gulf of Mexico, **b** Common murre in the Prince William Sound and **c** Pigeon guillemot in the PWS. Vertical bars show the 95% credible interval from the Monte Carlo Simulations. The larger error bars for seabirds in the Gulf of Mexico are primarily due to differences in density (D) used in the Monte Carlo Simulations (cf. Table 4.3)

on the modelled oil drift data (Table 4.7). This is mainly due to considerably longer exposure time of oil in the satellite data than in the modelled oil drift data. Compared to the performance boundaries, all simulations resulted in mortality in the "within" category.

4.2.2.2 Marine Mammals

The estimated mortality with ERA Acute for whales in the GOM was considerably underestimated compared to the field assessment and the performance boundaries. It is believed that this was primarily due to the physiological factor p_{phy} (probability of dying given contact with oil film on the sea surface thicker than 10 μm) firstly being set too low for toothed and baleen whales.

Cetaceans have in general been regarded as little vulnerable towards oil spills and the original p_{let} factor ($p_{phy} \times p_{beh}$) was 0.1%, based on early development work for the ERA Acute model and similar environmental risk analyses methods (e.g. French-McCay 2004, 2009; Østbye et al. 2003; Spikkerud et al. 2004; Spikkerud et al. 2010). Using the factors from previous work, the highest estimated ERA Acute mortality for bottlenose dolphins in the Gulf of Mexico using modelled data was 228 individuals and the highest estimate using field data was 1959 individuals, considerably underestimating the reported mortality.

During more recent development of ERA Acute, the factors were therefore re-evaluated for toothed and baleen whales (Stephansen et al. 2018) based on further scientific studies and preliminary reporting of high whale mortality from the Macondo accident (DHOS) (e.g. Deepwater Horizon Natural Resource Damage Assessment Trustees 2016). The increase in acute mortality estimated for the common bottlenose dolphin using the refined factors is illustrated in Fig. 4 12. The overall mean mortality increases from 1,128 (95% CI = 695–1,708) to 9,796 (95% CI = 6,256–14,092) individuals.

Currently recommended ERA Acute parameters are presented in Table 4.4 and are used in the results (Table 4.7), based on these calibrated vulnerability numbers for p_{phy} and $p_{beh.}$.

The mortality with the modelled oil drift data is considerably lower than the estimated mortality using the field data, with only 6 and 3% within the low and high-performance limits for the common bottlenose dolphin and the Bryde's whale, respectively (Table 4.7). The main reason for the large differences in the estimated ERA Acute mortality for satellite and model oil drift data is a considerably shorter exposure time of harmful oil in the grid cells in the modeled oil drift data (cf. Table 4.6).

For harbor seal in the PWS, the estimated mortality with ERA Acute was 195 individuals (95% CI = 28–685) resulting in an average mortality on the low side compared to the performance boundaries (Fig. 4.13a). The reasons for the large variation in the harbor seal results are partly due to relatively large uncertainty in the model parameters (cf. Table 4.4) and also possibly due to a scattered distribution in the resource dataset. The estimated mortality of or sea otters in PWS with ERA Acute

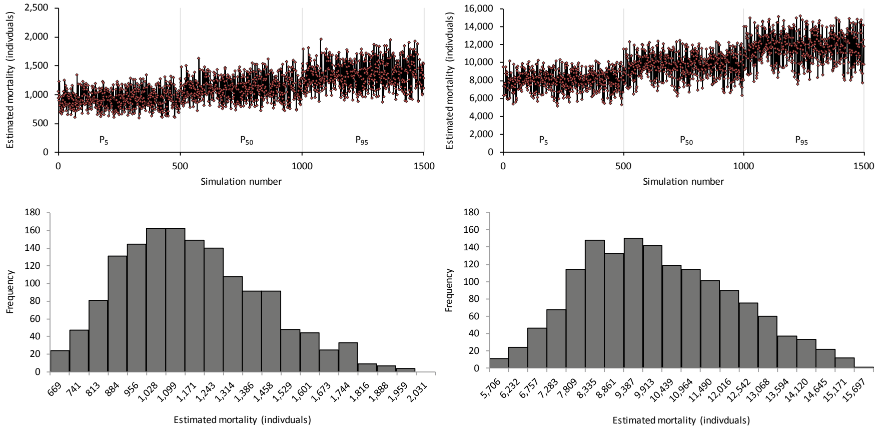


Fig. 4.12 Calibration of the individual vulnerability factors (p_{phy} and p_{beh}) for toothed and baleen whales. Impact is calculated based on oil drift statistics derived from satellite data. Result from 1500 simulations using original individual vulnerability factors (left panels) and corresponding results using the new individual vulnerability factors (right panels) for all three datasets (P_5, P_{50} and P_{95})

was 2,374 individuals (95% CI = 1,205–3,997), resulting in an average mortality within the performance boundaries (Fig. 4.13b).

