Persistence and biodegradation of oil at the ocean floor following *Deepwater Horizon*

Sarah C. Bagby[a,b], Christopher M. Reddy[c], Christoph Aeppli[d], G. Burch Fisher[e], and David L. Valentine[a,b,1]

[a]Department of Earth Science, University of California, Santa Barbara, CA 93106; [b]Marine Science Institute, University of California, Santa Barbara, CA 93106; [c]Department of Marine Chemistry and Geochemistry, Woods Hole Oceanographic Institution, Woods Hole, MA 02543; [d]Bigelow Laboratory for Ocean Sciences, East Boothbay, ME 04544; and [e]Jackson School of Geosciences, University of Texas at Austin, Austin, TX 78712

Edited by Donald E. Canfield, Institute of Biology and Nordic Center for Earth Evolution, University of Southern Denmark, Odense M., Denmark, and approved November 20, 2016 (received for review June 22, 2016)

The 2010 *Deepwater Horizon* disaster introduced an unprecedented discharge of oil into the deep Gulf of Mexico. Considerable uncertainty has persisted regarding the oil's fate and effects in the deep ocean. In this work we assess the compound-specific rates of biodegradation for 125 aliphatic, aromatic, and biomarker petroleum hydrocarbons that settled to the deep ocean floor following release from the damaged Macondo Well. Based on a dataset comprising measurements of up to 168 distinct hydrocarbon analytes in 2,980 sediment samples collected within 4 y of the spill, we develop a Macondo oil "fingerprint" and conservatively identify a subset of 312 surficial samples consistent with contamination by Macondo oil. Three trends emerge from analysis of the biodegradation rates of 125 individual hydrocarbons in these samples. First, molecular structure served to modulate biodegradation in a predictable fashion, with the simplest structures subject to fastest loss, indicating that biodegradation in the deep ocean progresses similarly to other environments. Second, for many alkanes and polycyclic aromatic hydrocarbons biodegradation occurred in two distinct phases, consistent with rapid loss while oil particles remained suspended followed by slow loss after deposition to the seafloor. Third, the extent of biodegradation for any given sample was influenced by the hydrocarbon content, leading to substantially greater hydrocarbon persistence among the more highly contaminated samples. In addition, under some conditions we find strong evidence for extensive degradation of numerous petroleum biomarkers, notably including the native internal standard 17α(H),21β(H)-hopane, commonly used to calculate the extent of oil weathering.

Deepwater Horizon | biodegradation | oil spills | hydrocarbon | petroleum biomarkers

On 20 April 2010, a blowout from the Macondo Well in the Gulf of Mexico (GOM) caused an explosion on the *Deepwater Horizon* (DWH) mobile offshore drilling unit that ultimately led to its sinking and the deaths of 11 crewmembers. From the time of the blowout until the well was capped on 15 July 2010, petroleum fluids flowed continuously from the Macondo Well, with environmental emission estimates of 4.1 million barrels of oil and 1.7×10^{11} g natural gas (1–3). The spill was noteworthy not only for its volume but also for its distance offshore and its depth: Oil and gas entered the ocean at a water depth of ∼1,500 m and then partitioned between the deep ocean and the sea surface. This partitioning may have varied over time because of reservoir depressurization and deliberate interventions such as the shearing of the riser pipe and the application of chemical dispersant at the wellhead (4–6). In all, approximately half of the oil ascended to the ocean surface (1, 7), where it was skimmed or flared by response teams, trapped in sinking particles by marine oil snow sedimentation and flocculent accumulation (8, 9), washed ashore, or left exposed to the canonical weathering processes of evaporation, biodegradation, and photooxidation (7, 10). The rest remained in the deep ocean. Because the DWH event was the first major spill to occur in the deep ocean, the processes determining the fate of this oil were largely unknown.

In the wake of the spill, water-column data shed light on the physical partitioning of the submerged oil. Many compounds containing <10 carbon atoms (e.g., natural gas, benzene and its alkylated analogs, cycloalkanes, and branched alkanes) dissolved in seawater to form deep, aqueous plumes (2, 6, 11–16); in the first weeks of the spill, dissolution is also expected to have influenced the distribution of two- and three-ring polycyclic aromatic hydrocarbons (PAHs), particularly naphthalene and its alkylated analogs, and to a lesser extent fluorene, phenanthrene, and anthracene and their alkylated analogs. Hydrocarbons that remained undissolved became trapped in the deep ocean in a suspension of small (less than ∼100 μm) droplets of liquid oil that lacked the buoyant force to rise through the water column. These droplets remained concentrated close to the well's coordinates (2, 11, 12, 15), but modeling suggests that droplet size drove further vertical partitioning, with droplets >50 μm mixing upwards by August 2010 and smaller droplets remaining suspended in the deep ocean (7, 17, 18). Some suspended oil was eventually deposited to the seafloor, likely via oil–mineral aggregates or microbial flocs (8, 19, 20), with intense contamination within ∼5 km of the well (21–26). Surficial sediments near the well were found to carry >1,000-fold–elevated concentrations of dioctyl sodium sulfosuccinate (4), an active ingredient of the chemical dispersant applied at the wellhead, and to exhibit a radiocarbon deficit consistent with oil deposition (27). We recently identified a 3,200-km^2 deposition footprint stretching southwest from the wellhead (28). This footprint, marked by substantial heterogeneity in the oil mass of deposited particles, was estimated to account for ∼4–31% of the submerged oil (28).

Significance

The *Deepwater Horizon* event led to an unprecedented discharge of ∼4.1 million barrels of oil to the Gulf of Mexico. The deposition of ∼4–31% of this oil to the seafloor has been quantified previously on a bulk basis. In this work, we assess the extent of degradation over 4 y postspill for each of 125 petroleum hydrocarbons that contaminated the seafloor. As expected, chemically simpler compounds broke down more quickly than complex compounds, but degradation rates also depended on environmental context: Breakdown often was faster before seafloor deposition than after and for oil trapped in small droplets than for oil in large particles. These results provide a basis to predict the long-term fate of seafloor oil.

Author contributions: S.C.B., C.M.R., C.A., and D.L.V. designed research; S.C.B. and G.B.F. performed research; S.C.B., G.B.F., and D.L.V. analyzed data; and S.C.B., C.M.R., C.A., G.B.F., and D.L.V. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

Data deposition: All raw data analyzed in this article are publicly available via the Natural Resource Damage Assessment website (https://dwhdiver.orr.noaa.gov/explore-the-data/). The cleaned and filtered dataset, together with all analysis code used in the paper and a README file describing the use of the code, has been bundled with all necessary R packages to make a functional research compendium. This compendium has been deposited with Figshare, with the DOI 10.6084/m9.figshare.4001262.

[1]To whom correspondence should be addressed. Email: valentine@geol.ucsb.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1610110114/-/DCSupplemental.

Superimposed on these changes in physical distribution, hydrocarbons trapped in the deep ocean were subject to biologically mediated loss processes (i.e., biodegradation) (12, 16, 18, 29–31). For sparingly soluble hydrocarbons, biodegradation is expected to serve as the primary cause of weathering in the deep ocean because other key weathering processes (e.g., evaporation, photooxidation) depend on atmospheric and solar exposure. Although hydrocarbon recalcitrance to biodegradation is expected to scale roughly with molecular mass and steric complexity (32), the rates at which specific hydrocarbons are metabolized vary based on myriad environmental variables: temperature, salinity, pressure, oxygen concentration, pH, solar exposure, availability of nutrients and other sources of organic matter (33), water availability, access to substrate, solid-phase interactions, competition, predation, and inhibition (34). Although early reports addressed the degradation of some low- to moderate-molecular-mass compounds in the water column (12, 14, 16), the degradation rates of the petroleum hydrocarbons constituting Macondo oil after seafloor deposition are unknown. However, it is these rates, and the factors that control them, that will largely determine the long-term fate and biological impacts of the spill on the GOM seafloor.

In this work we address the fate of oil that was deposited on the floor of the deep ocean following the DWH event. We use publicly available data from the Natural Resource Damage Assessment (NRDA) process to identify samples contaminated by oil from the Macondo Well conservatively and to analyze the rate and extent of biodegradation for 125 hydrocarbon compounds spanning 4 y postspill. Based on the results, we identify key factors that modulated biodegradation, finding that a dependence on the intensity of contamination overlaid the expected trends in chemical structure and complexity.

Results

Identification of Macondo-Contaminated Samples. In previous work, we established that surficial GOM sediments not contaminated with Macondo oil typically bear $<$75 ng/g of the native internal standard $17\alpha(H),21\beta(H)$-hopane and that the large majority of contamination lies within 40 km of the wellhead, in a footprint extending southwest from the wellhead (28). However, this region is also the site of ongoing natural seepage, and inclusion of any seeped oils could severely distort analysis of the time course of seafloor Macondo oil weathering. Thus, the first goal in our analysis was to identify a subset of NRDA samples in which the detected petroleum hydrocarbons originate unambiguously from the Macondo Well.

This task is complicated by there being no "smoking gun" compound that is present in Macondo Well oil but absent from local seeps. Nonetheless, the relative abundances of the suite of recalcitrant native petroleum compounds commonly referred to as "biomarkers" constitute a chemical fingerprint for Macondo Well oil. We used 12 diagnostic biomarker ratios capturing a broad range of these compounds' structural diversity to develop an aggregate Macondo dissimilarity index (hereafter, MDI), as detailed in the *SI Appendix*, *SI Text*. The spatial structuring of MDI fingerprint results (Fig. 1 and *SI Appendix*, Fig. S2), particularly the appearance of distinct patterns of biomarker ratios deeper in the sediment column where seep oil is expected to be the dominant source (25, 26), suggests that the matching process successfully identified samples contaminated by Macondo oil against a low-level background of chemically distinct seep oils.

Of the 2,980 sections of sediment cores collected in the NRDA sampling effort (*SI Appendix*, *SI Text*), 443 matched the MDI fingerprint and hereafter are referred to as "Macondo-contaminated samples." However, these samples originated from sediment depths of up to 5 cm. Although spilled oil may be pushed downcore in the course of sampling, it also can be mixed downward in sediment by bioturbation in situ (35), exposing it to different oxygen concentrations and microbial communities for an unknown period. To reduce heterogeneity in the dataset, we exclude all samples with upper depth $\geq$0.5 cm from further consideration, yielding a conservative final set of 312 samples of Macondo-contaminated surficial sediments.

These samples were collected at 272 stations in the course of 14 cruises from September 2010 to June 2014, at seafloor depths ranging from 1,029 to 1,912 m (median, 1,494 m; interquartile range, 1,394–1,568 m). Notably, only ∼5% of these Macondo-contaminated samples were collected at depths consistent with "bathtub ring" deposition by deep plumes impinging on the continental slope, and ∼95% lie within the deeper, more heavily contaminated fallout plume (Fig. 1). The contamination in this region has previously been shown to derive from oil from the deep plume rather than from oil that rose to the surface, weathered, and sank again to the seafloor (28). In that work, we established 28 ng/g as the mean background concentration of hopane in surficial sediments $\geq$40 km from the wellhead; 97% of surficial sediments sampled at this distance from the well contained $<$75 ng/g hopane. In the Macondo oil–contaminated surficial sediments we identify here with the MDI, hopane concentrations range from 43.8 to 14,100 ng/g, indicating a $>$320-fold range in the severity of contamination. Among the petroleum hydrocarbons measured in these samples, we identified a set of 125 aliphatic compounds, PAHs, and polycyclic isoprenoid and triaromatic sterane petroleum biomarkers with sufficiently high data quality (*SI Appendix*, *SI Text*) to permit kinetic analysis. The resulting dataset comprises 34,769 measurements, a median of 310 measurements per compound.

Initial Observations of Hydrocarbon Weathering Kinetics. To begin our analyses of these compounds, we used hopane as an internal standard, first normalizing each analyte concentration to the hopane concentration in that sample (36, 37) and then normalizing that ratio to the corresponding ratio in source oil. We refer to these doubly normalized values as the "fraction remaining" of a given compound. On initial examination of single-compound datasets, we observed what appeared to be a contamination-level dependence in the change in fraction remaining over time (*SI Appendix*, Fig. S3*A*). To test this apparent pattern, we binned samples by hopane concentration (low, $<$400 ng/g; moderate, 400–750 ng/g; high, $\geq$750 ng/g) as a proxy for contamination level and asked whether the distribution of fraction-remaining data differed significantly ($P < 0.05$, Kolmogorov–Smirnov test) across bins (*SI Appendix*, Fig. S3 *B* and *C*). We found significant differences between the low- and high-contamination bins for 94% of the compounds and significant differences between the low- and moderate-contamination bins for 60% of the compounds. This result raises the strong possibility that oil-particle size influences weathering rates, perhaps because of the relatively limited bioavailability of oil in large particles, oxygen or nutrient drawdown within large particles, or distinct deposition histories for different particle-size classes. In light of this result, we assessed biodegradation rates separately for the low-, moderate-, and high-contamination bins. We had sufficient data (*Methods*) to analyze 353 (of a possible 372) compound–contamination bin datasets.

To a first approximation, biodegradation is expected to follow simple exponential decay kinetics from a starting point of fraction remaining = 1 at t_0. However, each particle of deposited oil has weathered in two distinct physical environments, first in suspension and then on the seafloor, with considerable differences in key environmental variables including advection, temperature, oxygen and nutrient concentration, pressure, and, potentially, microbial community abundance and composition. Thus, we considered the possibility that the best fit might be a biphasic (broken-stick) model, with independent decay rates in the two phases. It should be noted that, with sediment and source-oil data only, we have end-members for the first phase of such a model, but we lack direct measurements to define the trajectory of that phase. Initial loss could have obeyed first-order kinetics or a more complex trajectory—e.g., a rapid physical process associated with discharge,

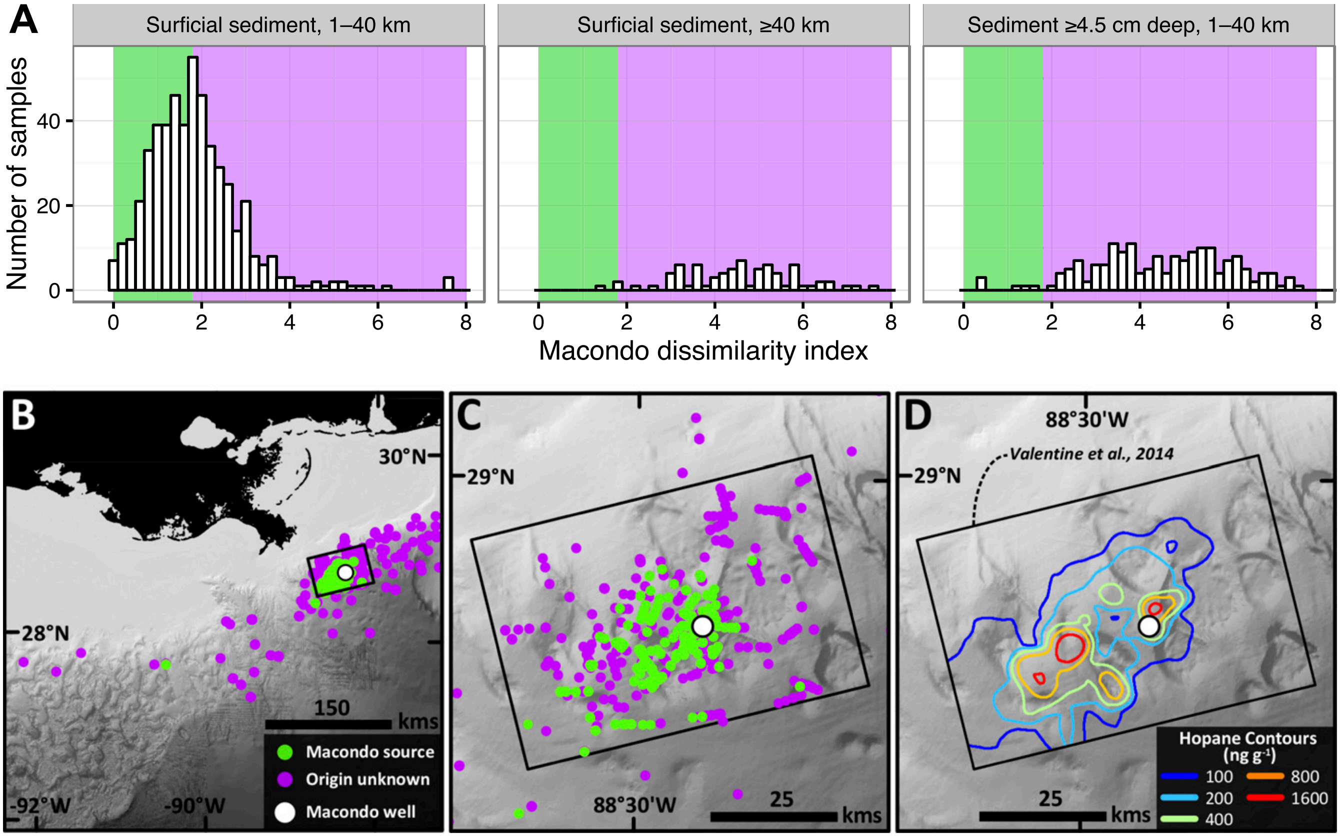

Fig. 1. Application of the MDI to NRDA sediment samples. (*A*) Spatial distribution of MDI values. (*Left*) Surficial sediments (upper depth = 0 cm) collected 1–40 km from the wellhead. (*Center*) Surficial sediments collected ≥40 km from the wellhead. (*Right*) Downcore sediments (upper depth ≥4.5 cm) collected 1–40 km from the wellhead. Samples falling in the green region (MDI <1.8) are consistent with Macondo oil. (*B*) Bathymetric chart of the region around the wellhead showing MDI results for each sample collected. Green symbols, MDI <1.8; purple symbols, MDI ≥1.8. (*C*) Zoomed view of *B* showing detail in the immediate vicinity of the wellhead. (*D*) Footprint of seafloor oil deposition in the immediate vicinity of the wellhead as detected by hopane-concentration anomalies in previous work (28).

followed by biodegradation and/or dissolution from suspended particles. As a third possibility, we considered a simple exponential decay model in which the *y* intercept was allowed to vary freely.

Because oil continued to flow from the wellhead for 87 d, there is a large intrinsic uncertainty in the length of time a given particle of deposited oil was exposed to weathering before sample collection. To cope with this uncertainty explicitly, for each analyte at each contamination level, we generated 100 pseudoreplicate datasets with different randomized time offsets of 0–87 d added to each data point. We fit our three models to each pseudoreplicate, using the Bayesian information criterion (BIC) to choose the best fit, and we report the median of best-fit predictions.

To assess the global quality of our fits, we examined the so-called "pull" distribution, i.e., the distribution of (fitted slope)/(error on fit) (38). The high tails in these distributions, most notably at low and moderate contamination levels, indicate a pathology in the hopane-normalized dataset (*SI Appendix, SI Text* and Fig. S4). High pull values were especially common among aliphatic compounds of at least 29 carbons, compounds that, when normalized to hopane, appear to show increasing fraction remaining over time (*SI Appendix*, Fig. S4 *D* and *E*). The simplest explanation for this behavior is that hopane is not, in fact, conservative for these conditions but rather is more labile than the most recalcitrant aliphatics and aromatics.

Such hopane lability is not unusual. Substantial hopane degradation has been observed before in laboratory settings (39) and in other environments; studying the OSSA II spill in the Bolivian Altiplano, Douglas et al. (40) found that the long-chain alkane *n*-C40 provided a more conservative basis for normalization than hopane. The longest-chain aliphatic present in Macondo oil at sufficient concentrations to be useful for normalization is octatriacontane, *n*-C38. Accordingly, we renormalized our data to *n*-C38 concentrations and repeated our analysis. (To avoid confusion, we continue to categorize samples as showing low, moderate, or high contamination using the bins established above.) The resulting pull distributions were far less distorted (*SI Appendix*, Fig. S4*B*), indicating that normalization to *n*-C38 offered a far more reliable basis for fitting than normalization to hopane. In the remainder of this work we refer exclusively to analysis of *n*-C38–normalized concentration data.

Factors Controlling Seafloor Hydrocarbon Weathering Rates. We performed 35,300 head-to-head model comparisons (100 pseudoreplicates for each of 353 datasets), finding strong statistical support (ΔBIC ≥6) for any model in just under half the comparisons (17,115). Strikingly, the biphasic (broken stick) model was the best-fit model in every such case. Extensive early loss was far more common among pseudoreplicates best fit by the biphasic model than among those best fit by the single-phase models (*SI Appendix*, Fig. S5). The biphasic model was disproportionately likely to be the best-fit model in the low- and moderate-contamination bins and among aliphatic and aromatic compounds (*SI Appendix*, Fig. S5). A subset of data and fits is presented in *SI Appendix*, Fig. S6.

These results are consistent with a model in which all compounds are subject to two phases of weathering, but the transition between phases is obscured when the first phase is either retarded by chemical recalcitrance or low diffusivity or is limited by large particle size. Particle size could influence first-phase weathering either through bioavailability, in that hydrocarbons trapped in larger particles could be comparatively inaccessible to oil-degrading microbes, or through deposition dynamics, e.g., if larger particles tended to be deposited more rapidly, after less exposure to the plume's comparatively favorable conditions for biodegradation. Because the earliest samples in the sediment dataset were collected 160 d postexplosion, we cannot distinguish between these possibilities.