4.2.3 Impact in the Shoreline Compartment

The estimated length of impacted coastline by ERA Acute using modelled oil drift data is longer for shoreline fauna than for flora. The result shows a satisfactory validation for both flora and fauna, between 55 and 85% of simulations are within performance boundaries for the two cases (DNV GL, Acona 2020). Table 4.8 summarize impact results for the EVOS and DHOS cases. The validation results for EVOS are shown in Fig. 4.14 and results for DHOS in Fig. 4.15.

The mean ERA Acute impact from all simulations is located within the performance boundaries for flora and fauna for both DHOS and EVOS. The average ERA Acute impact for shoreline fauna is similar as the reported cumulative oiling from observations: 2,225 ± 659 km SD versus 2,117 km for DHOS and 423 ± 203 km SD km versus 404 km for heavy to moderate oiling for EVOS (Gundlach et al., 1991, GEO 1994, Deepwater Horizon Natural Resource Damage Assessment Trustees, 2016).

The distribution of impact along the coastline and to a large degree, the most affected shorelines types in ERA Acute, corresponds to the estimates derived from the field data (DNV GL, Acona 2020). For EVOS, the calculated average shoreline fauna impact in ERA Acute is in line with Heavy + Moderate (HM) impact values for ESI 1 (Exposed Rocky Shores), ESI 2 (Exposed Wave-cut Platforms) and ESI 7 (Exposed

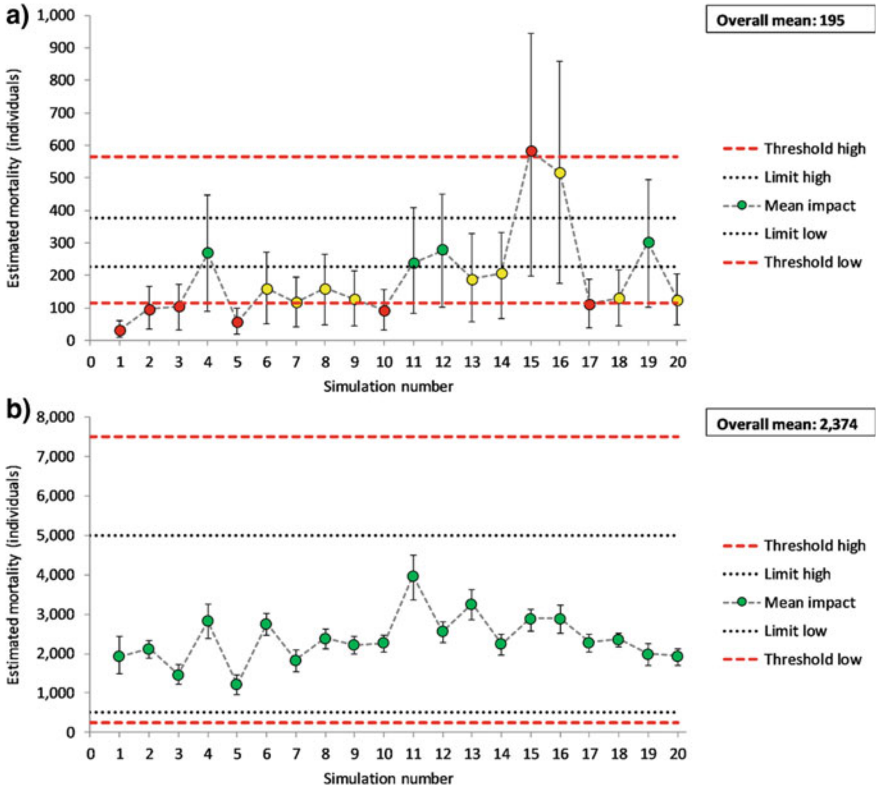


Fig. 4.13 Estimated mortality for marine mammals in the PWS from the 20 oil drift simulations performed in OSCAR shown as the mean and 95% credible intervals. **a** Harbor seal, **b** Sea otter