Hydrocarbon molecular mass and structure typically influence biodegradation rates, with progressively slower degradation with increasing molecular mass, ring number, and alkyl branching (32, 34, 41–44). We tested the validity of these relationships for the seafloor by comparing the residual fraction for each hydrocarbon remaining in the sediment 4 y after the spill began. Among the compounds examined, carbon skeletons range from nine to 37 atoms (aliphatics, 9–37; aromatics, 9–22; biomarkers, 23–35) and vary in complexity from the straight-chain aliphatic *n*-C9 to the pentacyclic, multiply substituted biomarker pentakishomohopane. This analysis provides an unparalleled window into the disposition of oil following the DWH event, in that the extent of biodegradation is quantified simultaneously for 125 petroleum hydrocarbons across wide-ranging contamination levels. The results of this analysis clearly show the influence of molecular mass and structure on the extent of biodegradation (Fig. 2, Table 1, and *SI Appendix*, Table S1).

Among straight-chain aliphatic compounds, the extent of degradation after 4 y changes sharply at chain lengths of 28 or 29 carbons (Fig. 3 and Table 1). Among longer chains, across all contamination levels, the fraction remaining after 4 y increases steadily with chain length, reaching 100% for *n*-C37; by contrast, degradation is almost entirely complete for shorter chains (Fig. 3 and Table 1). Indeed, shorter-chain compounds are largely lost by 160 d at low and moderate contamination levels; at high contamination levels, the loss of these compounds at 160 d is substantial but far from complete (*SI Appendix*, Fig. S7). Branching of the carbon backbone is expected to slow biodegradation (32); this effect is not detectable in the light- and moderate-contamination bins, but at high contamination levels the branched compounds show significantly less biodegradation (24–62% contamination remaining) by 160 d postexplosion than their straight-chain counterparts (6–33%). Even in these highly contaminated samples, the degradation of branched-chain aliphatics is largely complete by 4 y postexplosion (Figs. 2 and 3).

Among aromatic compounds, chemical complexity begins to retard biodegradation at lower molecular masses (Figs. 4 and 5, Table 1, and *SI Appendix*, Table S1). At low and moderate levels of contamination, biodegradation is largely complete by 160 d for compounds of <16 or 17 carbons but not for larger compounds (Table 1 and *SI Appendix*, Figs. S8 and S9). Postdeposition biodegradation is detectable for most larger compounds at low and moderate levels of contamination (low contamination: 19 of 21 compounds; moderate contamination: 13 of 21 compounds). At high contamination levels, the largest compound for which biodegradation is nearly complete (<5% remaining) by 160 d is the 14-carbon PAH phenanthrene, and only half of the smaller compounds are degraded to <5% remaining by 160 d (*SI Appendix*, Figs. S8 and S9). Although postdeposition biodegradation is detectable for 15

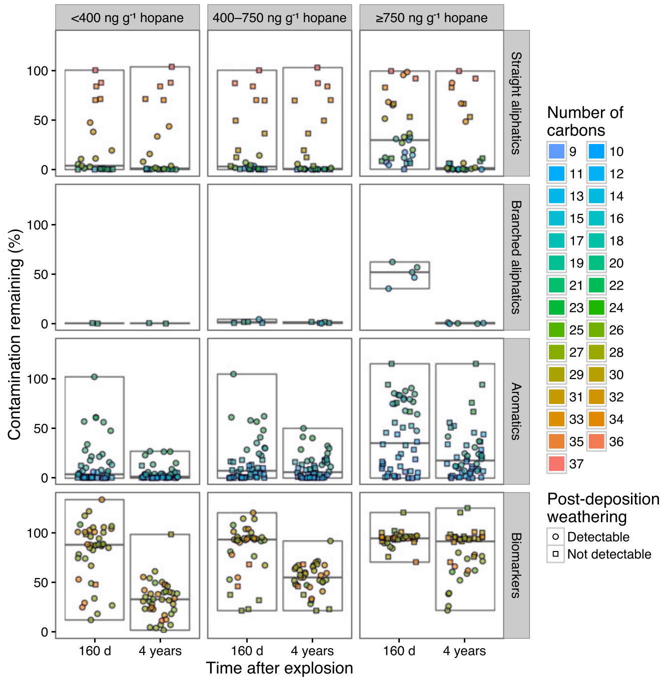

Fig. 2. Overview of the relationship between carbon skeleton size and structure and the extent of biodegradation at 160 d and 4 y postexplosion. Symbols are colored by the number of carbons in the skeleton; symbol shape indicates whether postdeposition biodegradation was or was not detectable for each compound. Results are the median of 100 pseudoreplicate fits for each compound–contamination bin dataset.

Table 1. Size-dependent differences in persistence of aliphatic and aromatic compounds

		Percent remaining after 160 d				Percent remaining after 4 y			
Class	Hopane, ng/g	Boundary position*		Among smaller analytes	Among larger analytes	Boundary position*		Among smaller analytes	Among larger analytes
Aliphatics	<400	C28	Median	0.6	70.2	C29	Median	0.4	70.8
			Range	0.1–5.5	11.2–100.4		Range	0.1–3.7	15.8–103.8
	400–750	C28	Median	1.1	69.8	C28	Median	0.5	69.8
			Range	0.1–7.2	12.5–100.3		Range	0.1–3.6	12.5–103
	≥750	C13	Median	6.1	36.5	C29	Median	0.6	74.8
			Range	0.4–17.5	6.2–99.5		Range	0–11.4	48.5–99.5
Aromatics	<400	C17	Median	0.5	26.2	C20	Median	0.9	23.3
			Range	0–6.5	7.1–102.3		Range	0–12.6	15.3–26.9
	400–750	C16	Median	1.2	28.7	C16	Median	1.2	18.3
			Range	0.1–12.5	6.5–105.1		Range	0.1–12.5	6.5–50.4
	≥750	C14	Median	5.1	75.1	C14	Median	5.1	24.9
			Range	0.1–44.2	11.9–115.5		Range	0.1–44.2	7.3–115.5

*Boundary position is the number of carbons below which loss is so extensive that no trend in percent remaining vs. carbon number can be distinguished and above which there is a clear size trend in percent remaining. The boundary between these regimes can be seen most clearly in Fig. 3 and *SI Appendix*, Fig. S7, where the y axes are ordered by carbon number. Higher chemical lability and lower contamination levels favor more extensive biodegradation, pushing the boundary between regimes to larger carbon numbers.

of 32 larger aromatic compounds at high contamination levels, none of these compounds is degraded to <5% remaining by 4 y postexplosion (Figs. 4 and 5 and Table 1). High contamination also highlights the pattern of relatively fast loss of unsubstituted parent PAHs and slower loss of alkylated daughter PAHs (Figs. 4 and 5 and *SI Appendix*, Figs. S8 and S9).

The largest and most complex hydrocarbons analyzed are the biomarkers: polycyclic isoprenoids including hopanes and diahopanes, terpanes, and steranes and diasteranes, as well as triaromatic steranes. As discussed above, our initial analysis of hopane-normalized data pointed to some degree of biomarker weathering. As analysis of *n*-C38–normalized data makes clear, the large majority of biomarkers analyzed are subject to biodegradation on the timescale studied, consistent with previous observations of Macondo oil in slicks, oiled sands, and deep-sea corals (Fig. 6 and *SI Appendix*, Table S1) (45–47). Biodegradation is limited in all contamination bins before deposition (*SI Appendix*, Fig. S10; median predictions at 160 d: low contamination, 89% remaining; moderate contamination, 94% remaining; high contamination, 95% remaining) but is extensive in the low- and moderate-contamination bins after deposition (median predictions at 4 y: low contamination, 33%; moderate contamination, 56%; high contamination, 92%). For biomarkers, we find no clear relationship between carbon number and degradation rate (Figs. 2 and 6), likely because some of the larger biomarkers (e.g., homohopanes) bear relatively labile alkyl chain substituents (45).

Although these findings are consistent with established biodegradation patterns, they contradict the biodegradation trends reported by Hazen et al. in early analysis of water-column samples (12). Their work suggested half-lives of 0.6–9.5 d for alkane compounds in suspended oil, with no discernible trend related to molecular mass or methyl branching. The discrepancy between that work and the results obtained here led us to reanalyze the dataset they studied. We find two significant shortcomings in their analysis. First, although they set out to determine biodegradation rates within the deep plume of oil, the depth of the plume varied from place to place. By using all samples collected at water depths of 1,099–1,219 m, regardless of station, they included numerous samples collected outside local concentration peaks. Because hydrocarbon concentration dropped sharply above and below the plume, inclusion of these out-of-plume samples introduced a low bias to the resulting half-lives. Second, heightening this bias, they analyzed raw concentration data, without normalizing to the concentration of a conservative compound. Analysis of normalized, in-plume data for the alkanes studied by Hazen et al. (12) (*SI Appendix*, Fig. S11) reveals half-lives 10-fold to ∼50-fold longer than previously reported values (*SI Appendix*, Fig. S12). We find no evidence for the extraordinarily rapid, structure-independent degradation rates originally reported.

Discussion

Dissimilarity Fingerprinting Advances Deep-Sea Oil Spill Forensics. Although a Macondo fingerprint for use in oily samples has been developed recently (48), the MDI fingerprint developed for this work represents a significant forensic advance for the study of the DWH spill in sediments. Our approach is independently supported by the good spatial agreement found between the deposition footprint as defined by the MDI and the footprint as previously defined by hopane (28) and natural abundance radiocarbon (27) anomalies. The MDI offers an advantage over these methods in its ability to distinguish between seeped and spilled oil in individual samples: Among sediment samples that do not meet our MDI threshold, a distinct and coherent fingerprint emerges at greater distances from the Macondo Well and lower depths in the sediment column, likely representing weathered oil that originated in natural seeps. This sensitivity suggests that comparable dissimilarity approaches may be useful for analysis of future spills.

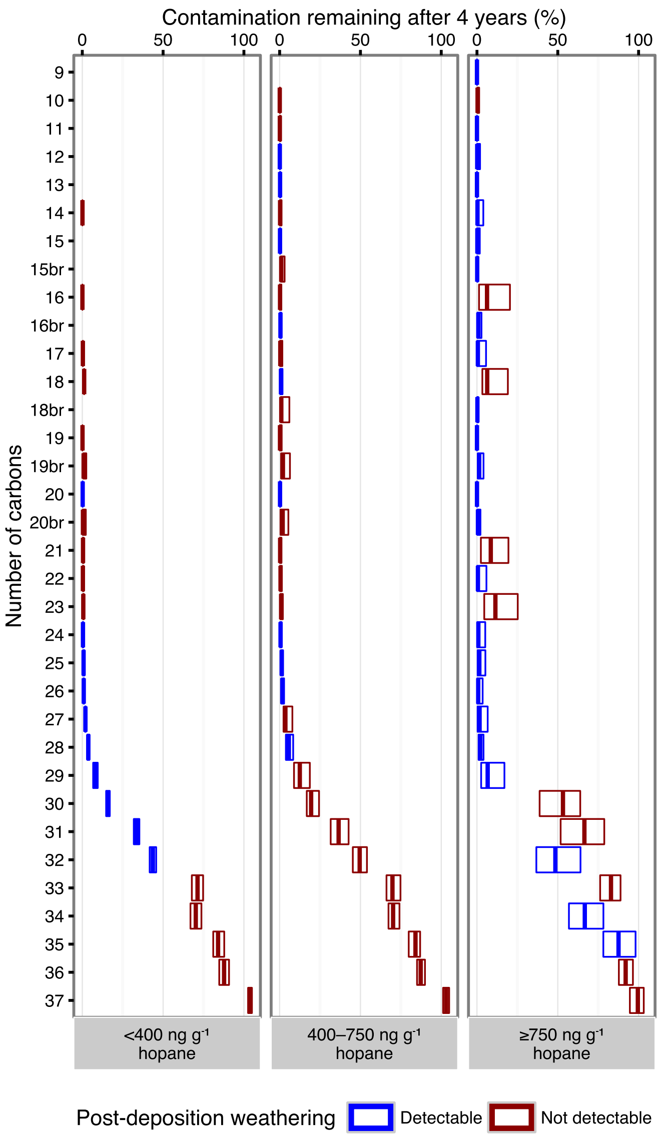

Fig. 3. Percent of aliphatic compounds remaining at 4 y postexplosion, ordered by chain length. Branched compounds are indicated by "br" on the *y* axis. Compounds for which biodegradation was detectable after deposition are shown in blue, with crossbars indicating the fitted value and boxes indicating the 95% confidence interval (CI) of the median fit result. Compounds for which postdeposition biodegradation was not detectable are shown in red, with vertical bars indicating the median and boxes indicating the interquartile range of measured values.

In light of the degradation of biomarkers described above, it is reasonable to ask on what timescale a biomarker-based fingerprint can remain diagnostic. Starting with Macondo oil, we calculated the projected time-course changes in the ratios used to calculate MDI (*SI Appendix*, Fig. S13). We find that in sediments with low contamination the MDI as described here should remain useful for ∼5.4 y; it should remain useful for ∼10 y in moderately contaminated sediments and for ∼5 y in highly contaminated sediments. Notably, the biomarkers whose loss causes the MDI to drift over time differ across contamination levels: Steranes dominate the loss of discrimination at low contamination levels, hopanes at moderate contamination levels, and triaromatic steranes at high contamination levels (*SI Appendix*, Fig. S13*B*). Although the MDI's useful lifetime might be extended by relaxing the threshold to account for drift or by modifying the chosen set of biomarker ratios

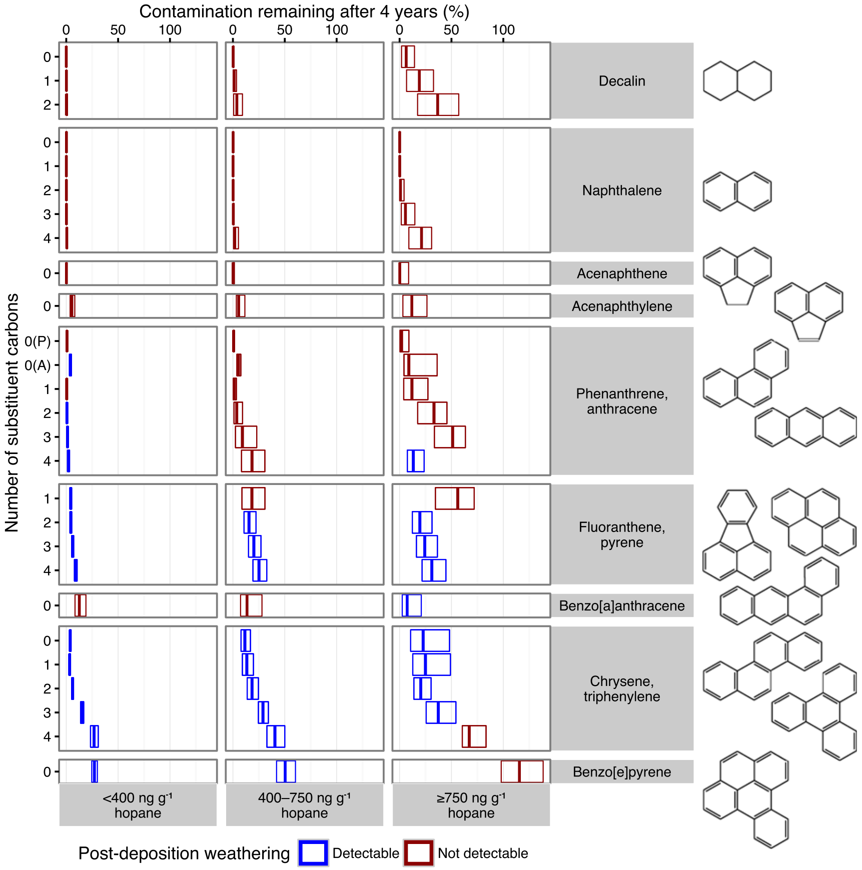

Fig. 4. Percent remaining of aromatic compounds with six-membered rings 4 y postexplosion. (Decalin is not aromatic but is included here.) Panels are ordered by increasing carbon skeleton size from top to bottom and within each panel by increasing number of carbon substituents. Where multiple carbon skeletons are shown for a single group at right, the compounds in that category were not separately resolved in chemical analysis, unless otherwise indicated on the *y* axis. 0(P), unsubstituted phenanthrene only; 0(A), unsubstituted anthracene only. Compounds for which biodegradation was detectable after deposition are shown in blue, with crossbars indicating the fitted value and boxes indicating the 95% CI of the median fit result. Compounds for which post-deposition biodegradation was not detectable are shown in red, with vertical bars indicating the median and boxes indicating the interquartile range of measured values.

(45), these changes would likely come at the cost of increasing false positives.

Pseudoreplicates Can Address Uncertainties in Weathering. Although Monte Carlo methods are common in other fields, they have not typically been used in oil-spill assessment. Here, we used ensembles of pseudoreplicates with added noise in the time coordinate to address a major uncertainty in the dataset: For the oiled particle(s) collected in each sediment sample, how much time elapsed between wellhead emission and sample collection? Although many spills are contained far more quickly than the DWH event, no oil spill is without its uncertainties. We can increase our confidence in our analyses of these events and set bounds on the range of possible outcomes by modeling the uncertainties explicitly.

Hopane Is Not Always Conserved. For more than 20 y, hopane has been widely used as a conservative internal standard (36, 37) for quantification of oil weathering after spills. Indeed, we have previously treated hopane as conservative and have used the seafloor hopane anomaly as a basis to calculate the corresponding contamination burden as ∼4–31% of the oil from the deep plume (28). The analysis we present here supports the conclusion that hopane does not behave uniformly as a conservative biomarker in Macondo oil deposited to the seafloor but rather undergoes significant biodegradation at low and moderate contamination levels. Two-thirds of the surficial samples identified by the MDI fall into the low-contamination class, and for these samples only 39% of hopane remained at 4 y postexplosion. An additional 19% of samples fall into the moderate-contamination class; in these samples, 64% of hopane remained after 4 y. However, hopane is relatively persistent (95% remaining after 4 y) in the highly contaminated samples, supporting its use a conservative marker in heavily contaminated environments. These results add to other studies (39, 40, 49, 50) that redefine views on hopane's fidelity and utility as an internal standard. In light of the research community's crucial public role in assessing the damage wrought by past and future spills, this mounting evidence strongly suggests that best practices are due for revision. Although the use of some internal standard is essential, hopane should not be assumed to be the best choice for timescales of months to years but rather should be assessed for utility on a case-by-case basis.

In this spirit, we have updated our previous hopane-based estimate of seafloor oil contamination from the DWH event (28). Working within the same 3,200-km^2 study area considered in that study (28), we applied the most robust kriging model identified there [empirical Bayesian kriging (EBK) model EBK-C] to (*i*) *n*-C38 concentrations and (*ii*) projections of original hopane concentration. The resulting interpolated deposition footprints are in good agreement, accounting for, respectively, ∼13.7 and <14.7% of the oil from the deep plume. The comparatively small disparity between these estimates and the previous EBK-C estimate of ∼12% (28) likely results from the freshness of samples used in interpolation (collected ≤1.5 y postexplosion) and from the recalcitrance of hopane in heavily contaminated samples.

Multiple Factors Control Biodegradation of Macondo Oil. The contamination-level bins used in the present work were chosen empirically, based on exploratory analysis of the fraction-remaining data. It is noteworthy, then, that these bins correspond neatly to the level of contamination expected from the different particle-size classes suggested by previous modeling work (28). Contamination with <400 ng/g hopane is consistent with the deposition of a single oiled particle from the smallest predicted size class (∼0.024 g oil); contamination with ≥750 ng/g hopane is consistent with the deposition of more than one particle of the larger classes. We observed more scatter in the fraction-remaining data in the 400–750 ng/g

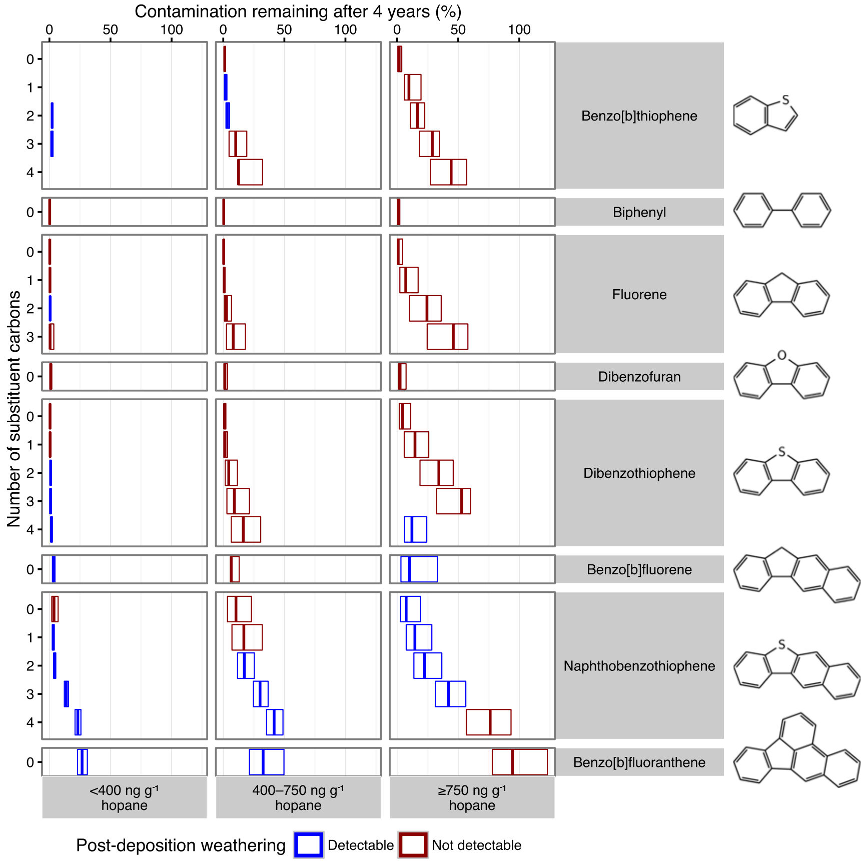

Fig. 5. Percent of aromatic compounds with five-membered rings remaining 4 y postexplosion. (Biphenyl is included here because of its structural resemblance to fluorene, dibenzofuran, and dibenzothiophene.) Panels are ordered by increasing carbon skeleton size from top to bottom and within each panel by increasing number of carbon substituents. Compounds for which biodegradation was detectable after deposition are shown in blue, with crossbars indicating the fitted value and boxes the 95% CI of the median fit result. Compounds for which postdeposition biodegradation was not detectable are shown in red, with vertical bars indicating the median and boxes the interquartile range of measured values.

hopane bin than in either the <400 ng/g or the ≥750 ng/g hopane bins; this moderate-contamination level can arise either from two or three particles of the smallest size class or from one of the particles in the low tail of the second size class. The observed kinetic heterogeneity therefore might reflect a mixture of light- and heavy-contamination–like behaviors across different samples. Alternatively, particles in this concentration range may represent two populations deposited on the seafloor with different histories.