Tidal Flats) found in GEO (1994). For ESI 4 (Coarse-grained Sand Beaches) and ESI 9 (Sheltered Tidal Flats), ERA Acute numbers are lower than impact reported by surveys, although surveyed impact for this ESI type is limited to only a few km. For ESI 5 (Mixed Sand and Gravel Beaches), ERA Acute numbers are also too low, while they are overestimated for ESI 6 (Gravel, Cobble, Boulder Beaches). The difference might be explained by the classification in the ERA Acute VEC dataset versus the surveyed shoreline as the combined impact for these two ESI rankings are in line with surveyed numbers for Heavy + Medium + Light (HML) oiling. For ESI 8 (Sheltered Rocky Shores) and ESI 10 (Marshes), ERA Acute numbers are in between HM and HML survey numbers. For DHOS, the most affected habitat types in ERA Acute calculations are freshwater marshes, swamps (ESI 10ABC) and scrub-shrub wetlands, mangroves (ESI 10DE) and the ERA Acute calculated shoreline lengths within the five affected states was like the reported shoreline lengths of oiling in the NRDA.

A plausible reason for low impact values in some of the modelled oil trajectory simulations is that the oil drift (induced by wind and current) in these simulations is

Table 4.8 Estimated impact with ERA Acute, the performance boundaries and classification of the estimated impacts for VECs in the shoreline compartment according to the performance boundaries. The number are given as percentage of the simulations falling within the different performance boundaries rounded to nearest whole number

Case	VEC	Estimated impact with ERA acute			Performance boundaries				Classification of ERA acute				
		Mean	$P_{2.5}$[a]	$P_{97.5}$[a]	T low	L low	Within	L high	T low (%)	L low (%)	Within (%)	L high (%)	T high (%)
EVOS	Flora	423	4	144	20	39	185	340	25	10	65	0	0
	Fauna	73	62	732	86	165	788	1,446	5	15	80	0	0
DHOS	Flora	1,137	149	1,695	563	1,161	2,117	3,307	10	35	55	0	0
	Fauna	2,225	601	2,988	704	1,451	2,646	4,134	5	10	55	30	0

[a] Since only 20 simulations have been performed, $P_{2.5}$ and $P_{97.5}$ is equal to the minimum and maximum values

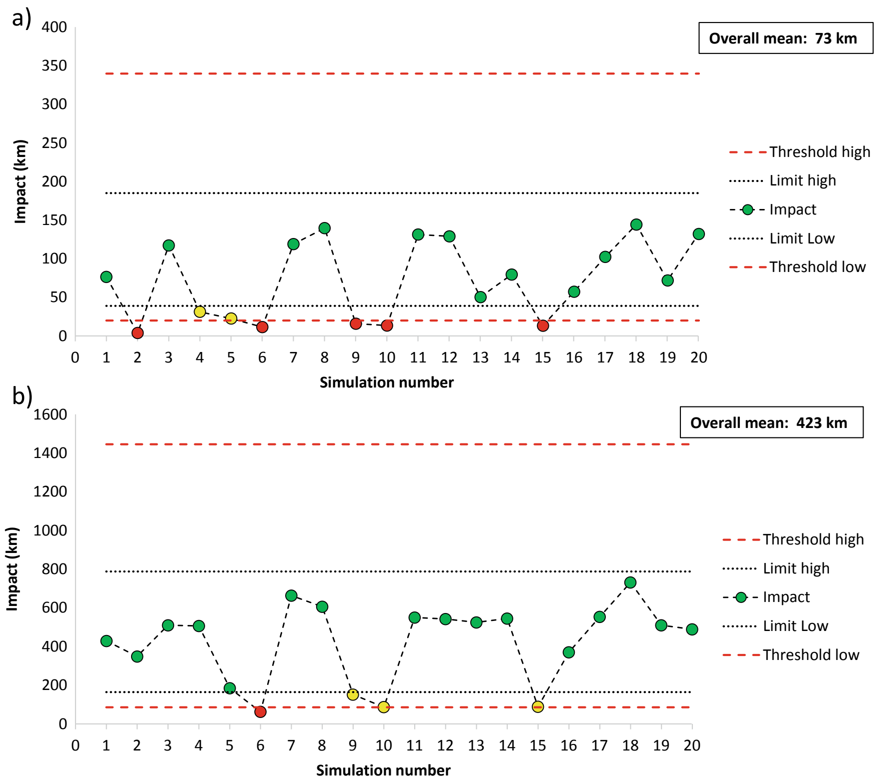


Fig. 4.14 Estimated impact calculated in ERA Acute v.1.1.0.27 for shoreline from the 20 oil drift simulations performed for the EVOS case in OSCAR. **a** Flora, **b** Fauna