The effect of contamination level on the biodegradation rate reported here is consistent with reports from other environmental settings [e.g., the boulder armoring that protected oil from biodegradation following the *Exxon Valdez* disaster (51), beach sands (52), and bioremediation studies as reviewed previously (53)]. Notably, however, previous examples of this phenomenon have all involved larger spatial scales and higher concentrations. The contamination effect we observe suggests that a similar phenomenon also operates on the approximately millimeter scale and within oil volumes of ~0.01–1 mL (28).

Contra Hazen et al. (12), and consistent with independent metatranscriptomic evidence (54), we find clear evidence for the expected relationship between chemical size and complexity and biodegradation rate. This relationship is clearest in the aliphatic and aromatic compounds analyzed and is most obscure among the biomarkers. The observed rates of diasterane biodegradation are particularly variable, consistent with previous observations in salt marshes (46). This variability is also consistent with previous observations (45, 47), and the observed concentration dependence provides a framework for interpreting such variable sterane deficits.

The robust distinction between the two phases of loss for samples with low and moderate contamination suggests that controls on weathering differed before and after deposition. We hypothesize that this effect arises from a relatively rapid microbial response to freshly suspended oil droplets followed by a marked reduction in microbial metabolism after droplets aggregated and settled to the sea floor, where biodegradation might be limited by insufficient access to a terminal oxidant or nutrients. Among highly contaminated samples, predeposition biodegradation could be limited either by the faster deposition of larger particles, limiting their exposure to in-plume weathering conditions, or by larger particles' low surface area:volume ratio, limiting bioavailability. In the latter case, particles might spread upon deposition, allowing biodegradation to proceed.

Two limitations of the biphasic kinetic model should be emphasized. First, the distribution of deposition times, i.e., of breakpoints between phases, is not known. We chose to fix the breakpoint uniformly at t = 160 d postexplosion because that represents the earliest date from which we have Macondo-contaminated sediment samples. The modal breakpoint could be earlier, and, as noted above, could differ for different particle-size classes. Second, the first phase of degradation is characterized only by its modeled endpoints, i.e., source oil and the earliest sediment samples. Complex kinetics could lurk in the first phase; we can make claims only about the total extent of biodegradation before seafloor deposition, not about the time-dependence of biodegradation in this window.

Stout and Payne (25) have recently argued that the predominant signals in DWH sediment chemistry data are distance-related: The farther from the well an oily particle was deposited, the greater is the extent of biodegradation. They hypothesize that biodegradation proceeds more rapidly in suspension than after sedimentation, so that hydrocarbons in oily particles that were carried further (and thus remained in suspension longer) are systematically more degraded than those in particles deposited closer to the wellhead. This model is consistent with our finding that the majority of pseudoreplicates are best fit by a two-phase biodegradation model, with faster degradation before deposition.

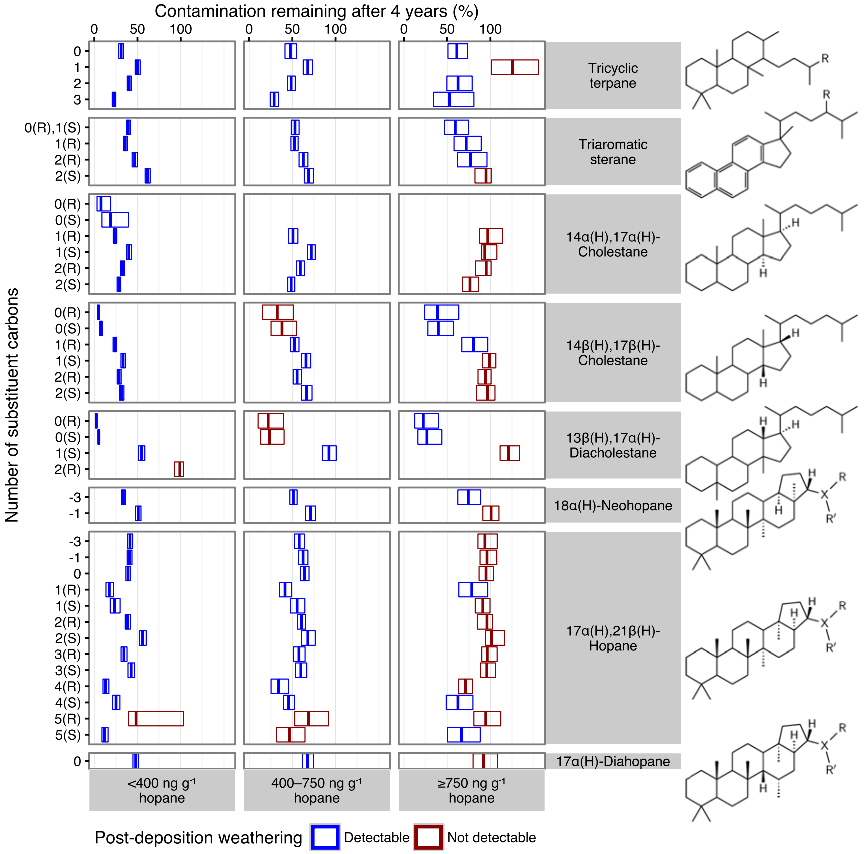

Fig. 6. Percent of biomarker compounds remaining 4 y postexplosion. Panels are ordered by increasing carbon skeleton size from top to bottom and within each panel by increasing number of carbon substituents, with (R) and (S) substituent stereochemistry displayed separately. Among neohopanes and hopanes, −3 indicates the tris-nor compounds and −1 indicates the nor compounds; 1(R) and 1(S) through 5(R) and 5(S) indicate homohopanes through pentakishomohopanes. Compounds for which biodegradation was detectable after deposition are shown in blue, with crossbars indicating the fitted value and boxes indicating the 95% CI of the median fit result. Compounds for which postdeposition biodegradation was not detectable are shown in red, with vertical bars indicating the median and boxes indicating the interquartile range of measured values.

In light of the variability of currents in the region (29) and the unknown deposition history of oiled particles in different samples, our analysis does not account explicitly for distance from the wellhead; the breakpoint between the first and second phase of biodegradation is treated as constant for all samples. Thus, if a distance signal exists, it should be detectable in the residuals of our fits: Macondo-contaminated samples collected farther from the wellhead should have systematically more negative residuals (i.e., greater degradation than the model predicts, occurring during the longer-than-average time to deposition) than Macondo-contaminated samples collected closer to the wellhead. To check, we examined the relationship between model residuals and distance from the wellhead, limiting the analysis to pseudoreplicates with at least moderate support (ΔBIC ≥2) for the best-fit model.

We found a significant negative slope in all contamination bins for aliphatic and aromatic compounds. The more recalcitrant biomarkers show a smaller negative slope at low contamination levels, a negligible negative slope at moderate contamination levels, and a negligible positive slope at high contamination levels (*SI Appendix*, Fig. S14). This result is consistent with our observation that two-phase kinetics dominate for aliphatic and aromatic compounds but not for biomarkers: The length of time spent in suspension should matter more when the difference between pre- and postdeposition rates is larger. An additional nuance is that the distance effect is weakest whenever predeposition biodegradation is either very fast or very slow. This diminution in distance effect could reflect a Goldilocks effect: For labile compounds in small particles, even those deposited closest to the well remained in suspension for multiple degradation half-lives, whereas for recalcitrant compounds in large particles, even those deposited farthest were in suspension for only one or two. Both cases should damp the distance signal.

Implications for Response Efforts in Future Deep Spills. Future spill-response efforts may be informed by two key findings from the present work: first, that biodegradation is much faster in suspension than after deposition for most compounds studied; and second, that contamination level is a key control on degradation rates. These findings suggest that the lasting benthic impact of deep-sea spills may be minimized by measures that drive oil to stay suspended in smaller droplets for longer—an intended mode of action of the 2.9×10^6 L of chemical dispersant applied directly at the wellhead during the spill. While it is not known to what extent dispersant drove oil into microdroplets that biodegraded while remaining suspended in the ocean's interior, the identification of the dispersant's active ingredient in the deep ocean intrusion layers (6) and in benthic oil deposits (4) suggests that the dispersant did remain in suspension with the oil. Furthermore, roller-tank experiments with Macondo oil (19) demonstrated that dispersant delayed the formation of marine snow, perhaps through a direct influence on the surface-layer properties of oil particles or an effect on the microbial release of aggregation-promoting exudates. Together with the present findings, these observations suggest that subsea dispersant application contributes to a net acceleration of biodegradation.

However, other lines of evidence cloud this picture. A recent study questions the efficacy of the subsea dispersant application in modulating droplet size (55), and the variable impacts of dispersant on biodegradation are at the center of an ongoing debate (56, 57). The effect of dispersant itself on the benthos is not well understood, but components of dispersant have been found to persist in sediments and in fragile deep-sea coral communities on a scale of years (4). Decisions regarding the use of dispersant in future spills thus need to weigh not only endpoint hydrocarbon concentration but also context: The net environmental impact of years-long exposure to high local concentrations of undispersed sediment-bound oil may or may not be more severe than the combined effects of short-term exposure of the deep water column to microdroplets of oil and dispersant and long-term exposure of sediments to their residues.

Conclusions

Our compound-specific analysis confirms expected chemical structure trends in biodegradation rates but holds several surprises: First, that biomarkers, including hopane, are subject to substantial biodegradation after deposition; second, that biodegradation patterns differ markedly depending on the extent of contamination; and third, that biodegradation was typically much faster in the short window while oil particles remained suspended than it was subsequently on the deep seafloor. These results provide a basis for predicting the ongoing biodegradation of Macondo oil on the floor of the Gulf of Mexico, inform the ongoing debate about the merits of subsea dispersant use, and argue for caution when using hopane as an internal standard for oil-spill research at long timescales.

Methods

Data used in this work are freely available from the National Oceanic and Atmospheric Administration (NOAA), as part of the NRDA of the DWH event. Data were downloaded from the NRDA data site, now at https://dwhdiver.orr.noaa.gov/explore-the-data/, and included all chemistry data in the sediment category through 8 May 2015 and all chemistry data in the subsurface sediment, oil, and water categories as of 2 May 2014. All data used in this study are included in the research compendium deposited with Figshare (https://figshare.com), DOI 10.6084/m9.figshare.4001262. Bathymetric data were downloaded 11 June 2014 from the NOAA GEODAS server (www.ngdc.noaa.gov/mgg/gdas/gd_designagrid.html); xyz values are from the ETOPO1 dataset (https://www.ngdc.noaa.gov/mgg/global/global.html) at 1-min resolution.

Analysis was performed in R (v. 3.3.1), using the packages broom (v. 0.4.1), e1071 (v. 1.6-7), NADA (v. 1.5-6), survival (v. 2.39-5), geosphere (v. 1.5-5), sp (v. 1.2-3), lubridate (v. 1.5.6), assertr (v. 1.0.2), reshape2 (v. 1.4.1), dplyr (v. 0.5.0), tidyr (v. 0.6.0), plyr (v. 1.8.4), and marmap (v. 0.9.5). Figures were produced in R with packages png (v. 0.1-7), Cairo (v. 1.5-9), extrafont (v. 0.17), gridExtra (v. 2.0.1), hexbin (v. 1.27.1), cowplot (v. 0.6.2), scales (v. 0.4.0), and ggplot2 (v. 2.1.0). All code used in the analysis is deposited with the Figshare compendium (DOI 10.6084/m9.figshare.4001262).

Data Reduction. We filtered the dataset to remove all stations with seafloor depth of less than 200 m, all samples identified as "Filter/Particulate," and all bottom-water samples. See *SI Appendix*, *SI Text* for a complete description of sample curation. We included only analyses with reported quality codes of Not Available, J, and U. Analyses with quality code U were censored as the interval [0, detection limit (dl)], even where the reported concentration exceeded the dl. For analyses with quality code J, concentrations were censored as the interval (0, dl) in the small number of cases (n = 149 in the full dataset, n = 2 remaining in the final dataset) where the reported concentration fell below the reported dl and were uncensored otherwise. Analyses with quality code Not Available were uniformly treated as quantitative, i.e., uncensored.

Some samples had been measured repeatedly at different calibration scales, often yielding a nondetect at high dl and a quantitative result at low dl. In these cases, the nondetects were excluded from further analysis. In some cases, a single physical sample was split for multiple measurements of an analyte using different analytical methods, with one method marked by Alpha Analytical as disfavored and flagged with an "X" suffix in the labrep. We discarded these analyses (*SI Appendix*, *SI Text*). Methods that the NRDA later designated as disfavored were the only methods used to analyze the 40 biomarker compounds in 256 samples (n = 13 in the final dataset) collected within 8 mo of the explosion. Because few or no other data were available for these compounds in this period (*SI Appendix*, Fig. S15), we included these measurements in our analysis. Comparison fits excluding these measurements indicated that these data points did not skew the results (*SI Appendix*, *SI Text* and Figs. S16 and S17).

Identification of Macondo Oil by the MDI. We assessed 83 biomarker ratios (45, 58) for inclusion in the MDI score, choosing 12 based on their data completeness, discriminating ability, and chemical diversity (*SI Appendix*, Table S2). We defined a set of reference samples (*SI Appendix*, Table S3) consisting of (*i*) the 11 samples (pooling physical splits) from study name Chem–Source Oil 2010 and (*ii*) the 12 samples identified in the NRDA Sediment and Subsurface datasets at <1 km from the wellhead, upper depth 0 cm, and hopane ≥750 ng/g (i.e., ≥10-fold above background), collected <1 y after the spill midpoint. This composite source was chosen to account for potential variability in the measured biomarker ratios of Macondo oil such as might be caused by changes in fluid composition during discharge, matrix effects impacting extraction from sediment, and analytical uncertainty. We defined a per-ratio dissimilarity metric (*SI Appendix*, *SI Text*, Eq. S1) and calculated the MDI as the average penalty over n informative (noncensored) ratios (*SI Appendix*, *SI Text*, Eq. S2). We excluded samples with n <8 from further analysis. See *SI Appendix*, *SI Text* and Fig. S1 for details of the choice of the cutoff value MDI <1.8.

Data Normalization and Contamination Binning. To control for differences in absolute oil mass, we analyzed biodegradation kinetics in terms of the concentration of each compound remaining in a sample relative to the concentration of an internal reference compound and normalized that ratio to the corresponding ratio in Macondo Well (MW) source oil (*SI Appendix*, *SI Text*). The result is the fraction remaining: for an analyte a, $F_{a,samp} = (C_{a,samp}/C_{ref,samp})/(C_{a,MW}/C_{ref,MW})$. Samples for which the measurement of the reference compound was censored were excluded from further analysis.

As reference compounds for normalization, we considered both the biomarker hopane, expected to be recalcitrant, and n-C38, the longest detectable aliphatic compound. We calculated the pull for each pseudoreplicate as the best-fit slope divided by the fitted error on the slope and examined the positive arm of the pull distributions for each contamination level and compound class (*SI Appendix*, *SI Text* and Fig. S4). Pull SDs were calculated as the rms of positive pull values (*SI Appendix*, *SI Text*). After examination of the skewed pull distributions (*SI Appendix*, Fig. S4) resulting from fitting hopane-normalized data, we proceeded with analysis of n-C38–normalized data instead. See *SI Appendix*, *SI Text* for complete description of data normalization.

Visual inspection of fraction-remaining time-course plots suggested a systematic influence of contamination level on fraction remaining. We used the two-sample Kolmogorov–Smirnov test to compare empirical cumulative distribution functions across contamination bins (low, <400 ng/g hopane; moderate, 400–750 ng/g hopane; high, ≥750 ng/g hopane), finding significant differences ($P < 0.05$) for the large majority of compounds. We applied these contamination bins to all further analysis.

Construction and Head-To-Head Comparison of Fits. We attempted to fit all 353 compound–contamination bin datasets that include ≥10 uncensored data points and ≤80% censored observations. Three low-contamination datasets [13β(H),17α(H)-20R-ethyldiacholestane and the 20R and 20S enantiomers of 14α(H),17α(H)-cholestane] spanned a time range of <100 d in sediment (samples collected at 160–239 d postexplosion). All other datasets spanned >200 d in sediment, with 337 of the 353 datasets spanning ∼3.5 y. Because the dataset includes nondetects, we fit using maximum likelihood-based survival analysis (R package survival), assuming a Gaussian distribution for the log-transformed response variable.

For each dataset, we fit 100 pseudoreplicates with uniformly distributed noise added to the time axis to reflect the 87-d uncertainty in time from emission to sampling. We used the BIC to perform head-to-head comparisons of three models on each pseudoreplicate's log-transformed fraction-remaining data: a single-phase linear model with the y intercept permitted to vary freely; a single-phase linear model with the y intercept fixed at 0; and a piecewise-linear (broken-stick) biphasic model with breakpoint fixed at 160 d after the explosion. For the biphasic model, the first phase is constrained only by its endpoints, i.e., source oil at $t = 0$ and the earliest sediment samples at $t = 160$ d. For each pseudoreplicate, we identified the best-fit model as the model that minimized the BIC. See *SI Appendix*, *SI Text* for details of extracting predictions from pseudoreplicate ensembles.

EBK Estimation of Seafloor Oil Contamination Burden. To revise our estimate of the seafloor contamination burden, we repeated the EBK process used in ref. 28 with the most robust model from that analysis, there designated EBK-C. To facilitate comparison, we used the same set of 534 samples here as in our previous work, with the exception that 14 of the samples had no concentration data available for n-C38. The exclusion of these points from the original hopane analysis had no net effect on the previously published contamination burden. To calculate the excess n-C38 in the footprint area, we used a background concentration of 82 ng/g n-C38, calculated as the mean n-C38 concentration in surficial sediment samples that pass our quality filters, have an MDI >1.8, and were collected ≥40 km from the wellhead. For the EBK-C run using projected hopane concentrations in unweathered oil, we assumed the maximum possible age for the oil at the time of sampling (i.e., that all oil collected had emerged from the wellhead on the day of the explosion). For each sample that matched the Macondo fingerprint, we used the median best-fit model from the appropriate contamination bin to calculate the mass of hopane that would have been present in that sample absent biodegradation. Hopane concentrations from samples with an MDI ≥1.8 were held constant. As in ref. 28, we calculated the excess hopane burden using a background concentration of 28 ng/g hopane.

ACKNOWLEDGMENTS. We thank Elizabeth Porta (Alpha Analytical) for useful advice, Deborah French-McCay (Applied Science Associates) for preparation and coordination of NRDA cruises, and the crew and scientists of the 18 NRDA cruises whose sediment samples we analyzed. This work was supported, in part, by National Science Foundation Grants OCE-1333162, OCE-0961725, EAR-0950600, and OCE-1046144 (to D.L.V.) and OCE-1333148 (to C.M.R.) and by Simons Foundation Grant 385324 (to D.L.V.).