not very representative for the incidents, exemplified by the impact area in simulation no. 6 and 15 for EVOS, with wind mainly from southwest during the modelled oil spill and an impact area limited to the Valdez Bay area. In DHOS, much of the oil was apparently trapped in a large stationary eddy on the northern part of the Loop Current that would not necessary be present each year and therefore is not reproduced by the oil drift simulations used as input to ERA Acute.

4.3 Discussion of the Validation

Estimating the extent of injury on natural resources has historically been a contentious, uncertain, and politically charged process. The testing and validation of the ERA Acute model is based on data that have a high degree of uncertainty. This includes oil drift data, data on distribution and abundance of VECs, historical field assessments of injury and the establishment of model parameters. A probabilistic

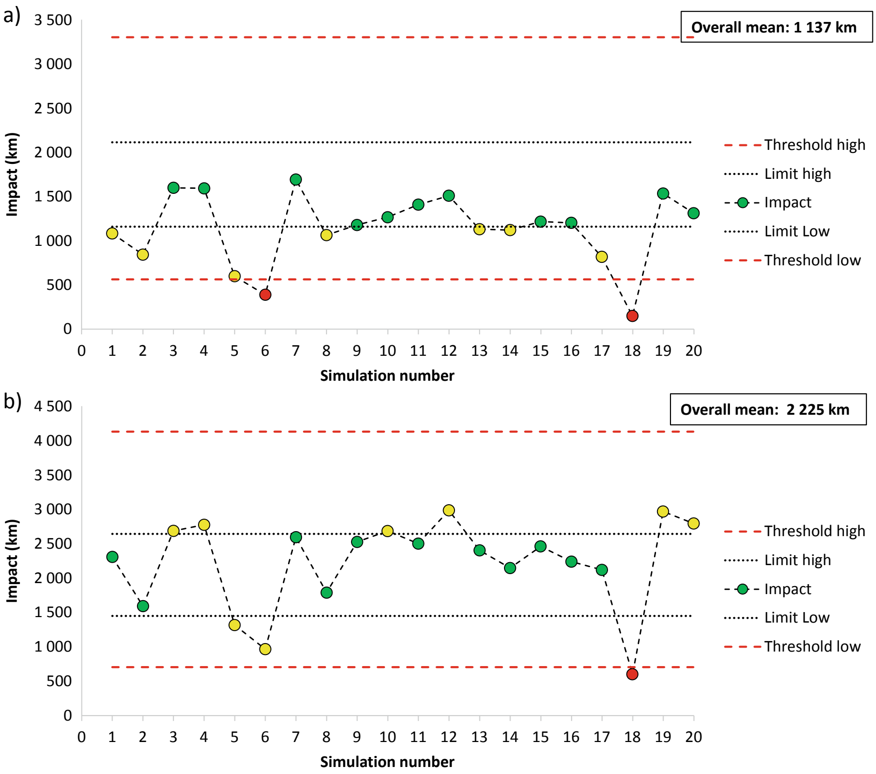


Fig. 4.15 Estimated impact calculated in ERA Acute v.1.1.0.27 for shoreline from the 20 oil drift simulations performed for the DHOS case in OSCAR. **a** Flora, **b** Fauna

approach was used to include some of this uncertainty, including VEC densities and distribution, individual vulnerability towards oil, and for model oil drift—uncertainty in the oil drift parameters.

The oil drift model used as input (OSCAR) performed reasonably well compared to field data estimates, taking into consideration uncertainty in blowout rates, reference oil types and resolution in the driver data and analysis grid. Modelled oil drift data are an important input to ERA Acute (cf. Sect. 1.5.1) and different metocean conditions constitute a significant source of variability in the prediction of spreading of oil between modelled data and actual spill incidents. Therefore, if the oil spill cases used in the validation had occurred at a different time, for example a year earlier, it is likely that the oil trajectory would be different. Much of the oil from the 2010 DHOS was apparently trapped in a large stationary eddy on the northern part of the Loop Current (cf. Wilson et al. 2010 and references therein) that would not necessary be present a different year. If, in the modelling, oil is transported out to sea instead of to the shoreline due to special weather conditions, the impact for the shoreline will be greatly underreported compared to the reported data from the incident.