1. McNutt MK, et al. (2012) Review of flow rate estimates of the Deepwater Horizon oil spill. *Proc Natl Acad Sci USA* 109(50):20260–20267.
2. Reddy CM, et al. (2012) Composition and fate of gas and oil released to the water column during the Deepwater Horizon oil spill. *Proc Natl Acad Sci USA* 109(50): 20229–20234.
3. Camilli R, et al. (2012) Acoustic measurement of the Deepwater Horizon Macondo well flow rate. *Proc Natl Acad Sci USA* 109(50):20235–20239.
4. White HK, et al. (2014) Long-term persistence of dispersants following the Deepwater Horizon oil spill. *Environ Sci Technol Lett* 1(7):295–299.
5. Socolofsky SA, et al. (2015) Intercomparison of oil spill prediction models for accidental blowout scenarios with and without subsea chemical dispersant injection. *Mar Pollut Bull* 96(1-2):110–126.
6. Kujawinski EB, et al. (2011) Fate of dispersants associated with the Deepwater Horizon oil spill. *Environ Sci Technol* 45(4):1298–1306.
7. Ryerson TB, et al. (2012) Chemical data quantify Deepwater Horizon hydrocarbon flow rate and environmental distribution. *Proc Natl Acad Sci USA* 109(50): 20246–20253.
8. Passow U, Ziervogel K, Asper V, Diercks A (2012) Marine snow formation in the aftermath of the Deepwater Horizon spill in the Gulf of Mexico. *Environ Res Lett* 7(3): 035301.
9. Daly KL, Passow U, Chanton J, Hollander D (2016) Assessing the impacts of oil-associated marine snow formation and sedimentation during and after the Deepwater Horizon oil spill. *Anthropocene* 13:18–33.
10. Aeppli C, et al. (2012) Oil weathering after the Deepwater Horizon disaster led to the formation of oxygenated residues. *Environ Sci Technol* 46(16):8799–8807.
11. Diercks A-R, et al. (2010) Characterization of subsurface polycyclic aromatic hydrocarbons at the Deepwater Horizon site. *Geophys Res Lett* 37(20):L20602.
12. Hazen TC, et al. (2010) Deep-sea oil plume enriches indigenous oil-degrading bacteria. *Science* 330(6001):204–208.
13. Joye SB, MacDonald IR, Leifer I, Asper V (2011) Magnitude and oxidation potential of hydrocarbon gases released from the BP oil well blowout. *Nat Geosci* 4(3):160–164.
14. Kessler JD, et al. (2011) A persistent oxygen anomaly reveals the fate of spilled methane in the deep Gulf of Mexico. *Science* 331(6015):312–315.
15. Spier C, Stringfellow WT, Hazen TC, Conrad M (2013) Distribution of hydrocarbons released during the 2010 MC252 oil spill in deep offshore waters. *Environ Pollut* 173: 224–230.
16. Valentine DL, et al. (2010) Propane respiration jump-starts microbial response to a deep oil spill. *Science* 330(6001):208–211.
17. Lindo-Atichati D, et al. (2016) Simulating the effects of droplet size, high-pressure biodegradation, and variable flow rate on the subsea evolution of deep plumes from the Macondo blowout. *Deep Sea Res Part II Top Stud Oceanogr* 129:301–310.
18. North EW, et al. (2015) The influence of droplet size and biodegradation on the transport of subsurface oil droplets during the Deepwater Horizon spill: A model sensitivity study. *Environ Res Lett* 10(2):024016.
19. Passow U (2014) Formation of rapidly-sinking, oil-associated marine snow. *Deep-Sea Res PT II* 129:232–240.
20. Stoffyn-Egli P, Lee K (2002) Formation and characterization of oil–mineral aggregates. *Spill Sci Technol Bull* 8(1):31–44.
21. Montagna PA, et al. (2013) Deep-sea benthic footprint of the Deepwater Horizon blowout. *PLoS One* 8(8):e70540.
22. Liu Z, Liu J, Zhu Q, Wu W (2012) The weathering of oil after the Deepwater Horizon oil spill: Insights from the chemical composition of the oil from the sea surface, salt marshes and sediments. *Environ Res Lett* 7(3):035302.
23. Liu Y, MacFadyen A, Ji Z-G, Weisberg RH, eds (2011) *Monitoring and Modeling the Deepwater Horizon Oil Spill: A Record-Breaking Enterprise* (American Geophysical Union, Washington, DC), 10.1029/GM195.
24. Kimes NE, et al. (2013) Metagenomic analysis and metabolite profiling of deep-sea sediments from the Gulf of Mexico following the Deepwater Horizon oil spill. *Front Microbiol* 4:50.
25. Stout SA, Payne JR (2016) Macondo oil in deep-sea sediments: Part 1 - sub-sea weathering of oil deposited on the seafloor. *Mar Pollut Bull* 111(1-2):365–380.
26. Stout SA, Payne JR, Ricker RW, Baker G, Lewis C (2016) Macondo oil in deep-sea sediments: Part 2 - Distribution and distinction from background and natural oil seeps. *Mar Pollut Bull* 111(1-2):381–401.
27. Chanton J, et al. (2015) Using natural abundance radiocarbon to trace the flux of petrocarbon to the seafloor following the Deepwater Horizon oil spill. *Environ Sci Technol* 49(2):847–854.
28. Valentine DL, et al. (2014) Fallout plume of submerged oil from Deepwater Horizon. *Proc Natl Acad Sci USA* 111(45):15906–15911.
29. Valentine DL, et al. (2012) Dynamic autoinoculation and the microbial ecology of a deep water hydrocarbon irruption. *Proc Natl Acad Sci USA* 109(50):20286–20291.
30. Redmond MC, Valentine DL (2012) Natural gas and temperature structured a microbial community response to the Deepwater Horizon oil spill. *Proc Natl Acad Sci USA* 109(50):20292–20297.
31. Dubinsky EA, et al. (2013) Succession of hydrocarbon-degrading bacteria in the aftermath of the deepwater horizon oil spill in the gulf of Mexico. *Environ Sci Technol* 47(19):10860–10867.
32. Gros J, et al. (2014) Resolving biodegradation patterns of persistent saturated hydrocarbons in weathered oil samples from the Deepwater Horizon disaster. *Environ Sci Technol* 48(3):1628–1637.
33. Slater GF, White HK, Eglinton TI, Reddy CM (2005) Determination of microbial carbon sources in petroleum contaminated sediments using molecular ^{14}C analysis. *Environ Sci Technol* 39(8):2552–2558.
34. Leahy JG, Colwell RR (1990) Microbial degradation of hydrocarbons in the environment. *Microbiol Rev* 54(3):305–315.
35. Grossi V, Massias D, Stora G, Bertrand J-C (2002) Burial, exportation and degradation of acyclic petroleum hydrocarbons following a simulated oil spill in bioturbated Mediterranean coastal sediments. *Chemosphere* 48(9):947–954.
36. Prince RC, et al. (1994) 17.α.(H)-21.β.(H)-hopane as a conserved internal marker for estimating the biodegradation of crude oil. *Environ Sci Technol* 28(1):142–145.
37. Venosa AD, Suidan MT, King D, Wrenn BA (1997) Use of hopane as a conservative biomarker for monitoring the bioremediation effectiveness of crude oil contaminating a sandy beach. *J Ind Microbiol Biotechnol* 18(2–3):131–139.
38. Demortier L, Lyons L (2002) Everything you always wanted to know about pulls. CDF Note 5776, version 2.10. Available at inspirehep.net/record/1354911/files/. Accessed June 14, 2016.
39. Douglas GS, Hardenstine JH, Liu B, Uhler AD (2012) Laboratory and field verification of a method to estimate the extent of petroleum biodegradation in soil. *Environ Sci Technol* 46(15):8279–8287.
40. Douglas GS, Owens EH, Hardenstine J, Prince RC (2002) The OSSA II pipeline oil spill: The character and weathering of the spilled oil. *Spill Sci Technol Bull* 7(3):135–148.
41. Heitkamp MA, Cerniglia CE (1987) Effects of chemical structure and exposure on the microbial degradation of polycyclic aromatic hydrocarbons in freshwater and estuarine ecosystems. *Environ Toxicol Chem* 6(7):535–546.
42. Wammer KH, Peters CA (2005) Polycyclic aromatic hydrocarbon biodegradation rates: A structure-based study. *Environ Sci Technol* 39(8):2571–2578.
43. Bossert ID, Bartha R (1986) Structure-biodegradability relationships of polycyclic aromatic hydrocarbons in soil. *Bull Environ Contam Toxicol* 37(4):490–495.
44. Wardlaw GD, Nelson RK, Reddy CM, Valentine DL (2011) Biodegradation preference for isomers of alkylated naphthalenes and benzothiophenes in marine sediment contaminated with crude oil. *Org Geochem* 42(6):630–639.
45. Aeppli C, et al. (2014) Recalcitrance and degradation of petroleum biomarkers upon abiotic and biotic natural weathering of Deepwater Horizon oil. *Environ Sci Technol* 48(12):6726–6734.
46. Overton EB, Miles MS, Meyer BM, Gao H, Turner RE (2014) Oil source fingerprinting in heavily weathered residues and coastal marsh samples. Proceedings of the International Oil Spill Conference 2014(1):2074–2082.
47. White HK, et al. (2012) Impact of the Deepwater Horizon oil spill on a deep-water coral community in the Gulf of Mexico. *Proc Natl Acad Sci USA* 109(50):20303–20308.
48. Stout S (2016) Oil spill fingerprinting method for oily matrices used in the Deepwater Horizon NRDA. *Environ Forensics* 17(3):218–243.
49. Frontera-Suau R, Bost FD, McDonald TJ, Morris PJ (2002) Aerobic biodegradation of hopanes and other biomarkers by crude oil-degrading enrichment cultures. *Environ Sci Technol* 36(21):4585–4592.
50. Carls MG, Holland L, Irvine GV, Mann DH, Lindeberg M (2016) Petroleum biomarkers as tracers of Exxon Valdez oil. *Environ Toxicol Chem* 35(11):2683–2690.
51. Short JW, et al. (2007) Slightly weathered Exxon Valdez oil persists in Gulf of Alaska beach sediments after 16 years. *Environ Sci Technol* 41(4):1245–1250.
52. Del'Arco JP, de França FP (2001) Influence of oil contamination levels on hydrocarbon biodegradation in sandy sediment. *Environ Pollut* 112(3):515–519.
53. Swannell RP, Lee K, McDonagh M (1996) Field evaluations of marine oil spill bioremediation. *Microbiol Rev* 60(2):342–365.
54. Mason OU, et al. (2012) Metagenome, metatranscriptome and single-cell sequencing reveal microbial response to Deepwater Horizon oil spill. *ISME J* 6(9):1715–1727.
55. Aman ZM, Paris CB, May EF, Johns ML, Lindo-Atichati D (2015) High-pressure visual experimental studies of oil-in-water dispersion droplet size. *Chem Eng Sci* 127: 392–400.
56. Kleindienst S, et al. (2015) Chemical dispersants can suppress the activity of natural oil-degrading microorganisms. *Proc Natl Acad Sci USA* 112(48):14900–14905.
57. Prince RC, Coolbaugh TS, Parkerton TF (2016) Oil dispersants do facilitate biodegradation of spilled oil. *Proc Natl Acad Sci USA* 113(11):E1421–E1421.
58. Wang Z, et al. (2007) *Oil Spill Environmental Forensics*, eds Wang Z, Stout SA (Elsevier, Burlington, MA), pp 73–146.

SI Appendix for "Persistence and biodegradation of oil at the ocean floor following *Deepwater Horizon*"

SC Bagby, CM Reddy, C Aeppli, GB Fisher, and DL Valentine

1 SI Text

1.1 Data source

Oil, sediment, and subsurface sediment chemistry datasets were downloaded 2014-05-29; the water chemistry dataset was downloaded 2014-06-04. According to the NRDA download site (`http://www.gulfspillrestoration.noaa.gov/oil-spill/gulf-spill-data/`), the most recent previous update to each dataset was 2014-05-02. The NRDA website was updated with additional data on 2015-05-08. The updated surface sediment data was downloaded 2015-07-16 and added to the analysis pipeline.

In September 2015, the NRDA download site was redesigned. Raw data downloads may now be obtained via the DIVER interface at `https://dwhdiver.orr.noaa.gov/explore-the-data`; choose Download Data > Samples > Chemistry Exports to see the most recent versions of the water, oil, and sediment datasets NRDA has made available. The versions fed into our analysis, after preliminary cleaning (see README) but before quality filtering, are included in the research compendium in the `data/` directory.

Bathymetric data were downloaded 2014-06-11 from the NOAA GEODAS server, `http://www.ngdc.noaa.gov/mgg/gdas/gd_designagrid.html`. xyz values are from the ETOPO1 dataset at 1-minute resolution. For details of custom grid generation, see `data/geodas_GOMhopan-8104/help.htm` and `data/geodas_GOMhopan-8104/readme.txt` in the accompanying research compendium (DOI: 10.6084/m9.figshare.4001262).

1.2 Data filtering

1.2.1 Data reduction and metadata calculation

To minimize heterogeneity in the dataset, we filter the dataset to include only measurements made by Alpha Analytical (Mansfield, MA). In addition, we remove any measurements for which latitude and/or longitude are not reported. For analysis of sediment data, we exclude bottom water/coretop floc samples, as well as samples collected onto filters.

For the remaining records we calculate the following:

- Depth at sampling site (per NOAA's ETOPO1 dataset, at 1-minute resolution)
- Distance to the wellhead, in km (using the 'Meeus' great circle distance calculation)

- Number of days between the *Deepwater Horizon* explosion (2010-04-20) and sampling
- Number of days between the midpoint of the spill (2010-06-02) and sampling
- Number of days between the well closure (2010-07-16) and sampling

We filter out samples collected at sites shallower than 200 m water depth.

1.2.2 Data quality and binning

We filter out all values with qualcodes other than "U" or "U,NSR", "J" or "J,NSR", and "Not Available" or "NSR". The "U" flag indicates a nondetect, and we take this qualcode flag as more authoritative than the reported concentration. Thus, for all U values, *even if* the reported concentration is above the detection limit dl, we treat the data as censored to a concentration interval of $[intL, intR] = [0, dl]$. For measurements with "J" or "Not Available" qualcode flags, we define concentration intervals as

$$[intL, intR] = \begin{cases} [0, dl] & : C < dl \\ [C, C] & : C \geq dl \end{cases}$$

Groups of replicate analyses may include any combination of quantitative measurements and nondetects. In a number of instances, samples were analyzed repeatedly at different calibration scales, such that the results from a single physical sample include, e.g., (1) labrep 1, a nondetect with reporting limit $rl \approx 80000$ and concentration $conc < rl$ and (2) labrep 2, a quantitative measurement with $rl \approx 8000$ and $conc > rl$. In these cases, the high-calibration-range nondetects should *not* be included in averaging and further data processing; that uncertainty is not informative when the same sample, analyzed at a more sensitive scale, yielded a quantitative measurement. We filter out all such nondetects. By contrast, when a group of replicates includes nondetects obtained at or below the highest calibration range that yielded a quantitative measurement, those nondetects are informative, and are not discarded. Likewise, if a group of replicates includes no quantitative measurements, that sample represents a genuine nondetect, and as such should not be filtered out; in such cases we keep the lowest-rl measurement only, as the measurement that constrains the underlying concentration to the narrowest interval.

In some cases multiple measurements were made using different analytical methods, with one method marked by Alpha as disfavored and flagged with an "X" suffix in the labrep field. Following the above filtering step, we ask whether any groups of measurements of the same physical sample were obtained using more than one method; for any that are, we discard results derived from the disfavored method. In the resulting dataset, each analyte is measured by only one method per sample, but in some cases more than one method was used within the complete dataset for an analyte. See "Assessing fits", below, for discussion of the limited impact of these mixed-method datasets on our analysis.

1.3 Identification of Macondo oil

To move forward with analysis of biodegradation kinetics, we needed to define a set of samples we can classify as contaminated by Macondo oil. This sample identification should be conservative, so that (insofar as possible) we are studying biodegradation in oil released from the Macondo well during a known time period; we would rather weaken our statistics by cutting the size of the dataset than skew the fits by including extraneous samples.

We used biomarker concentration ratios as a basis for identifying Macondo-contaminated samples. Previous work (*1*, *2*) has identified a large number of potentially informative biomarker ratios. We calculated per-sample values of 83 such ratios (see this manuscript's research compendium for the complete list) and used exploratory data visualization (not shown) to assess their potential utility in the Macondo dissimilarity index, choosing 12 that encompass a broad chemical range and were well represented in the dataset.

We compared the ratio values obtained in candidate samples to those found in a reference set of samples consisting of

- Chem–Source Oil 2010 (11 samples)
- All those samples from the Sediment dataset that were collected <1 km from the wellhead, <1 year after the spill midpoint, with upper depth = 0 cm *and* hopane ≥ 750 ng g^{-1} (12 samples)

The latter group—close, surficial, highly contaminated samples—-are those that have the highest a priori likelihood being contaminated with Macondo oil. Although it might be argued that they raise the risk of circular reasoning, we included them to counteract the possibility that extraction efficiencies will differ systematically between the source oil sample and the sediment samples. We choose 750 ng g^{-1} hopane as the lower threshold for this reference set because it is an order of magnitude above the high cut-off value we have previously established for background hopane concentrations in the region of the wellhead (*3*).

We defined a dissimilarity function to use in the fingerprint:

$$d(r_{ij}) = \begin{cases} \min(\frac{\log(r_{ij}/m_i)}{\log(l_i/m_i)}, 10) & : r_{ij} < l_i \\ 0 & : l_i \le r_{ij} \le u_i \\ \min(\frac{\log(r_{ij}/m_i)}{\log(u_i/m_i)}, 10) & : r_{ij} > u_i \end{cases} \tag{1}$$

where r_{ij} is the value of ratio i ($1 \le i \le 12$) in sample j, l_i is the second-lowest observed value of ratio i in the reference set, m_i is the median observed value of ratio i in the reference set, and u_i is the second-highest observed value of ratio i in the reference set. For each sample j, we counted the number n of informative (non-censored) ratios, calculated $d(r_{ij})$ for all i, sum over i, and divided by the number of informative ratios to get an average dissimilarity:

$$D_j = \frac{1}{n}\sum_{i=1}^{n} d(r_{ij}) \tag{2}$$

(For samples with no censored data, $n = 12$; for samples with censored data, $n < 12$.) Note that, because this is a dissimilarity metric, *smaller* values indicate a *closer* match to the reference set.

Low hydrocarbon content in many sediment samples led to measured values below the limit of quantitation, i.e., censored values. While it is essential to include nondetects in the degradation analysis, they are not informative for fingerprinting. To get useful output from this metric, we required samples to have informative (non-censored) ratios for at least 8 of the 12 ratios examined. This excluded a substantial portion of the dataset: out of the 2980 sediment samples in which hydrocarbon content was measured, only 1555 met our threshold of $\geq$ 8 fingerprint ratios with uncensored data.

To define a threshold for identification of Macondo oil under this metric, we first examined the distribution of D_j for sufficiently informative surficial samples (upper depth 0 cm) with <75 ng g^{-1} hopane, i.e., those that are indistinguishable from background on the basis of oil quantity. (We excluded samples collected at $\geq$475 d after the spill midpoint, because the timescale on which weathering might be expected to make Macondo oil in these samples unrecognizable was unknown.) Among these samples, the fourth percentile falls at $D_j \approx 1.8$ (Fig. S1). By contrast, a threshold of 1.8 lies at the 73^{rd} percentile of sufficiently informative surficial samples with $\geq$75 ng g^{-1} hopane (again, excluding late samples from the reckoning), suggesting good separation between the samples we expect a priori to be Macondo oil and those that are comparatively unlikely to be.

We defined the working set of samples contaminated with Macondo oil as those meeting *all* of the following criteria:

- Upper depth in sediment 0 cm
- $\geq$8 informative ratios (of 12 examined)
- $D_j < 1.8$

1.4 Final data-quality trimming

1.4.1 Sample curation

We flag a final handful of samples for removal prior to kinetic analysis:

- Four sediment samples, of which three pass the MDI cutoff, are from the region expected to be contaminated by drilling mud, whose hydrocarbon content could confound analysis (*4*).
- We remove one source oil sample (fldsampid GU2988-A0521-O9805, sampleid O006D) because no hopane measurement is reported.
- As shown in Fig. S18, one sediment sample is a clear outlier in visual comparison of n-C38 concentration data vs. hopane concentration data (fldsampid GU2790-A1023-SE301, sampleid S001). We exclude this sample from further consideration. Note that this sample was collected much farther away, 239 km from the wellhead, than the next most distant sample (51.5 km); it was also much shallower, at a seafloor depth of 828 m, than the next most shallow (1029 m). It should be noted that the very low hopane to n-C38 concentration ratio in this sample is consistent with extensive hopane degradation during longer than usual transport to this distant site, arguing for the general hypothesis of (*5*) but against those authors' contention that hopane is not degraded under these conditions.

1.4.2 Selection of compounds for which to analyze biodegradation

Using the final working set of Macondo-contaminated samples, we then chose which compounds to analyze based on data quality. We checked first for adequate data quality in the source oil measurements. Because sediment data is normalized to oil data, low-quality oil data has a substantial negative impact on analysis of biodegradation in sediments.

Some censoring is permissible in the source oil dataset: so long as <50% of measurements are censored, a median value can be calculated for use in normalization. However, for eight compounds, >50% of the measurements in source oil are censored; we removed these compounds from further analysis.

Chemname	Number of measurements	Number censored
Benzo(a)fluoranthene	10	10
Benzo(j+k)fluoranthene	10	9
Carbazole	12	9
Indeno(1,2,3-c,d)pyrene	12	8
Nonatriacontane (C39)	10	6
Perylene	12	10
Retene	11	11
Tetracontane (C40)	10	9

Three analytes of potential interest were not found in the source oil dataset and were thus excluded from the sediment dataset as well:

Chemname	Class
T22a-Gammacerane/C32-diahopane	Biomarkers
Solids, percent	Other
Total Extractable Hydrocarbons (C9-C44) (silca treated)	Other

Exploratory data analysis suggested that oil-normalized data grew unreliable for the compounds that were least abundant in source oil, with a mean abundance threshold of approximately 0.08 (relative to hopane concentration). We filtered out the 4 aromatic compounds and 13 biomarkers whose mean relative abundance in source oil fell below this empirical threshold.

Chemname	Class	Mean relative abundance
Benzo(g,h,i)perylene	Aromatics	0.036
Dibenzo(a,h)anthracene	Aromatics	0.040
Benzo(a)pyrene	Aromatics	0.050
Fluoranthene	Aromatics	0.070
18A(H)&18B(H)-Oleananes	Biomarkers	0.038
17A(H),21B(H)-25-Norhopane	Biomarkers	0.039
13A,17B-20S-Ethyldiacholestane	Biomarkers	0.041
C28 Tricyclic Terpane-22S	Biomarkers	0.045
C24 Tetracyclic Terpane	Biomarkers	0.047
C30 Tricyclic Terpane-22R	Biomarkers	0.053
C30 Tricyclic Terpane-22S	Biomarkers	0.053
C28 Tricyclic Terpane-22R	Biomarkers	0.059
C29 Tricyclic Terpane-22R	Biomarkers	0.059
C26 Tricyclic Terpane-22R	Biomarkers	0.063
30-Normoretane	Biomarkers	0.070
C29 Tricyclic Terpane-22S	Biomarkers	0.070
17A/B,21B/A 28,30-Bisnorhopane	Biomarkers	0.078

Finally, we assessed data availability and quality in the sediment samples. Two analytes, Octane (C8) and Total Extractable Matter (C9-C44), that were represented in the raw dataset are not represented in the contaminated samples. Exploratory data visualization suggested the exclusion of an additional set of compounds: pyrene (C0), likely complicated by additional source terms from burning; moretane; and all individually resolved substituted PAHs (excluding, e.g., 1-Methylnaphthalene, 2-Methylnaphthalene, etc., in favor of the pooled set of C1-naphthalenes).

Chemname	Class	Number of measurements
1-Methyldibenzothiophene	Aromatics	292
1-Methylnaphthalene	Aromatics	321
1-Methylphenanthrene	Aromatics	289
2,3,5-Trimethylnaphthalene	Aromatics	295
2,6-Dimethylnaphthalene	Aromatics	319
2-Methylanthracene	Aromatics	254
2-Methylnaphthalene	Aromatics	321
2/3-Methyldibenzothiophene	Aromatics	304
2/4-Methylphenanthrene	Aromatics	259
3-Methylphenanthrene	Aromatics	289
4-Methyldibenzothiophene	Aromatics	298
9-Methylphenanthrene	Aromatics	262
Pyrene	Aromatics	321
17b(H),21a(H)-Hopane (Moretane)	Biomarkers	320

A total of 125 analytes remained for further analysis. Note that the decalins are grouped here with the set denominated "Aromatics", as noted in Fig. 4 and Fig. S8; all compounds in the "Aliphatics" set are saturated acyclic hydrocarbons.

Chemtype	Count
Aliphatics	35
Aromatics	50
Biomarkers	40

1.4.3 Description of data quality in the reduced datasets

The minimal extent of censoring in the reduced oil dataset should provide a robust basis for normalization:

Replicates	Censored measurements	Number of cases
1	0	993
1	1	12
2	0	125
4	0	85
6	0	39

Similarly, the extent of censoring in the reduced sediment dataset should not preclude robust analysis of sediment samples:

Replicates	Censored measurements	Number of cases
1	0	27787
1	1	6561
2	0	414
2	1	7

1.5 Normalizing to reference compounds and oil

1.5.1 Initial normalization

For each sample, we calculated median values for the left and right interval bounds for each compound. This aggregation was trivial for the single-replicate samples, and nearly so for the dual-replicate samples.

For kinetic analysis, the sample median for each compound is normalized to the concentration of an internal reference, either hopane or *n*-C38, and then renormalized to the corresponding ratio in source oil. Source oil ratios were aggregated across all oil sample measurements, using the median or the robust regression on order statistics (ROS) estimate of the median to obtain a single point value.

To determine whether to use the median or the ROS estimate of the median value in source oil, we confirmed that the censoring level is below 50% and checked to see if all censored values are strictly less than the median upper bound. If they are, the median is calculated as usual. If any censored value(s) may exceed the median upper bound, we invoke ROS. In practice, only 6 compounds (anthracene, benzo(b)fluorene, pentatriacontane (*n*-C35), hexatriacontane (*n*-C36), heptatriacontane (*n*-C37), octatriacontane (*n*-C38)) had any censored measurements in the final set of source oil samples; of these, ROS was required for three.

Chemname	Number of measurements	Number censored	Median
Heptatriacontane (C37)	11	2	ROS
Hexatriacontane (C36)	11	2	ROS
Octatriacontane (C38)	11	2	ROS
Anthracene	11	4	Simple
Benzo(b)fluorene	15	1	Simple
Pentatriacontane (C35)	11	1	Simple

1.5.2 Assessing support for binning by contamination level

Exploratory visualization of the normalized data over time suggested a difference in biodegradation rates between the most and least heavily contaminated samples. We used the Kolmogorov-Smirnov test to check for significant differences in the distribution of fraction-remaining data across contamination bins, finding significant differences for the large majority of compounds.

Bin 1	Bin 2	Comparisons with $p < 0.05$
<400 ng g^{-1}	400–750 ng g^{-1}	78 of 124
<400 ng g^{-1}	≥750 ng g^{-1}	119 of 124
400–750 ng g^{-1}	≥750 ng g^{-1}	108 of 124

(Note that the results above are for *n*-C38-normalized data prior to the recensoring described below. The KS test results reported in the main text are as run after recensoring.)

1.5.3 Adjusting intR for selected censored compounds

Exploratory visualization of dual-normalized data revealed a large gap between the highest upper bounds of censored data and the lowest quantitative measurements for some analytes. Whereas, for other compounds, we typically observed some overlap between the lowest detected values and the highest upper bounds for censored observations, for these compounds the lowest detected (doubly normalized) value might be 2–30 times greater than the highest upper bound for a nondetect. This gap, often enormous in the log-transformed data, had a substantial negative impact on the quality of the downstream timeseries fits. The chemical similarity of the compounds affected by this phenomenon strongly suggests an analytical error, possibly waxing out. Rather than biasing the dataset by discarding these values, we looked for an appropriate alternative upper bound.

In principle, it is appropriate to censor everything at the highest censoring level for the dataset. But since samples were analyzed at a wide range of calibration levels, the highest detection limit for a compound was often orders of magnitude higher than the lowest quantitative measurements, so this approach would introduce a massive and unnecessary uncertainty. Instead, we grouped samples by contamination bin (to keep the possible calibration levels within roughly the same range) and qcbatch (quality control batch, the experimentally relevant grouping), identified the maximum reporting limit for each group, and used those values as the alternative detection limit (altdl) for their groups.

To determine when to apply this alternative limit, we grouped compounds by contamination bin and applied two filters. First, only compound-contamination bin datasets that include at least 3 censored observations and at least 3 uncensored observations were eligible for altdl. Second, for dual-normalized data from these eligible datasets, we compared the lowest uncensored measurement to the upper bounds of censored observations. We considered altdl necessary only where the lowest uncensored measurement exceeded ≥95% of censored observations' upper bounds. When any analyte passed both these filters in any one hopane concentration bin, we calculated and applied the alternative detection limit in *all* hopane concentration bins for that compound, to avoid apples-to-oranges comparisons. We then normalized the re-censored data. The compounds and number of observations affected by these steps in each contamination bin are summarized below. (Minor differences in re-censoring *n*-C38-normalized vs. hopane-normalized data reflect the fact that some samples were missing data for *n*-C38, and others for hopane, which can affect whether a given dataset was considered for altdl adjustment.)

Chemcode	Reference(s)	Low contamination (n recensored)	Medium contamination (n recensored)	High contamination (n recensored)
ACENAPTYLE	hop	22	1	11
AHCN_C18	hop, c38	142	25	6
AHCN_C20	hop, c38	50	8	3
AHCN_C29	hop, c38	15	3	0
ANTHRACENE	hop, c38	69	20	23
BNZFL_23	hop, c38	118	23	4
C1BZBTHIOP	hop, c38	193	36	13
C1DECALINS	hop, c38	108	9	0
C1FLUORENS	hop, c38	64	10	1
C1FLUORPYR	hop, c38	4	0	0
C1NTHPHNE	hop, c38	32	0	0
C1PHENANCS	hop, c38	24	3	0
C2BZBTHIOP	hop, c38	151	20	8
C2DBZTHIOP	hop, c38	74	5	0
C2DECALINS	hop, c38	135	11	0
C2FLUORENS	hop, c38	93	9	2
C2NAPHTHS	hop, c38	28	2	1
C2PHENANCS	hop, c38	24	3	0
C3BZBTHIOP	hop, c38	157	22	9
C3DBZTHIOP	hop, c38	96	6	0
C3FLUORENS	hop, c38	131	12	3
C3NAPHTHS	hop, c38	32	4	0
C3PHENANCS	hop, c38	37	0	0
C4BZBTHIOP	hop, c38	197	37	9
C4CHRYSENS	hop, c38	9	1	1
C4DBZTHIOP	hop, c38	102	3	0
C4NAPHTHS	hop, c38	73	10	0
C4NTHPHNE	hop, c38	5	0	0
C4PHENANCS	hop, c38	73	3	0
NAPTHALENE	hop	39	7	8
THHOP2S01	hop	5	0	2

Since censored compounds were filtered out of fingerprinting calculations, recensoring does not affect the results of MDI analysis.

1.6 Time-series analysis

1.6.1 Constructing fits

We applied two final filters before fitting: we only attempted to fit those compound-contamination bin datasets for which there are at least 10 uncensored datapoints, and for which no more than 80% of observations are censored. (In practice, the second criterion was a distinction without a difference, as the datasets filtered out are a subset of those caught by the first criterion.)

Hopane contamination level	Datasets removed
<400 ng g^{-1}	12
400–750 ng g^{-1}	4
≥750 ng g^{-1}	3

After applying this filter, we added the source oil t = 0 measurements to each dataset.

We fit all datasets to three candidate models: simple exponential decay, with y-intercept fixed at ln(fraction remaining) = 0; simple exponential decay, with y-intercept allowed to vary; or two-phase exponential decay. The physical basis for the latter model was that oily particles may have been subject to two phases of weathering, first suspended in the plume and then on the seafloor after deposition. Note that the dataset includes no measurements within the first, suspended phase, only end-members (source oil; the earliest sediment samples). Thus, we did not attempt to fit the breakpoint. The first sediment samples with Macondo-identified oil were collected at 160 d post-explosion, so we took this as the late bound for possible breakpoints. We did not attempt to draw conclusions regarding the kinetic form of the decay in the first phase, only regarding the extent of decay.

For each compound-contamination bin dataset, we built pseudoreplicates by adding uniformly distributed noise to the time axis to reflect the 87-day uncertainty in time from emission (sometime between the explosion and the well closure) to sampling (a known date). We fit the three models to each pseudoreplicate, and used the Bayesian information criterion (BIC) to choose the best-fit model for that pseudoreplicate. BIC penalizes models with more parameters, so that the two-phase model needs to more than trivially outperform the single-phase models in order to be called the best fit. We considered ΔBIC ≥ 6 to constitute strong support for a given model, and $2 \leq \Delta$BIC < 6 to constitute moderate support. We repeated this process (adding noise to create a pseudoreplicate, fitting three models, comparing BIC) 100 times per dataset.

1.6.2 Assessing fits

Pull distributions As an initial global check on fit quality, we examined the pull distribution. The "pull" of a fit is defined as the best-fit slope, divided by the fitted error on that slope. In general, the distribution of these pull terms across datasets is expected to be a Gaussian of unit width, and deviations from this form are diagnostic of an analytical problem. In the present case, if there had been no weathering, the pull distribution would be centered on zero, and all non-zero values would arise from statistical fluctuations. Weathering should distort the negative arm of this distribution: negative pull values will arise from an unknown mixture of (1) true negative slopes in datasets showing biodegradation and (2) fluctuations affecting zero-slope datasets. Thus the negative arm of the pull distribution is not useful for diagnosis of these fits. By contrast, positive pull values should still arise only from fluctuations affecting genuinely zero-slope datasets. As such, positive pull values should be drawn from a Gaussian with mean 0 and standard deviation 1, truncated at 0. Because the expected mean of this underlying distribution is zero, its standard deviation (the "pull s.d.") is given by the root-mean-square of observed positive pull values.

We calculated the pull s.d. for fits of all hopane-normalized and *n*-C38-normalized pseudoreplicate datasets for aliphatic, aromatic, and biomarker compounds at low, moderate, and high contamination levels. Whereas the pull s.d. of hopane-normalized datasets varied widely, indicating a pathology in the dataset, the pull s.d. values of *n*-C38-normalized datasets clustered around the expected value of 1 (Fig. S4). In particular, the pattern of steadily increasing pull with aliphatic chain length strongly suggests that, for *Deepwater Horizon* oil from the deep plume, hopane cannot be treated as a generally conservative biomarker. We proceeded with analysis of *n*-C38-normalized data instead.

Pseudoreplicate concordance The best-fit model and fit parameters may differ across pseudoreplicates for a given dataset. To summarize results and model predictions, we first asked, for each pseudoreplicate best fit, (1) whether the model predicts that less than 1% of the compound remained by 160 d and (2) whether the 95% confidence interval of the fitted slope excludes zero. (Note that the slope in question was that describing the period t $>$ 160 d. For the single-phase models, this was the only slope parameter, which was fitted to all data from t = 0 onward.)

If either case applied, that pseudoreplicate offered no meaningful evidence of degradation ongoing in sediment, and that pseudoreplicate's best prediction at all timepoints $>$ 160 d is the median of measured values, independent of time, with uncertainties given by the interquartile range of measured values.

Otherwise, that pseudoreplicate *did* support the conclusion that degradation continued after deposition, and its best prediction at a time of interest is the fitted value and 95% CI at that time. The confidence interval derives from two error terms: first, the standard error on the fitted value, which incorporates errors on all the fitted parameters (1–2 slopes and 0–1 intercepts); and second, typically much smaller, the error from uncertainty in sampling time, which we estimate as the standard error of the mean of fitted values across the ensemble of pseudoreplicates of the dataset. We added these errors in quadrature and calculated the 95% CI as fitted value ± 1.96 × combined error.

For each set of pseudoreplicates, we then identified the median result (and associated uncertainties) at each timepoint of interest. We tallied the number of pseudoreplicates that did or did not give evidence of ongoing degradation after deposition, and the number best fit by each model. In general, we found good agreement (i.e., concordant results from >95 of 100 pseudoreplicates) within sets of pseudoreplicates on both evidence for degradation and best-fit model:

Degradation predictions	<400 ng g^{-1}	400–750 ng g^{-1}	≥750 ng g^{-1}
Concordant	108	117	93
Discordant	4	3	28

Best-fit model	<400 ng g^{-1}	400–750 ng g^{-1}	≥750 ng g^{-1}
Concordant	103	119	94
Discordant	9	1	27

This consistency was notably weaker in the highest contamination bin, where slower loss made the degradation signal more vulnerable to noise in the time coordinate.

Looking at trends in model choice more broadly, we found that *only* the two-phase model receives strong support (ΔBIC ≥ 6), and it does so in nearly half of all pseudoreplicates (17115 of 35300).

Model	ΔBIC < 2	$2 \leq \Delta$BIC < 6	ΔBIC ≥ 6
onephase	3042	1240	0
onephase.int0	2084	7021	0
twophase	2274	2524	17115

Extensive early loss was the feature most commonly shared by pseudoreplicates best fit by the biphasic model (Fig. S5), with 25% of these predictions showing 1% contamination remaining by the model breakpoint of 160 d post-explosion, and 73% of predictions showing 25% contamination at 160 d. Early loss was far more limited in pseudoreplicates best fit by the single-phase models (Fig. S5), with 2% of fixed-intercept models and 1% of varying-intercept models predicting 25% contamination remaining at 160 d.

Mixed-method influence As noted in Methods, for 14 samples collected early in the spill response (Fig. S15) that both pass our MDI criterion and have *n*-C38 data available for normalization, biomarker measurements were made using exclusively a method that was subsequently disfavored. These measurements were included in the analysis above, but there remained the possibility that the change in methods over time could skew the results. We assessed the influence of these measurements by re-fitting these compound–contamination bin datasets with the 14 samples omitted. Although there were minor differences between the resulting predictions for compound-contamination bin datasets showing evidence of post-deposition degradation, >91% of median predictions at 160 days post-explosion, and >95% at 4 years, agree to within ± 2% whether or not these measurements are included (Fig. S16, Fig. S17). The single largest effect is a difference of 6% in the predicted fraction of C26 tricyclic terpane (22S) remaining after 4 years in highly contaminated samples. In only one case (tetrakishomohopane (22R) at moderate contamination) did the inclusion of these samples affect the determination of whether or not degradation continued after deposition.

1.7 Additional analyses

1.7.1 MDI projection

To assess the useful lifetime of the MDI approach to identifying Macondo oil contamination, we used the results of kinetic analysis to predict the changes in each of the 12 MDI biomarker ratios over time, at each contamination level. For each compound in a given ratio, we determined the predicted fraction remaining based on each pseudoreplicate at a series of times from 160 d to 10 years. We applied these predictions separately to the measured concentrations of the MDI compounds in the 11 distinct source oil samples, calculating 100 predicted values per ratio per sample per timepoint. We determined the median ratio value for each sample at each timepoint, and the median for the ensemble of samples. Because the dissimilarity metric is constructed such that the penalty increases for values both lower and higher than the reference range, we determined the median ratio value and calculated the associated dissimilarity, rather than directly determining the median dissimilarity score.

1.7.2 Hazen rate comparison

Hazen et al. (*6*) based their water-column rate calculations on samples collected at six sites:

- BM57 (t = 1 or t = 3) (per supporting information accompanying (*6*), nominal t = 1 was treated as t = 1.2)
- BM58 (t = 2 or t = 5)
- BM53 (t = 0)
- BM54 (t = 0)
- OV011 (t = 0)
- OV010 (t = 1 or t = 3) (per supporting information accompanying (*6*), nominal t = 1 was treated as t = 1.2)

No data from OV010 was included in the NRDA data release.

Although there are some inconsistencies in the fldsampids assigned in the earliest days of the spill response, the correspondence between NRDA samples and samples reported by Hazen as being in the plume could largely be reconstructed by examining the depths, sample dates, and IDs as reported by Hazen et al. We could not conclusively link NRDA sample BM640104 (likely a typo for BM064104) to Hazen sample BM64, upper depth 1099, but Hazen (a) lists BM064 as being in the plume, and (b) defines the plume boundaries as 1099–1219 m. As there was no other sample collected at 1099 m on the relevant cruises, we treated this correspondence as confirmed.

From examination of measured hydrocarbon concentrations at the multiple water depths sampled on each cast, it is clear that, while the plume lay within the defined 1099–1219 m depth horizon, it did not *span* this range. For a given cast, hydrocarbon concentrations were much higher at one depth within this range than at any others. That is, the set of measurements collected at 1099–1219 m water depth constitutes a mixture of plume and non-plume samples.

Having reconstructed the NRDA dataset that most closely corresponds to the samples on which Hazen et al. reported, we attempted to fit linear models to log-transformed data under several conditions:

- Normalized to hopane concentration, normalized to *n*-C38 concentration, or not normalized;
- Fast current (2 d transit time) or slow current (5 d); and

- All samples within the 1099–1219 m water depth horizon ("plume and non-plume"), or just the one or two peak samples from each cast within this range ("plume")

As there are no censored measurements among these samples for the analytes considered, fitting was performed with `lm()` rather than with survival methods.

1.7.3 Revision of the hopane footprint contamination estimate

The estimate of seafloor oil contamination published in (*3*) was based on hopane concentrations. If hopane was in fact being biodegraded, then the concentration data that fed into the interpolation will have produced an underestimate of the true contamination burden. We re-ran the interpolation as previously described, using *n*-C38 data instead of hopane data for the same set of samples as previously reported (with the exception of the few samples for which no *n*-C38 data is reported). All data was uncensored. To determine the background *n*-C38 concentration for use in calculation of excess *n*-C38 burden in contaminated sediments, we calculated the mean *n*-C38 concentration in surficial sediment samples that pass our quality filters, have MDI > 1.8, and were collected ≥ 40 km from the wellhead.

Separately, we used the fitted hopane biodegradation kinetics to project backwards from the sampling date to $t = 0$, i.e., to estimate how much hopane would originally have been present in the oil deposited to each seafloor sample collected within the contamination footprint. We used these projections as input for an additional kriging run, again using the same parameters as previously reported in (*3*).

1.7.4 Assessment of residuals-vs.-distance relationship

To examine the hypothesis (proposed in (*5*)) that distance from the wellhead is a primary control on extent of degradation, we included only best-fit models with at least moderate support in our analysis of model residuals vs. distance from wellhead. While fitting to a linear model indicated statistically significant slopes in all compound class–contamination bin datasets, the model residual vs. distance relationship is negligible for biomarkers at moderate and high contamination.

2 SI Tables

Table S1. Best-fit estimates of the length of time post-explosion required for individual compounds to degrade to <5% remaining, in each of the three contamination bins: <400 ng g^{-1} hopane, 400–750 ng g^{-1} hopane, and 750 ng g^{-1} hopane. Because fitted degradation rates are likely to become increasingly inaccurate with continued physical changes to the deposited oil over time (e.g., due to bioturbation), we report estimates only up to 10 years; we report >10 yr for more recalcitrant compounds. Where less than 5% remained in the earliest available sediment data (t = 160 d post-explosion), we report <160 d. n.a., >5% remaining at 160 d and no detectable ongoing loss. n.d., insufficient data to fit. (As in Fig. 4 and Fig. S8, decalin is grouped with aromatic compounds.)