An important oil drift parameter for seabirds and marine mammals is the exposure time, i.e. how long harmful oil is present in a grid cell. The oil drift model OSCAR estimated considerably shorter exposure time in the grid cells than the exposure time that what was derived from the satellite data (cf. Table 4.6). This difference is the main explanation for the relatively large difference in estimated seabird and marine mammal mortality using modelled oil drift data and oil drift data derived from satellite data in Table 4.7.

The modelled oil drift used as input to the validation study does not include oil spill response. The effect of the oil spill response on the field-estimated impact for the two incidents is not known but it is reasonable to assume that the oil spill response measures implemented during the DHOS reduced the mortality and impacted shoreline area significantly. French-McCay et al. (2018) and Bock et al. (2018) demonstrated that surface oil mass, volume and area were significantly reduced by mechanical recovery, in-situ burning, surface and subsea injection dispersant, and that the relative risks to shoreline-, surface wildlife- and most aquatic life VECs were reduced for a hypothetical deep-water oil well blowout in the Gulf of Mexico. In simulations where shoreline oiling occurred, oil spill response also resulted in less volume ashore and shorter length of shoreline affected. Including oil spill response in the OSCAR model, both offshore and in coastal areas, would have reduced ERA Acute calculated impact on shoreline habitats.

Adequately documenting tests of risk assessment models requires explicit performance criteria against which the model performance is compared (cf. Kirchner et al. 1996; Rykiel 1996). In this study we defined performance criteria based on injury estimates from incident damage assessments, as well as peer reviewed literature, from two oil spill cases. This approach was valuable for evaluating the impact estimated by ERA Acute and also for comparing the performance of two alternative impact functions for the surface compartment (cf. Sect. 3.4.1). However, the performance must be interpreted relative to the width of the boundaries. For instance, the large uncertainty in injury estimates for seabirds after the DHOS incident increases the likelihood of obtaining a high-performance score. The estimated average loss of approximately 150,000 seabirds by ERA Acute was within, but slightly low compared to the performance boundaries. The studies by Haney et al. (2014a, b) was criticized by Sackmann et al. (2015) who suggested that an underestimation of carcass transport probability to shorelines was leading to overestimation of bird deaths by an order of magnitude; a comment which was refuted in a response letter from Haney et al. (2015) (see also Beyer et al. 2016). When compared only against the injury estimates from the DHOS NRDA process, the estimated impact by ERA Acute is somewhat high (conservative), for impacts estimated using both modeled oil drift data and oil drift data derived from satellites.

References—Testing and Validation

Ballachey BE, Bodkin JL, DeGange AR (1994). An overview of sea otter studies. Mar Mamm Exxon Vald 47–59

Beyer J, Trannum HC, Bakke T, Hodson PV, Collier TK (2016) Environmental effects of the deepwater horizon oil spill: a review. Mar Pollut Bull 110:28–51. https://doi.org/10.1016/J.MARPOLBUL.2016.06.027

Bock M, Robinson H, Wenning R, French-McCay D, Rowe J, Walker AH (2018) Comparative risk assessment of oil spill response options for a deepwater oil well blowout: part II. Relat Risk Methodol Mar Pollut Bull 133:984–1000. https://doi.org/10.1016/j.marpolbul.2018.05.032

DNV GL, Acona (2020) ERA acute shoreline re-validation. Memo no: 1007584. 20 Oct 2020

Deepwater Horizon Natural Resource Damage Assessment Trustees (2016) Deepwater horizon oil spill: final programmatic damage assessment and restoration plan and final programmatic environmental impact statement. Chap. 4. Injury to Natural Resources

Drew GS, Piatt JF, Renner M (2015) User's guide to the North Pacific Pelagic Seabird Database 2.0

EVOSTC (2010) Exxon valdez oil spill restoration plan. 2010 Update: Injured resources and services. Exxon Valdez Oil Spill Trustee Council

EVOSTC (2013) Potential recovery of pigeon guillemot populations, naked island Group-Prince William Sound. Prepared by: U.S. Fish and Wildlife Service, U.S. Forest Service, Chugach National Forest, U.S. Animal and Plant Health Inspection Service , Wildlife Services and GAP Solutions, Inc

EVOSTC (2014) Exxon valdez oil spill restoration plan, 2014 Update: Injured Resources and Services. Exxon Valdez Oil Spill Trustee Council