Carbons	**Name**	**<400 ng g^{-1} hopane**	**400–750 ng g^{-1} hopane**	**≥750 ng g^{-1} hopane**
Aliphatics				
9	Nonane (C9)	n.d.	n.d.	~190 d
10	Decane (C10)	n.d.	<160 d	<160 d
11	Undecane (C11)	n.d.	<160 d	~290 d
12	Dodecane (C12)	n.d.	<160 d	<160 d
13	Tridecane (C13)	n.d.	<160 d	~1.1 yr
14	Tetradecane (C14)	<160 d	<160 d	~1 yr
15	2,6,10 Trimethyldodecane (1380)	n.d.	<160 d	~1.8 yr
15	Pentadecane (C15)	n.d.	<160 d	~1.4 yr
16	2,6,10 Trimethyltridecane (1470)	n.d.	<160 d	~2.3 yr
16	Hexadecane (C16)	<160 d	<160 d	n.a.
17	Heptadecane (C17)	<160 d	<160 d	~1.5 yr
18	Norpristane (1650)	n.d.	<160 d	~2 yr
18	Octadecane (C18)	<160 d	<160 d	n.a.
19	Nonadecane (C19)	<160 d	<160 d	~1.1 yr
19	Pristane	<160 d	<160 d	~2.8 yr
20	Eicosane (C20)	<160 d	<160 d	~1.2 yr
20	Phytane	<160 d	<160 d	~2.4 yr
21	Heneicosane (C21)	<160 d	<160 d	n.a.
22	Docosane (C22)	<160 d	<160 d	~1.5 yr
23	Tricosane (C23)	<160 d	<160 d	n.a.
24	Tetracosane (C24)	<160 d	<160 d	~2.1 yr
25	Pentacosane (C25)	~230 d	~1.2 yr	~2.6 yr
26	Hexacosane (C26)	<160 d	<160 d	~2.2 yr
27	Heptacosane (C27)	<160 d	<160 d	~2.7 yr
28	Octacosane (C28)	~3 yr	~4.5 yr	~3.1 yr
29	Nonacosane (C29)	~9 yr	n.a.	~4.5 yr
30	Triacontane (C30)	>10 yr	n.a.	n.a.
31	Hentriacontane (C31)	>10 yr	n.a.	n.a.
32	Dotriacontane (C32)	>10 yr	n.a.	>10 yr
33	Tritriacontane (C33)	n.a.	n.a.	n.a.
34	Tetratriacontane (C34)	n.a.	n.a.	>10 yr
35	Pentatriacontane (C35)	n.a.	n.a.	>10 yr
36	Hexatriacontane (C36)	n.a.	n.a.	n.a.
37	Heptatriacontane (C37)	n.a.	n.a.	n.a.
Aromatics				
9	Benzo(b)thiophene	n.d.	<160 d	<160 d
10	C1-Benzo(b)thiophenes	n.d.	~1.2 yr	n.a.
10	Naphthalene	<160 d	<160 d	<160 d
10	cis/trans-Decalin	<160 d	<160 d	n.a.
11	C1-Decalins	<160 d	<160 d	n.a.

Continued on next page

Carbons	**Name**	**<400 ng g^{-1} hopane**	**400–750 ng g^{-1} hopane**	**≥750 ng g^{-1} hopane**
11	C1-Naphthalenes	<160 d	<160 d	<160 d
11	C2-Benzo(b)thiophenes	~1.4 yr	~2.6 yr	n.a.
12	Acenaphthene	<160 d	<160 d	<160 d
12	Acenaphthylene	n.a.	n.a.	n.a.
12	Biphenyl	<160 d	<160 d	<160 d
12	C2-Decalins	<160 d	<160 d	n.a.
12	C2-Naphthalenes	<160 d	<160 d	<160 d
12	C3-Benzo(b)thiophenes	~220 d	n.a.	n.a.
13	C3-Naphthalenes	<160 d	<160 d	n.a.
13	C4-Benzo(b)thiophenes	n.d.	n.a.	n.a.
13	Dibenzofuran	<160 d	<160 d	<160 d
13	Dibenzothiophene	<160 d	<160 d	<160 d
13	Fluorene	<160 d	<160 d	<160 d
14	Anthracene	~2 yr	n.a.	n.a.
14	C1-Dibenzothiophenes	<160 d	<160 d	n.a.
14	C1-Fluorenes	<160 d	<160 d	n.a.
14	C4-Naphthalenes	<160 d	<160 d	n.a.
14	Phenanthrene	<160 d	<160 d	<160 d
15	C1-Phenanthrenes/anthracenes	<160 d	<160 d	n.a.
15	C2-Dibenzothiophenes	<160 d	<160 d	n.a.
15	C2-Fluorenes	<160 d	<160 d	n.a.
16	C2-Phenanthrenes/anthracenes	<160 d	<160 d	n.a.
16	C3-Dibenzothiophenes	<160 d	n.a.	n.a.
16	C3-Fluorenes	<160 d	n.a.	n.a.
17	Benzo(b)fluorene	~2.2 yr	n.a.	~5.5 yr
17	C1-Fluoranthenes/pyrenes	~3.6 yr	n.a.	n.a.
17	C3-Phenanthrenes/anthracenes	<160 d	n.a.	n.a.
17	C4-Dibenzothiophenes	~1.5 yr	n.a.	~5.7 yr
17	Naphthobenzothiophene	<160 d	n.a.	~4.6 yr
18	Benzo(a)anthracene	n.a.	n.a.	~4.6 yr
18	C1-Naphthobenzothiophenes	~2.6 yr	n.a.	~6.2 yr
18	C2-Fluoranthenes/pyrenes	~3.7 yr	>10 yr	~7.3 yr
18	C4-Phenanthrenes/anthracenes	~1.5 yr	n.a.	~5.9 yr
18	Chrysene + Triphenylene	~3.4 yr	~7.9 yr	~9.2 yr
19	C1-Chrysenes	~3.1 yr	>10 yr	~9.9 yr
19	C2-Naphthobenzothiophenes	~3.7 yr	>10 yr	~8 yr
19	C3-Fluoranthenes/pyrenes	~4.5 yr	>10 yr	~8.5 yr
20	Benzo(b)fluoranthene	>10 yr	>10 yr	n.a.
20	Benzo(e)pyrene	~8.5 yr	>10 yr	n.a.
20	C2-Chrysenes	~4.5 yr	>10 yr	~7.6 yr
20	C3-Naphthobenzothiophenes	~6.6 yr	>10 yr	>10 yr
20	C4-Fluoranthenes/pyrenes	~5.6 yr	>10 yr	>10 yr
21	C3-Chrysenes	~7.5 yr	>10 yr	>10 yr
21	C4-Naphthobenzothiophenes	~9.7 yr	>10 yr	n.a.
22	C4-Chrysenes	>10 yr	>10 yr	n.a.
Biomarkers				
23	C23 Tricyclic Terpane	>10 yr	>10 yr	>10 yr
24	C24 Tricyclic Terpane	>10 yr	>10 yr	n.a.
25	C25 Tricyclic Terpane	>10 yr	>10 yr	>10 yr
26	C26 Tricyclic Terpane-22S	>10 yr	>10 yr	>10 yr
26	C26,20R- +C27,20S- Triaromatic Steroid	>10 yr	>10 yr	>10 yr
27	13B(H),17A(H)-20R-Diacholestane	~2.4 yr	n.a.	~8 yr
27	13B(H),17A(H)-20S-Diacholestane	~4.1 yr	n.a.	~9 yr

Continued on next page

Carbons	Name	<400 ng g^{-1} hopane	400–750 ng g^{-1} hopane	≥750 ng g^{-1} hopane
27	14A(H),17A(H)-20R-Cholestane	~4.6 yr	n.d.	n.d.
27	14A(H),17A(H)-20S-Cholestane	~7.2 yr	n.d.	n.d.
27	14B(H),17B(H)-20R-Cholestane	~3.8 yr	n.a.	>10 yr
27	14B(H),17B(H)-20S-Cholestane	~5 yr	n.a.	>10 yr
27	17A(H)-22,29,30-TRISNorhopane-TM	>10 yr	>10 yr	n.a.
27	18A-22,29,30-Trisnorneohopane-TS	>10 yr	>10 yr	>10 yr
27	C27,20R-Triaromatic Steroid	>10 yr	>10 yr	>10 yr
28	13B,17A-20S-Methyldiacholestane	>10 yr	>10 yr	n.a.
28	14A,17A-20R-Methylcholestane	~8.4 yr	>10 yr	n.a.
28	14A,17A-20S-Methylcholestane	>10 yr	>10 yr	n.a.
28	14B,17B-20R-Methylcholestane	~8.3 yr	>10 yr	>10 yr
28	14B,17B-20S-Methylcholestane	>10 yr	>10 yr	n.a.
28	C28,20R-Triaromatic Steroid	>10 yr	>10 yr	>10 yr
28	C28,20S-Triaromatic Steroid	>10 yr	>10 yr	n.a.
29	13B,17A-20R-Ethyldiacholestane	n.a.	n.d.	n.d.
29	14A(H),17A(H)-20R-Ethylcholestane	>10 yr	>10 yr	n.a.
29	14A(H),17A(H)-20S-Ethylcholestane	>10 yr	>10 yr	n.a.
29	14B(H),17B(H)-20R-Ethylcholestane	~9.6 yr	>10 yr	n.a.
29	14B(H),17B(H)-20S-Ethylcholestane	>10 yr	>10 yr	n.a.
29	18A(H)-30-Norneohopane-C29TS	>10 yr	>10 yr	n.a.
29	30-Norhopane	>10 yr	>10 yr	n.a.
30	17A(H)-Diahopane	>10 yr	>10 yr	n.a.
30	17a(H),21b(H)-Hopane (Hopane)	>10 yr	>10 yr	n.a.
31	30-Homohopane-22S	~7.7 yr	>10 yr	n.a.
31	T22-C31-Homohopane (R)	~6.5 yr	>10 yr	>10 yr
32	30,31-Bishomohopane-22R	>10 yr	>10 yr	n.a.
32	30,31-Bishomohopane-22S	>10 yr	>10 yr	n.a.
33	30,31-Trishomohopane-22R	>10 yr	>10 yr	n.a.
33	30,31-Trishomohopane-22S	>10 yr	>10 yr	n.a.
34	Tetrakishomohopane-22R	~8.1 yr	>10 yr	n.a.
34	Tetrakishomohopane-22S	>10 yr	>10 yr	>10 yr
35	T34-Pentakishomohopane (S)	~8.1 yr	n.a.	>10 yr
35	T35-Pentakishomohopane (R)	n.a.	n.a.	n.a.

Table S2. Biomarker ratios used to calculate the Macondo dissimilarity index (MDI). A sample's MDI is the average of per-ratio penalty scores, assigned based on the divergence of each ratio in the sample from the measured range in reference oil (see SI Text).

Ratio	**Numerator**	**Denominator**
A	13β(H),17α(H)-20S-Diacholestane	13β(H),17α(H)-20R-Diacholestane
B	14β,17β-20S-Methylcholestane + 14β,17β-20R-Methylcholestane	17α(H),21β(H)-Hopane (Hopane)
C	14α(H),17α(H)-20S-Ethylcholestane	17α(H),21β(H)-Hopane (Hopane)
D	C25 Tricyclic Terpane	17α(H),21β(H)-Hopane (Hopane)
E	18α-22,29,30-Trisnorneohopane-TS	17α(H)-22,29,30-TRISNorhopane-TM
F	18α-22,29,30-Trisnorneohopane-TS	17α(H),21β(H)-Hopane (Hopane)
G	30,31-Bishomohopane-22S	17α(H),21β(H)-Hopane (Hopane)
H	30,31-Bishomohopane-22R	17α(H),21β(H)-Hopane (Hopane)
I	17α(H)-Diahopane	17α(H),21β(H)-Hopane (Hopane)
J	C26,20R- +C27,20S- Triaromatic Steroid	17α(H),21β(H)-Hopane (Hopane)
K	C28,20R-Triaromatic Steroid	17α(H),21β(H)-Hopane (Hopane)
L	C27,20R-Triaromatic Steroid	17α(H),21β(H)-Hopane (Hopane)

Table S3. Summary of samples in the MDI reference set. The sediment samples chosen are all of the surficial samples with measured hopane concentration $\geq$750 ng g^{-1} (i.e., $\geq$10-fold above background) that were collected within 1 km of the wellhead and within one year of the spill midpoint.

Study name	**Fldsampid**	**Sample IDs**	**Date**	**Distance (km)**
Chem–Source Oil 2010	GU2988-A0521-O9801	O002	2010-05-21	2.7
	GU2988-A0521-O9802	O003	2010-05-21	2.7
	GU2988-A0521-O9803	O004	2010-05-21	2.7
	GU2988-A0521-O9804	O005	2010-05-21	2.7
	GU2988-A0521-O9805	O006 O006D O006D2 O006D3	2010-05-21	2.7
	GU2988-A0521-O9871	O001	2010-05-21	2.7
	SO-20100814-Q4000-009	O001	2010-08-14	0.4
	SO-20100814-Q4000-025	O002	2010-08-14	0.4
	SO-20100814-Q4000-041	O003	2010-08-14	0.4
	SO-20100814-MASS-013	O004	2010-08-15	0.4
	SO-20100818-HP1-013	O001	2010-08-18	0.5
HOS Sweetwater Cruise 02 MAR 23-APR 24 2011	HSW2L2_FP0093_B0423_S_50_E2_859	S00E1	2011-04-23	1.0
	HSW2L2_FP0094_B0423_S_50_H2_868	S00H1	2011-04-23	0.4
	HSW2L2_FP0095_B0424_S_50_J2_875	S00J1	2011-04-24	0.2
Sarah Bordelon Cruise 09 MAY 23-JUN 13 2011	SB9-65-B0525-S-D038SW-HC-0026	S001	2011-05-25	0.3
	SB9-65-B0526-S-NF006MOD-HC-0419	S001	2011-05-26	0.9
	SB9-65-B0527-S-D040S-HC-0536	S001	2011-05-27	0.5
	SB9-65-B0527-S-D040S-HC-0576	S001	2011-05-27	0.5
	SB9-65-B0528-S-D034S-HC-0969	S001	2011-05-28	0.6
	SB9-65-B0528-S-D034S-HC-1008	S001	2011-05-28	0.6
	SB9-65-B0528-S-D034S-HC-1047	S001	2011-05-28	0.6
	SB9-65-B0529-S-ALTNF001-HC-1246	S001	2011-05-29	0.6
	SB9-65-B0529-S-ALTNF001-HC-1285	S001	2011-05-29	0.6

3 SI Figures

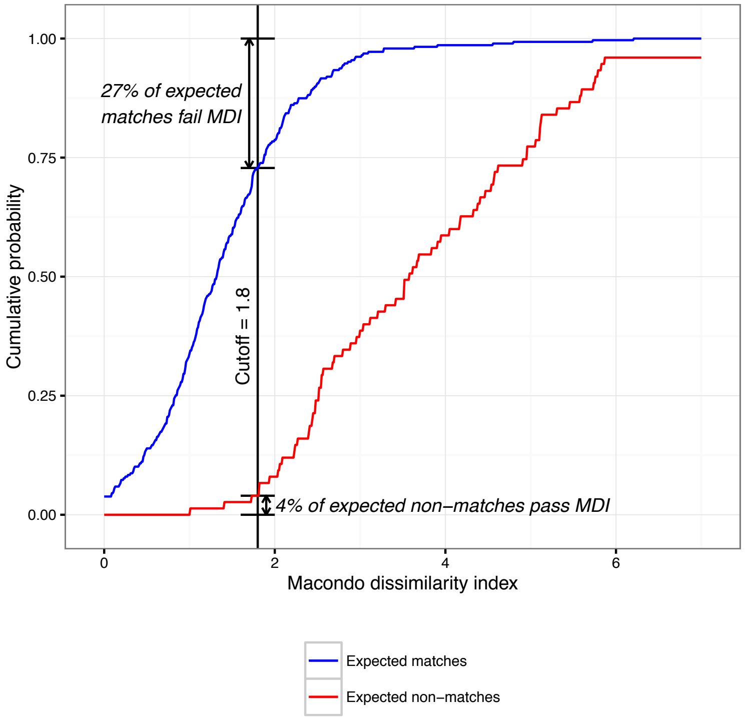

Figure S1: Empirical cumulative distribution functions for MDI scores of groups of samples expected to match the Macondo fingerprint (blue line; samples collected within 40 km of the wellhead, contaminated with above-background levels of hopane) and those expected not to match (red line; samples collected ≥ 1 km from the wellhead, with hopane concentration at or below the previously determined background concentration of 75 ng g^{-1}). The MDI cutoff value of 1.8 was chosen to reject 95% of expected non-matches.

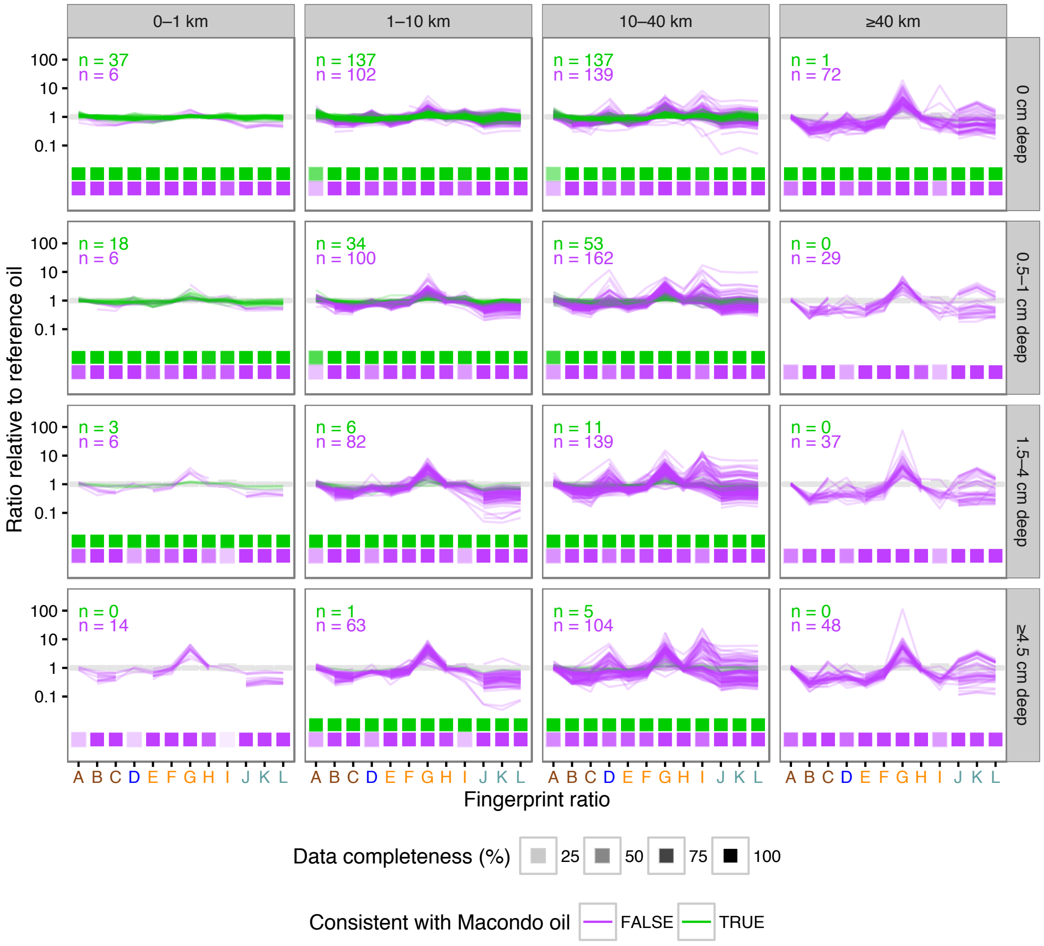

Figure S2: Spatial and depth distribution of fingerprint ratios used to calculate the Macondo dissimilarity index. Values plotted are normalized to the ratio's value in the Macondo reference set. Green, samples with MDI < 1.8; purple, samples with MDI ≥ 1.8. Squares show data completeness for each ratio (% of samples with measured concentrations within the range of quantitation); samples with <8 informative ratios were excluded from further analysis. Panels are binned by distance from the wellhead (from left to right, 0–1 km, 1–10 km, 10–40 km, ≥40 km) and upper depth in the sediment core (from top to bottom, 0 cm, 0.5–1 cm, 1.5–4 cm, ≥4.5 cm). Color of x-axis labels indicates type of biomarkers used in ratios A–L: brown (A–C), steranes; blue (D), terpanes; orange (E–I), hopanes; gray (J–L), triaromatic steroids. As throughout the paper, 17α(H),21β(H)-hopane is referred to as hopane. Ratios: A, 13β(H),17α(H)-20S-diacholestane / 13β(H),17α(H)-20R-diacholestane; B, (14β,17β-20S-methylcholestane + 14β,17β-20R-methylcholestane) / hopane; C, 14α(H),17α(H)-20S-ethylcholestane / hopane; D, C25 tricyclic terpane / hopane; E, 18α-22,29,30-trisnorneohopane (TS) / 17α(H)-22,29,30-trisnorhopane (TM); F, 18α-22,29,30-trisnorneohopane (TS) / hopane; G, 30,31-bishomohopane-22S / hopane; H, 30,31-bishomohopane-22R / hopane; I, 17α(H)-diahopane / hopane; J, C26,20R- + C27,20S-triaromatic steroid / hopane; K, C28,20R-triaromatic steroid / hopane; L, C27,20R-triaromatic steroid / hopane.

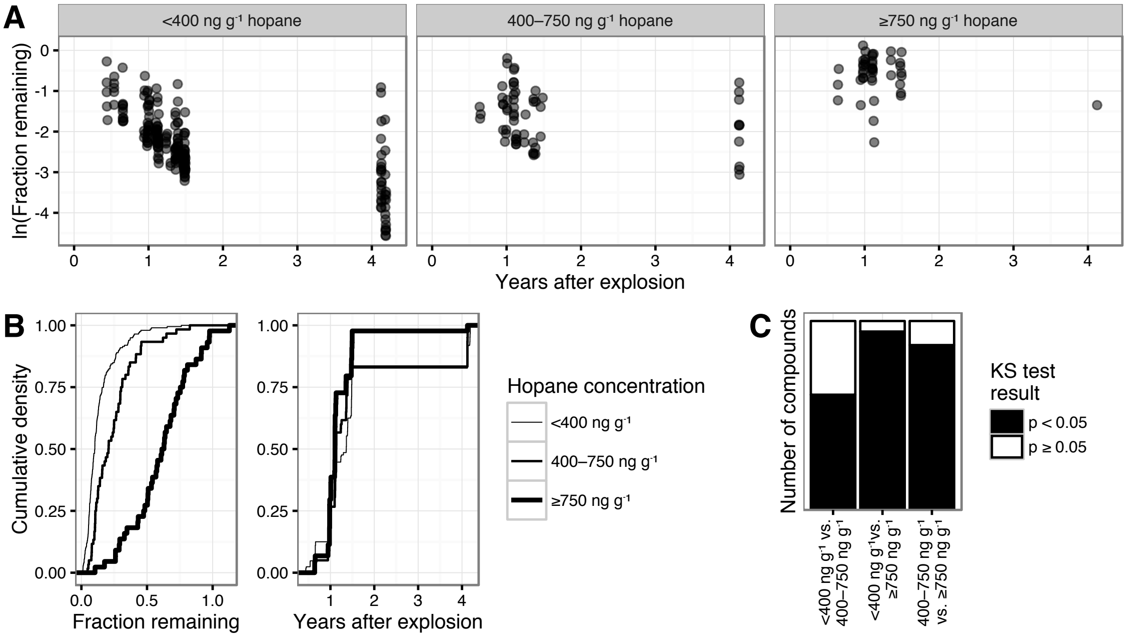

Figure S3: Evidence for contamination dependence in hydrocarbon loss rates. (A) Log-transformed fraction-remaining data for C1-chrysenes, binned by hopane concentration (left, <400 ng g^{-1}; center, 400–750 ng g^{-1}; right, $\geq$750 ng g^{-1}). (B) Empirical cumulative distribution functions for the fraction remaining (left) and days after explosion (right) for C1-chrysene data, showing that the apparent contamination dependence is not due to a systematic difference in sampling date. Line weights indicate contamination level: thin, <400 ng g^{-1} hopane; medium, 400–750 ng g^{-1}; thick, $\geq$750 ng g^{-1}. (C) Results of Kolmogorov-Smirnov tests comparing the empirical cumulative distribution functions of fraction-remaining data across hopane contamination bins. KS test results of $p < 0.05$ indicate evidence for a significant difference between the distributions.

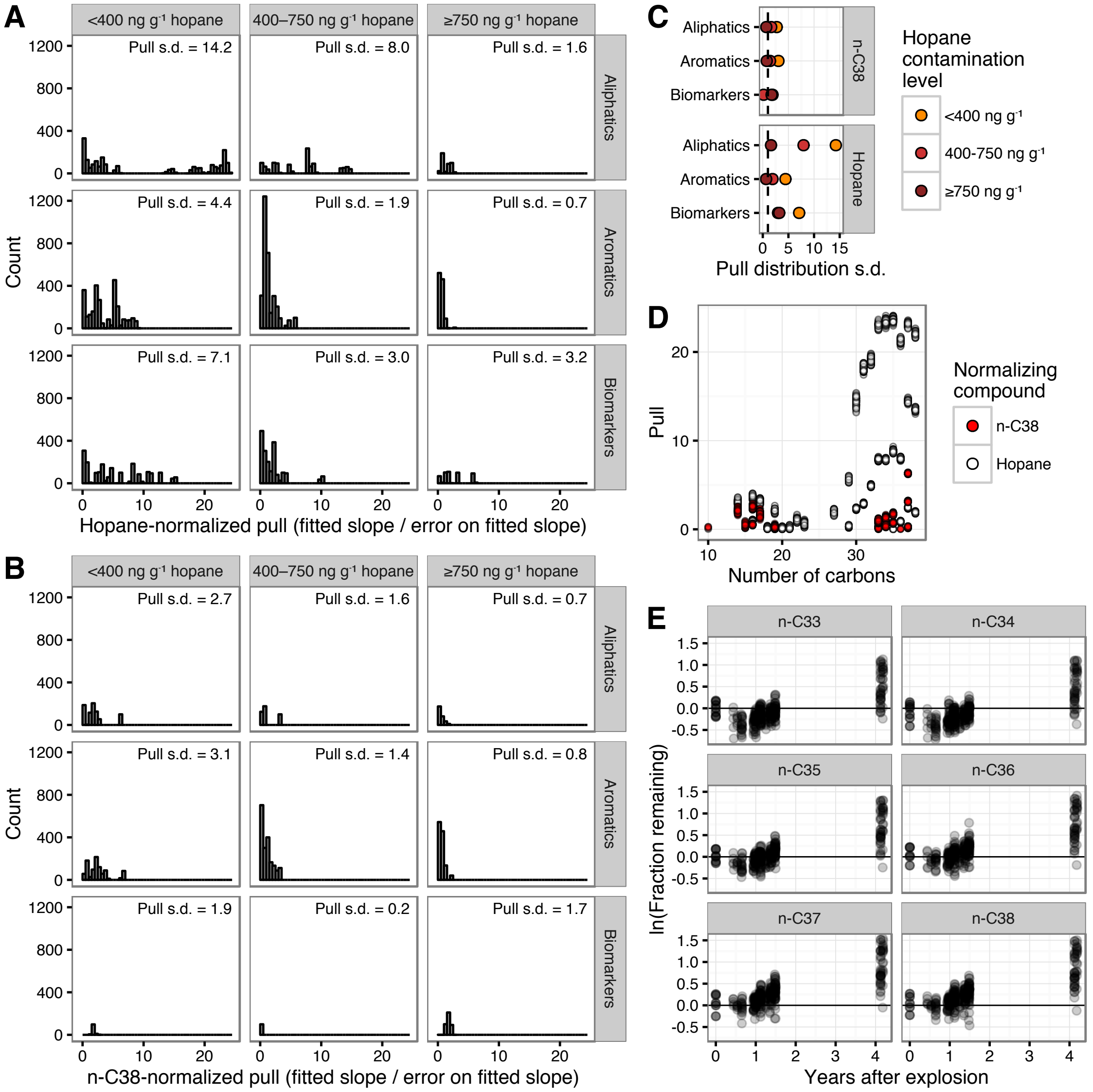

Figure S4: Analysis of positive pull distributions. (A) Distribution of positive pulls among hopane-normalized data. For well-behaved fits, positive pull distributions are expected to be the positive arm of Gaussian distributions centered on zero with unit standard deviation. (B) Distribution of positive pulls among *n*-C38-normalized hydrocarbon data. (C) Summary of positive pull distribution standard deviations, for fits to *n*-C38-normalized (top) and hopane-normalized (bottom) data. The expected standard deviation of a pull distribution for well-behaved fits is 1. (D) Relationship between pull (fitted slope / error on fitted slope) and chain length for aliphatic compounds, for fits to *n*-C38-normalized (red) and hopane-normalized (black) data. Results are shown for all best-fit pseudoreplicates with positive pull (0–100 pseudoreplicates per normalizing compound and contamination bin). (E) Log-transformed fraction-remaining data for long-chain aliphatics normalized to hopane, showing the large apparent increase in fraction remaining over time. Data from all contamination bins is plotted jointly.

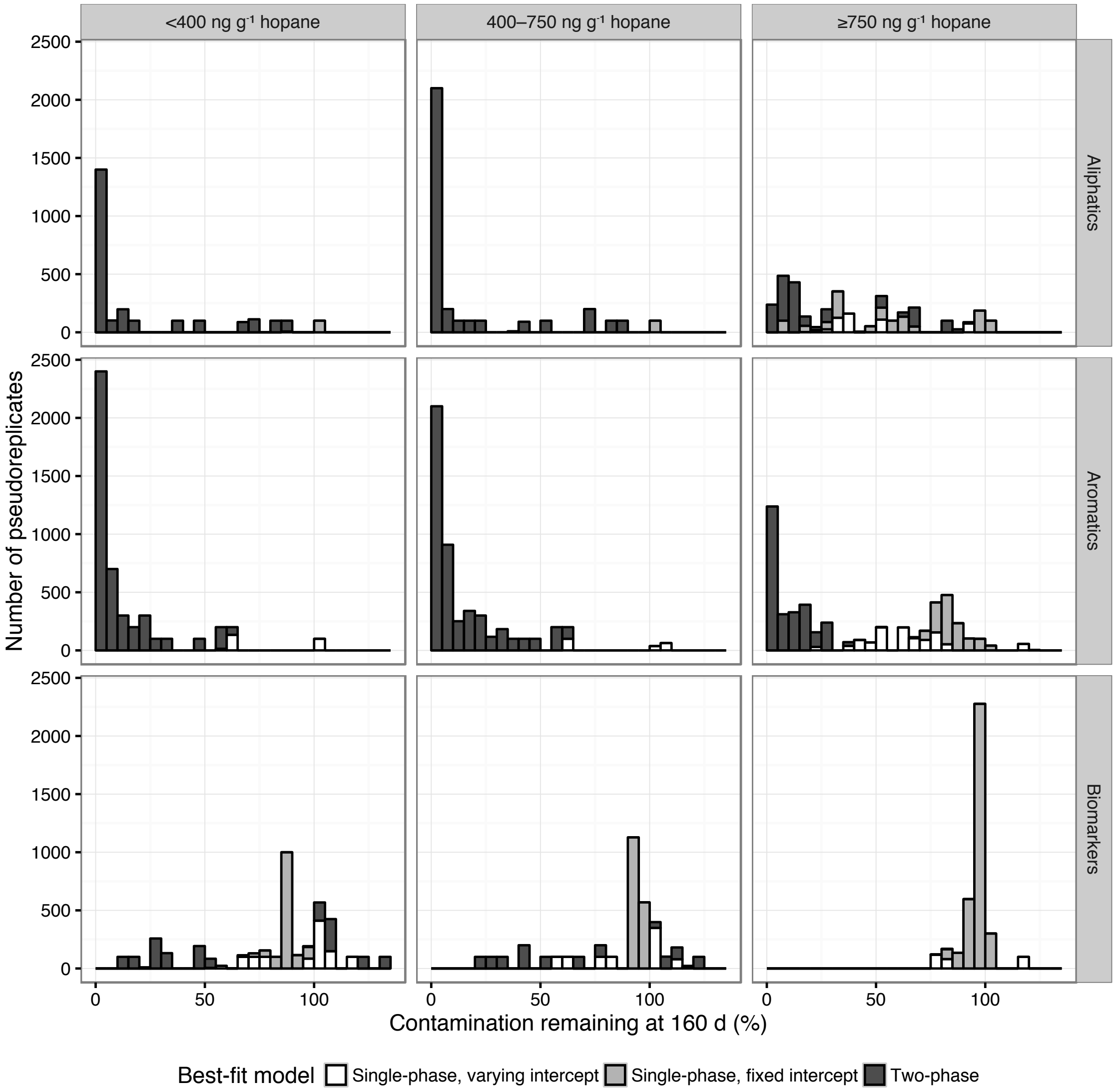

Figure S5: Distribution of best-fit models by fitted percent remaining at 160 d, by contamination level and compound type. Three models (single-phase, varying intercept; single-phase, fixed intercept; two-phase) were fitted to each of 100 pseudoreplicates of each dataset (see Methods); for each pseudoreplicate, the best-fit model was chosen by the Bayesian information criterion (BIC). The best-fit model could (and often did) differ across a given dataset's pseudoreplicates. Datasets with <50% remaining at 160 d were typically best fit by the two-phase model. All 100 best-fit models for each dataset's pseudoreplicates are plotted. White, pseudoreplicates best fit by the single-phase model with varying intercept; light gray, by the single-phase model with fixed intercept; dark gray, by the two-phase model.

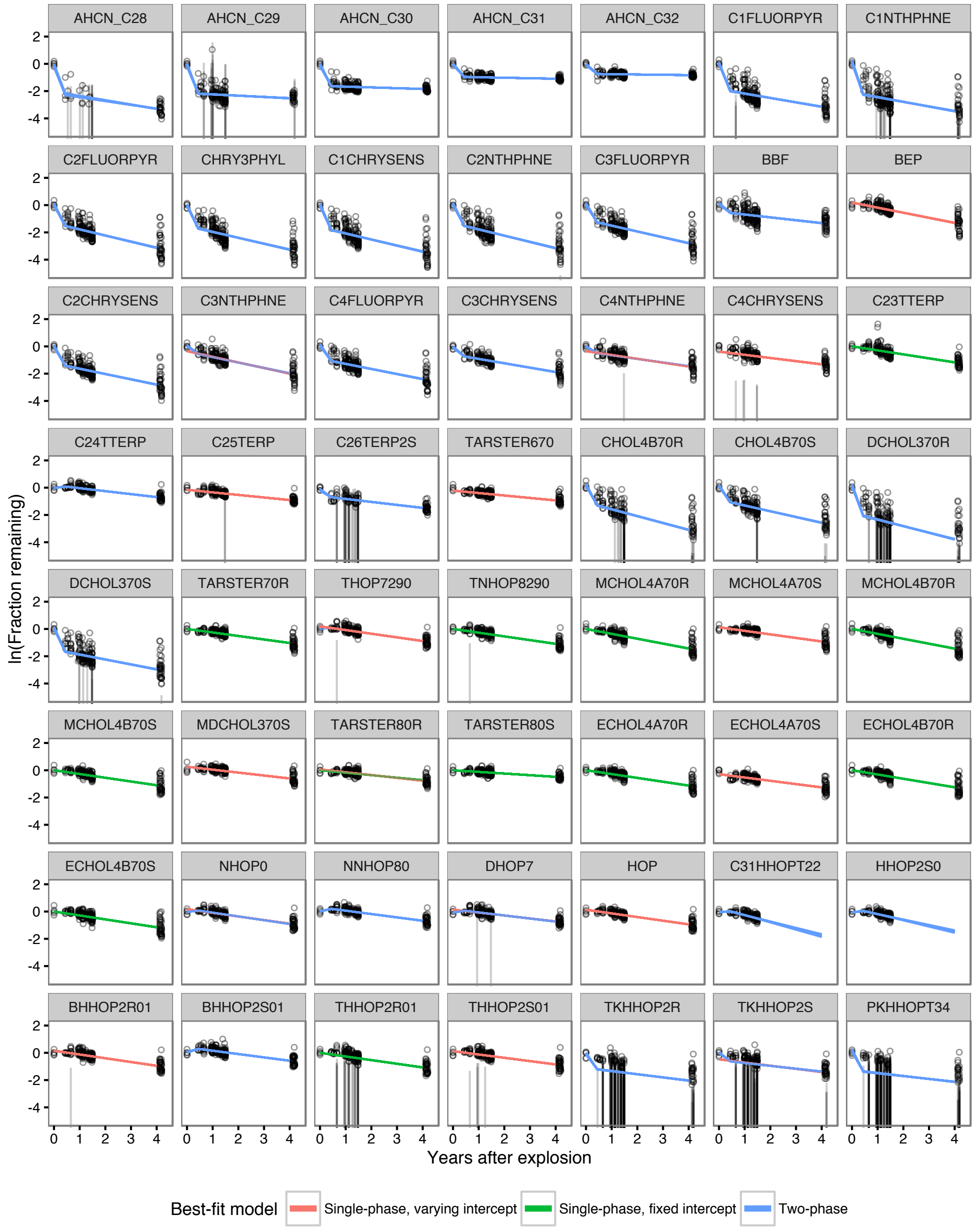

Figure S6: Data and fits for 56 of 58 low-contamination datasets in which ≥10% remained at 160 d and later biodegradation was detectable. The remaining two compounds (the 20R and 20S enantiomers of 14α(H),17α(H)-cholestane) are not shown because the available data spans a limited range (sediment samples collected 160–239 d post-explosion). Uncensored (quantitative) measurements are plotted as open circles. Censored (nondetect) measurements are plotted as thin black lines spanning the interval (from 0 to the detection limit) within which the concentration may lie. For fitting, the distribution of concentrations within these intervals was assumed to be lognormal (i.e., normal in the log-transformed dataset). For each dataset, all 100 pseudoreplicate fits are shown, colored by best-fit model: pink, single-phase with varying intercept; green, single-phase with fixed intercept; blue, two-phase.

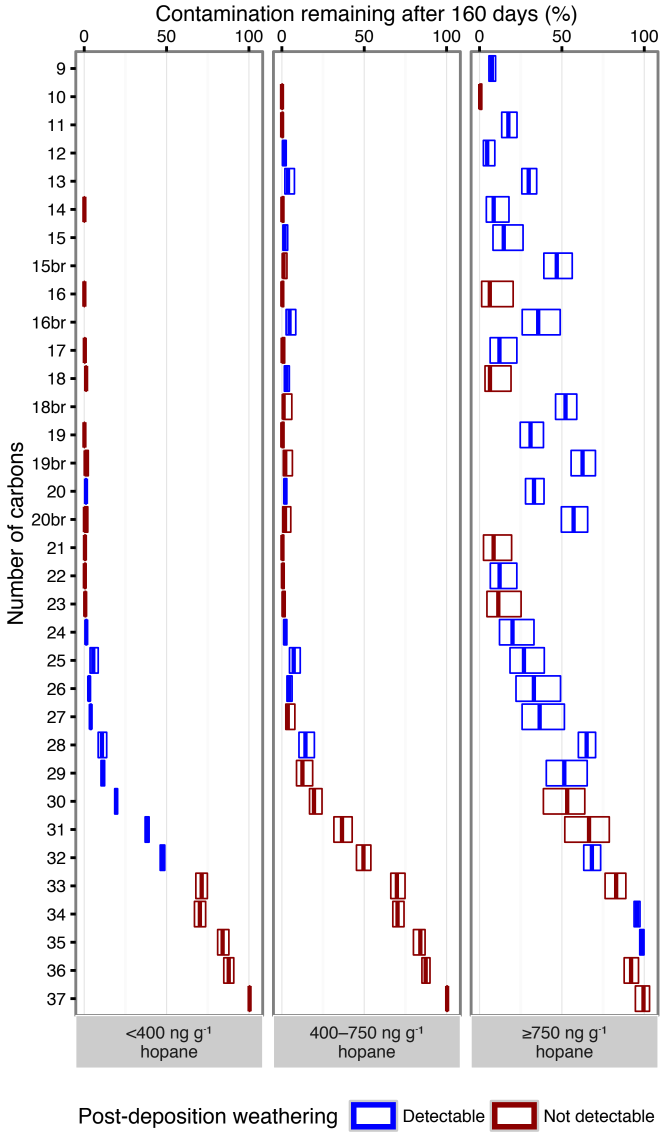

Figure S7: Percent remaining of aliphatic compounds at 160 d after the explosion, ordered by chain length. Branched compounds are indicated by 'br' on the y-axis. Compounds for which biodegradation was detectable after deposition are shown in blue, with crossbars indicating the fitted value and boxes the 95% CI of the median fit result. Compounds for which post-deposition biodegradation was not detectable are shown in red, with vertical bars indicating the median and boxes the interquartile range of measured values.

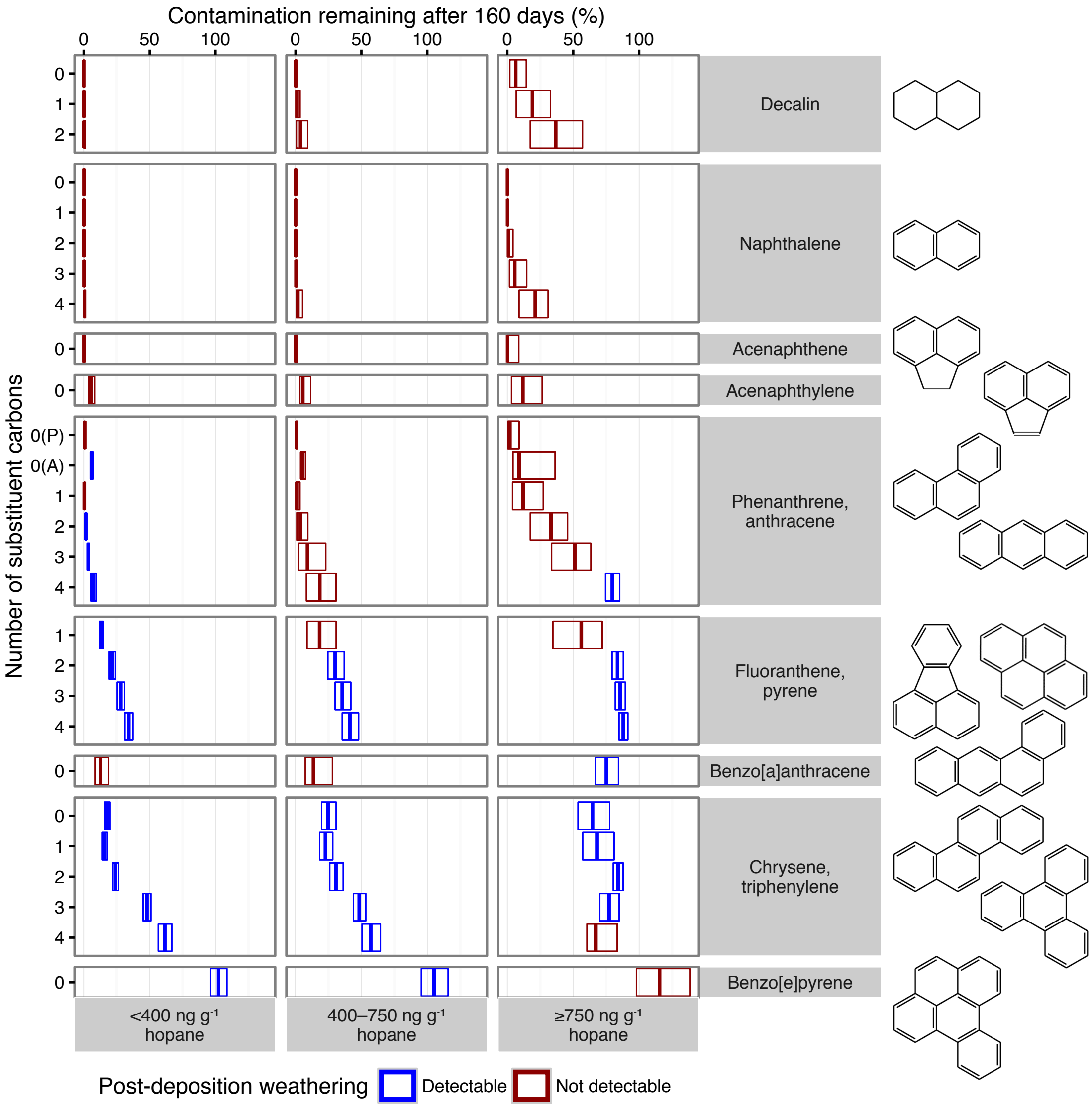

Figure S8: Percent remaining of aromatic compounds with six-membered rings, 160 d after the explosion. (Decalin is not aromatic but is included here.) Panels are ordered by increasing carbon skeleton size from top to bottom, and within each panel by increasing number of carbon substituents. Where multiple carbon skeletons are shown for a single group at right, the compounds in that category were not separately resolved in chemical analysis, unless otherwise indicated on the y-axis (0(P), unsubstituted phenanthrene only; 0(A), unsubstituted anthracene only). Compounds for which biodegradation was detectable after deposition are shown in blue, with crossbars indicating the fitted value and boxes the 95% CI of the median fit result. Compounds for which post-deposition biodegradation was not detectable are shown in red, with vertical bars indicating the median and boxes the interquartile range of measured values.