French DP, Rines HM (1997) Validation and use of spill impact modeling for impact assessment. In: International Oil Spill Conference. pp. 829–834

French-McCay D (2004) Oil spill impact modeling: Development and validation. Environ Toxicol Chem, 23(10):2441–2456

French-McCay D (2009) State-of-the-art and research needs for oil spill impact assessment modelling. In: Proceedings of the 32nd AMOP technical seminar on environmental contamination and response. Emergencies Science Division, Environment Canada, Ottawa, ON, Canada, pp 601–653

French-McCay D, Crowley D, Rowe JJ, Bock M, Robinson H, Wenning R, Walker AH, Joeckel J, Nedwed TJ, Parkerton TF (2018) Comparative risk assessment of spill response options for a deepwater oil well blowout. Oil Spill Modeling Mar Pollut Bull 133:1001–1015. https://doi.org/10.1016/j.marpolbul.2018.05.042

Frost KJ, Lowry LF (1994) Assessment of injury to harbor seals in Prince William Sound, Alaska, and adjacent areas following the Exxon Valdez oil spill. Alaska Department of Fish and Game, Wildlife Conservation Division

Garrott RA, Eberhardt LL, Burn DM (1993) Mortality of sea otters in Prince William Sound following the Exxon Valdez oil spill. Mar Mammal Sci 9:343–359

Garshelis DL (1997) Sea otter mortality estimated from carcasses collected after the Exxon Valdez oil spill. Conserv Biol 11:905–916

GEO (1994) The Exxon Valdez Oil spill natural resource damage assessment and restoration. In: A report on oiling to environmentally sensitive shoreline and oiling by Land Status. By the Exxon Valdez Oil spill damage assessment geoprocessing group. Alaska Department of Natural Resources & United States Fish and Wildlife Service. 4 Nov 1994

Gundlach E, Pavia EA, Robinson C, Gibeaut JC (1991) Shoreline surveys at the Exxon Valdez oil spill: the state of Alaska response. In: International Oil Spill Conference, pp 519–529

Haney JC, Geiger HJ, Short JW (2014a) Bird mortality from the Deepwater Horizon oil spill. II. Carcass sampling and exposure probability in the coastal Gulf of Mexico. Mar Ecol Prog. Ser 513:239–252

Haney JC, Geiger HJ, Short JW (2014b) Bird mortality from the Deepwater Horizon oil spill. I. Exposure probability in the offshore Gulf of Mexico. Mar Ecol Prog Ser 513:225–237

Haney JC, Geiger HJ, Short JW (2015) Bird mortality due to the Deepwater Horizon oil spill: reply to Sackmann & Becker. Mar Ecol Prog Ser 534:279–283

Hess NA, Ribic CA (2000) Seabird ecology. In: Davis RW, Evans WE, Würsig B (eds) Cetaceans, sea turtles and seabirds in the northern Gulf of Mexico: distribution, abundance and habitat associations. Vol II: Technical Report. Prepared by Texas A&M University, Galveston, TX and National Marine Fisheries Service, New Orleans, LA. OCS Study MMS 2000-003, pp. 275−316

Hoover-Miller A, Parker KR, Burns JJ (2001) A reassessment of the impact of the Exxon Valdez oil spill on harbor seals (*Phoca vitulina richardsi*) in Prince William Sound. Alaska Mar Mammal Sci 17:111–135

Kirchner JW, Hooper RP, Kendall C, Neal C, Leavesley G (1996) Testing and validating environmental models. Sci Total Environ 183:33–47

Limpert E, Stahel WA, Abbt M (2001) Log-normal distributions across the sciences: Keys and clues on the charms of statistics, and how mechanical models resembling gambling machines offer a link to a handy way to characterize log-normal distributions, which can provide deeper insight into variability and probability—normal or log-normal: that is the question. Bioscience 51:341–352

Lockyer C, Morris R (1990) Some observations on wound healing and persistence of scars in Tursiops truncatus. Reports Int Whal Comm (Special Issue 12):113–118

McFarlane RW, Lester LJ (2005) Determination of nearshore seabird density on the upper Texas coast. Use Radar Hum Obs to Assess Risk Bird Mortalities Potential Wind Turbine Install. Up. Texas Coast, Final Report. State Energy Conserv Off. Austin, TX

Nixon Z, Zengel S, Baker M, Steinhoff M, Fricano G, Rouhani S, Michel J (2016) Shoreline oiling from the Deepwater Horizon oil spill. Mar Pollut Bull 107:170–178. https://doi.org/10.1016/j.marpolbul.2016.04.003

NOAA (2017) ERMA Deepwater Gulf Response. A mapping application a mapping application that displays data from the Deepwater Horizon response and the natural resource damage assessment [WWW Document]. Natl Ocean Atmos Adm. https://gomex.erma.noaa.gov/erma.html

North Pacific Pelagic Seabird Database (NPPSD). https://alaska.usgs.gov/science/biology/nppsd/index.php Dec 2020

Østbye C, Moe KA, Brude OW, Spikkerud CS (2003) EIF Acute Concept Definition; Risk function & Model design. Alpha Memo 1162-03-I

Piatt JF, Anderson P (1996) Response of common murres to the Erron Valdez Oil Spill and and long-term changes in the Gulf of Alaska marine ecosystem. In: Conference paper, proceedings of the exxon valdez oil spill symposium: American fisheries society symposium vol 18, p 737

Piatt JF, Ford RG (1996) How many seabirds were killed by the Exxon Valdez oil spill? In: Conference paper, proceedings of the exxon valdez oil spill symposium: American fisheries society symposium, vol. 18, pp 712–719

Piatt JF, Lensink CJ, Butler W, Kendziorek M, Nysewander DR (1990) Immediate impact of the Exxon Valdez'oil spill on marine birds. Auk 107:387–397

Roberts JJ, Best BD, Mannocci L, Fujioka E, Halpin PN, Palka DL, Garrison LP, Mullin KD, Cole TVN, Khan CB, McLellan WM, Pabst DA, Lockhart GG (2016) Habitat-based cetacean density models for the U.S. Atlantic and Gulf of Mexico. Scientific Reports 6: 22615. https://doi.org/10.1038/srep22615. https://seamap.env.duke.edu/models/Duke-EC-GOM-2015/ Dec 2020

Rykiel EJ (1996) Testing ecological models: the meaning of validation. Ecol Modell 90:229–244

Sackmann BS, Becker, DS (2015) Bird mortality due to the Deepwater Horizon oil spill: comment on Haney et al. (2014a, b). Mar Ecol Prog Ser 534:273–277

Sanger GA, Cody MB (1994). Survey of pigeon guillemot colonies in Prince William Sound. Alaska, Exxon Vald. Oil Spill Restor. Project final report (Restoration Project 93034), US Fish Wildl Serv Anchorage, Alaska 4

SINTEF (2016). MEMW, (Marine Environmental Modeling Workbench) User Manual. Version 8.01. SINTEF Materials and Chemistry, Marine Environmental Technology. Trondheim, Norway

Spikkerud, CS, Brude, OW, Randall, M, Moe KA (2004) Criteria for choice and prioritisation of resources for levels II and III—Selection guidelines, Adaptation Guidelines for Resource Data, Application of Guidelines on Test Resource Data for EIF Levels II and III. Alpha report: 1190-01

Spikkerud, CS, Skeie, GM, Hoell E, Reed M, Brude OW, Bjørgesæter A (2010) ERA Acute, Oil spill risk assessment tool. Phase 1—design basis for model Level A. Akvaplan-niva AS report: 4531.01

Stephansen C, Bjørgesæter A, Brude OW (2018) WP2d Post-testing calibration of model. Changes to calculator, parameters and future research needs Report ID: ERA Acute 2D-1

Tasker ML, Jones PH, Dixon TIM, Blake BF (1984) Counting seabirds at sea from ships: a review of methods employed and a suggestion for a standardized approach. Auk 101:567–577

Udevitz MS, Ballachey BE and Bruden DL (1996) A population model for sea otters in western Prince William Sound. Exxon Valdez oil spill restoration project final report (restoration project 93043-3). Natl Biol Serv, Anchorage, AK. 34 p. https://evostc.state.ak.us/media/2526/2010-10100750a-final.pdf

US Coast Guard (2011) On scene coordinator report Deepwater Horizon Oil Spill. Submitted to the National Response Team Sept 2011

Wilson M, Graham L, Hale C, Maung-Douglass E, Sempier S, Skelton T, Swann L (2010) Deepwater Horizon: where did the oil go? oil spill science: sea grant programs of the Gulf of Mexico, no. GOMSG-G-17-006. https://gulfseagrant.org/