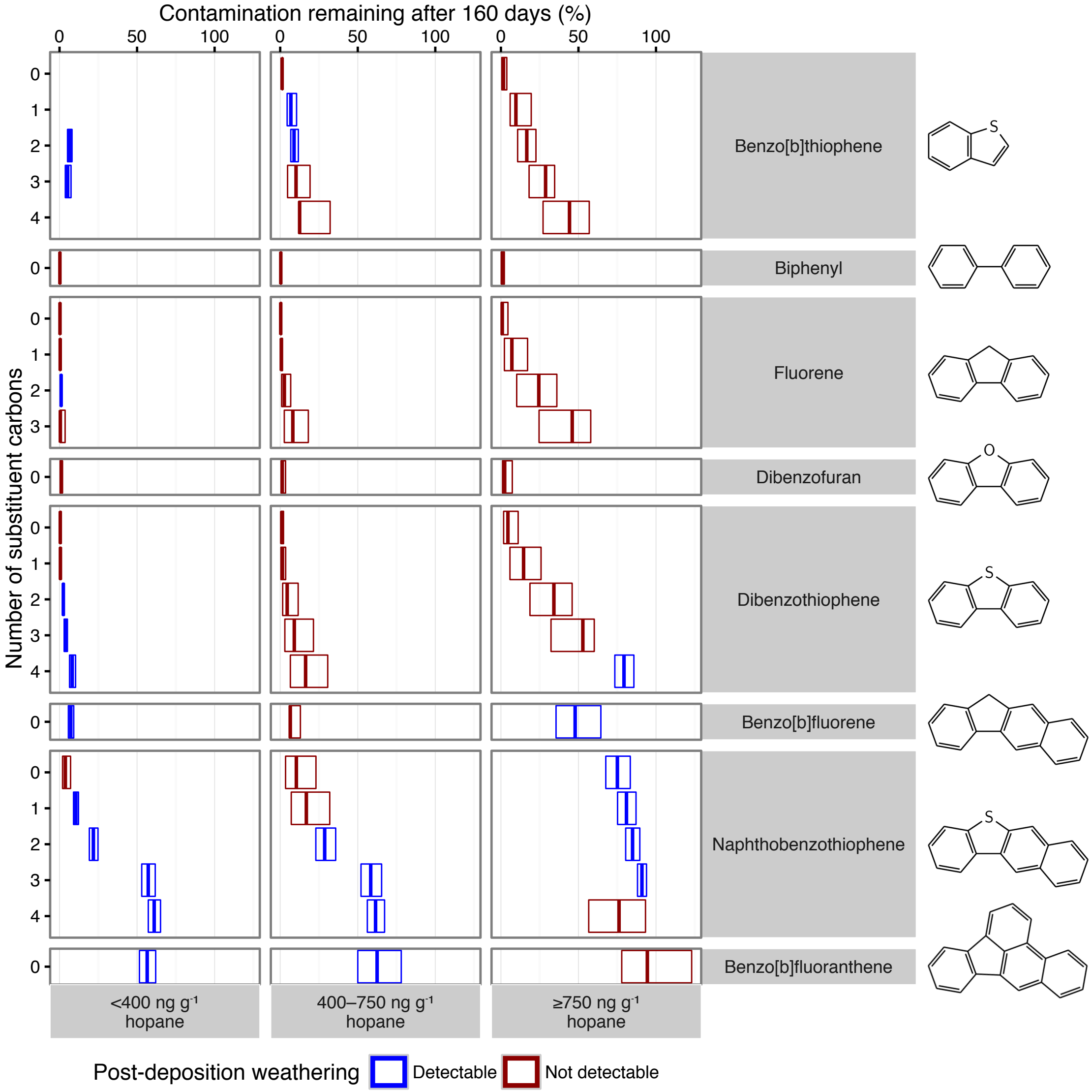

Figure S9: Percent remaining of aromatic compounds with five-membered rings, 160 d after the explosion. (Biphenyl is included here due to its structural resemblance to fluorene, dibenzofuran, and dibenzothiophene.) Panels are ordered by increasing carbon skeleton size from top to bottom, and within each panel by increasing number of carbon substituents. Compounds for which biodegradation was detectable after deposition are shown in blue, with crossbars indicating the fitted value and boxes the 95% CI of the median fit result. Compounds for which post-deposition biodegradation was not detectable are shown in red, with vertical bars indicating the median and boxes the interquartile range of measured values.

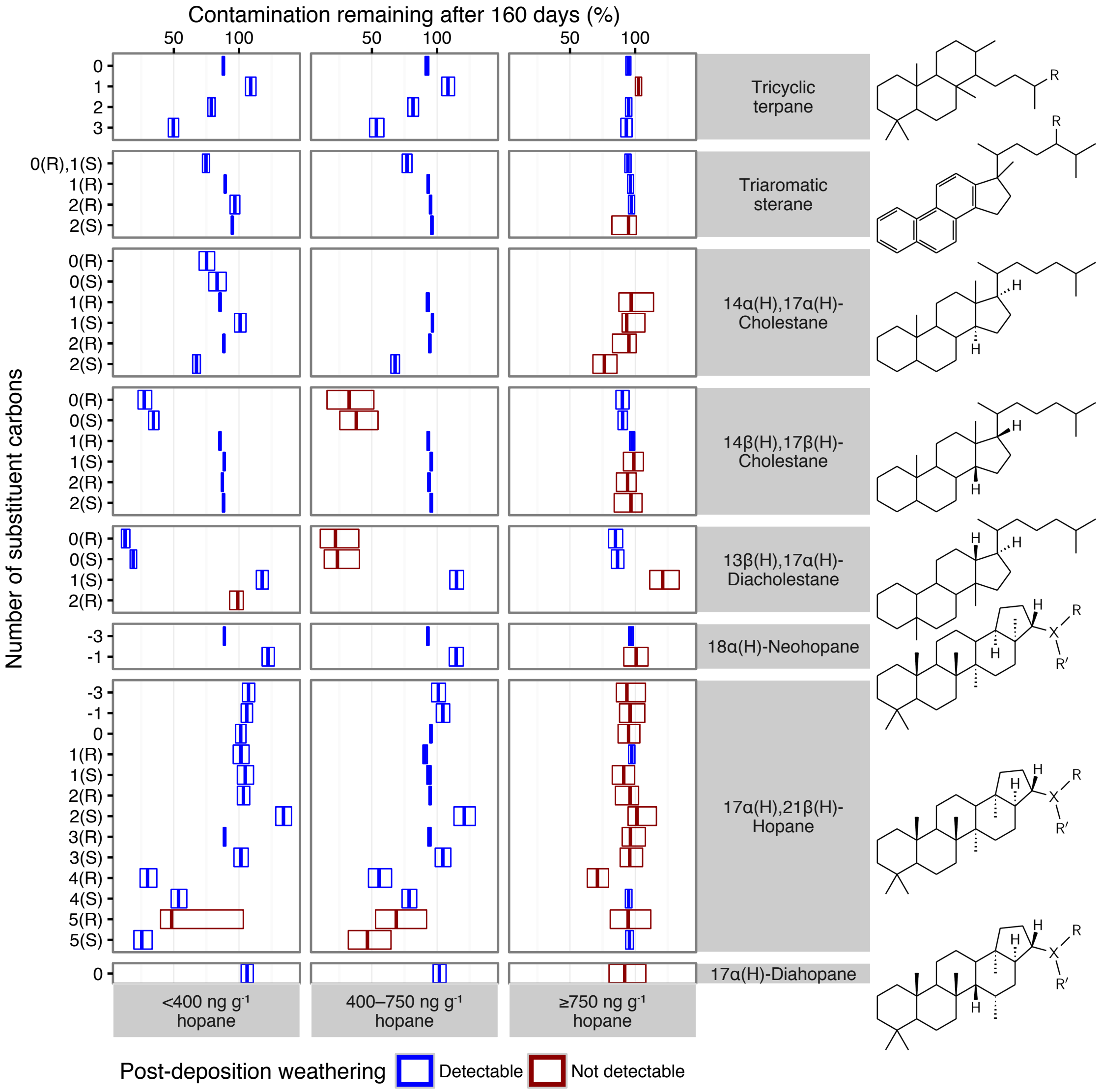

Figure S10: Percent remaining of biomarker compounds, 160 d after the explosion. Panels are ordered by increasing carbon skeleton size from top to bottom, and within each panel by increasing number of carbon substituents, with (R) and (S) substituent stereochemistry separately displayed. Among neohopanes and hopanes, -3 indicates the trisnor- compounds and -1 the nor- compounds; 1–5 indicates homohopanes through pentakishomohopanes. Compounds for which biodegradation was detectable after deposition are shown in blue, with crossbars indicating the fitted value and boxes the 95% CI of the median fit result. Compounds for which post-deposition biodegradation was not detectable are shown in red, with vertical bars indicating the median and boxes the interquartile range of measured values.

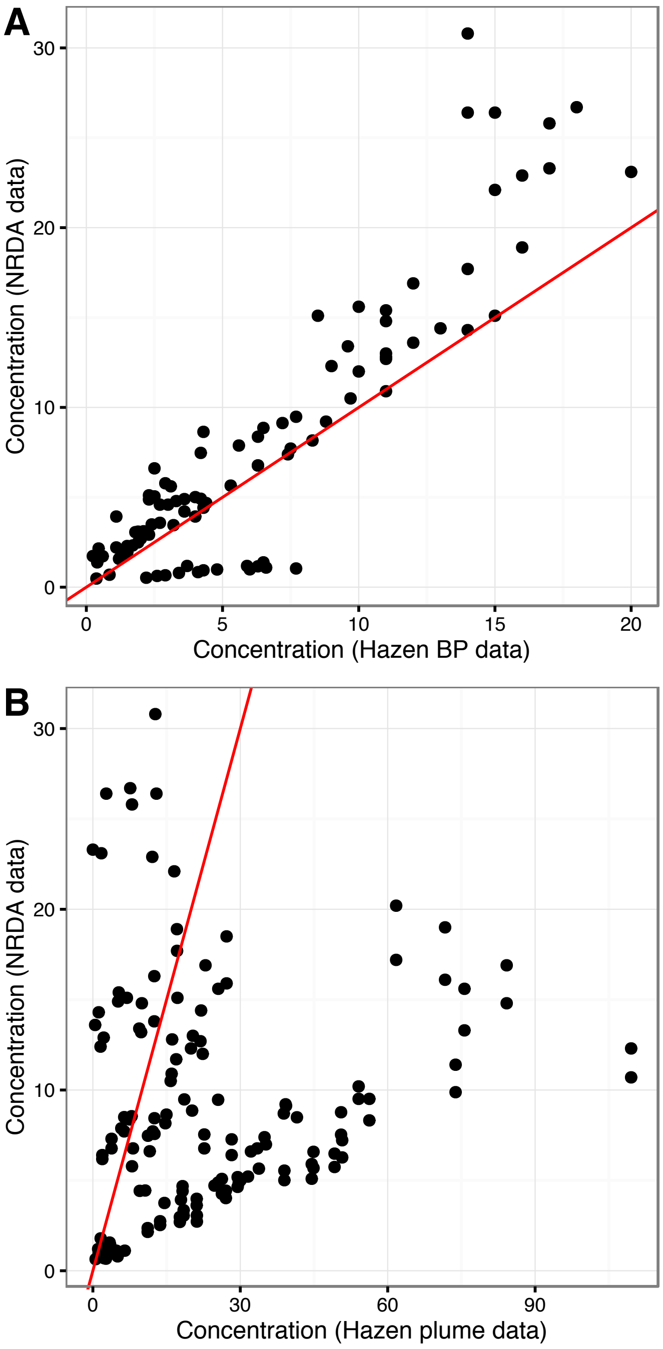

Figure S11: Comparison of datasets fitted in (*6*) and those fitted in the present re-analysis of this work. (A), current NRDA data vs. Hazen et al.'s "BP" dataset; (B), current NRDA data vs. Hazen et al.'s "plume" dataset. Current NRDA datapoints, from NRDA's water-column data release (see Methods), were matched to Hazen datapoints by station, depth, and sampling date. Differences reflect analysis of sample splits by different contract labs. Red line, $x = y$.

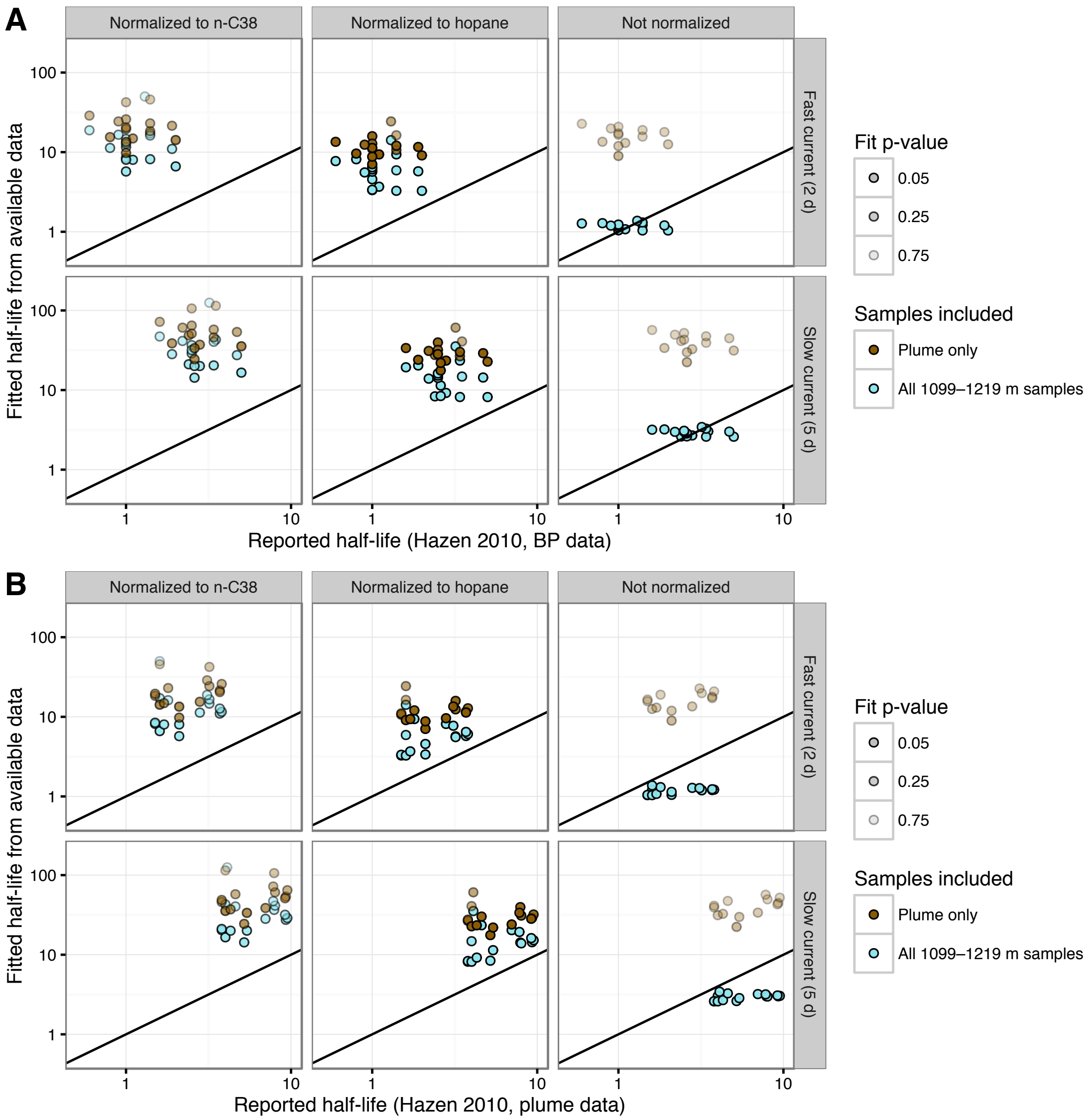

Figure S12: Comparison of fitted half-lives from the present re-analysis of water-column data and from (*6*), for (A) the Hazen "BP" dataset and (B) the Hazen "plume" dataset. Data were fitted after normalizing to *n*-C38 concentration; after normalizing to hopane concentration; or unnormalized. Datasets were fitted with samples from plume depth only (brown) or all water-column samples matching samples used by Hazen et al. (blue). Following Hazen et al., we consider both fast-current and slow-current models. Symbol opacity indicates fit quality. Fits with $p \geq 0.9$ are not shown. Black lines, $x = y$.

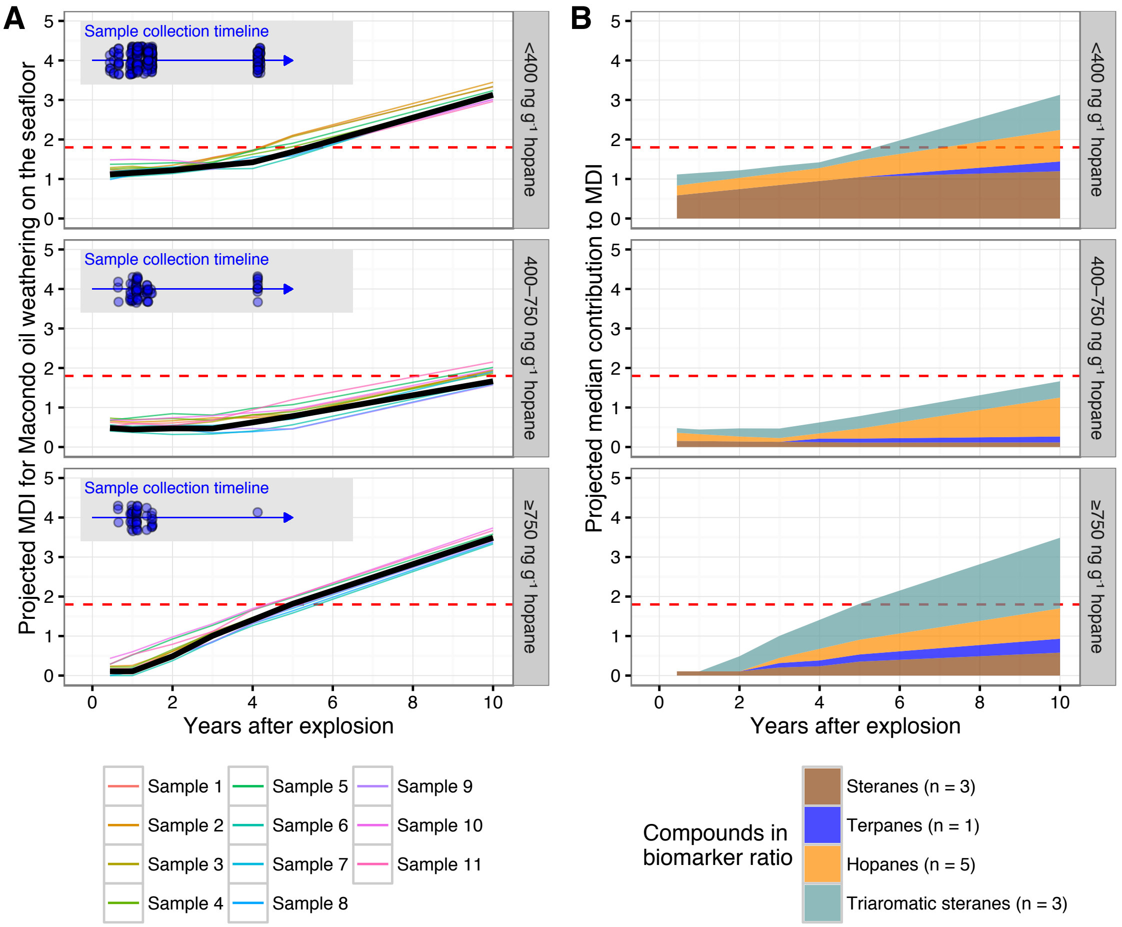

Figure S13: Projected changes in Macondo dissimilarity index for source oil weathering on the seafloor. (A) Projected MDI with uncertainty estimate. Heavy black line, median projection; thin colored lines, projections for the 11 source-oil samples. Inset: timeline of collection of samples (blue points) in each contamination bin, showing that the latest samples were collected before the MDI crosses the threshold value of 1.8 (dashed red line). Points are jittered vertically to minimize overplotting. (B) Projected MDI contributions of the different classes of biomarkers represented in the 12 fingerprint ratios. Median penalty scores accruing to the ratios in each class of compounds (see Table S2) were summed to calculate projected median MDI.

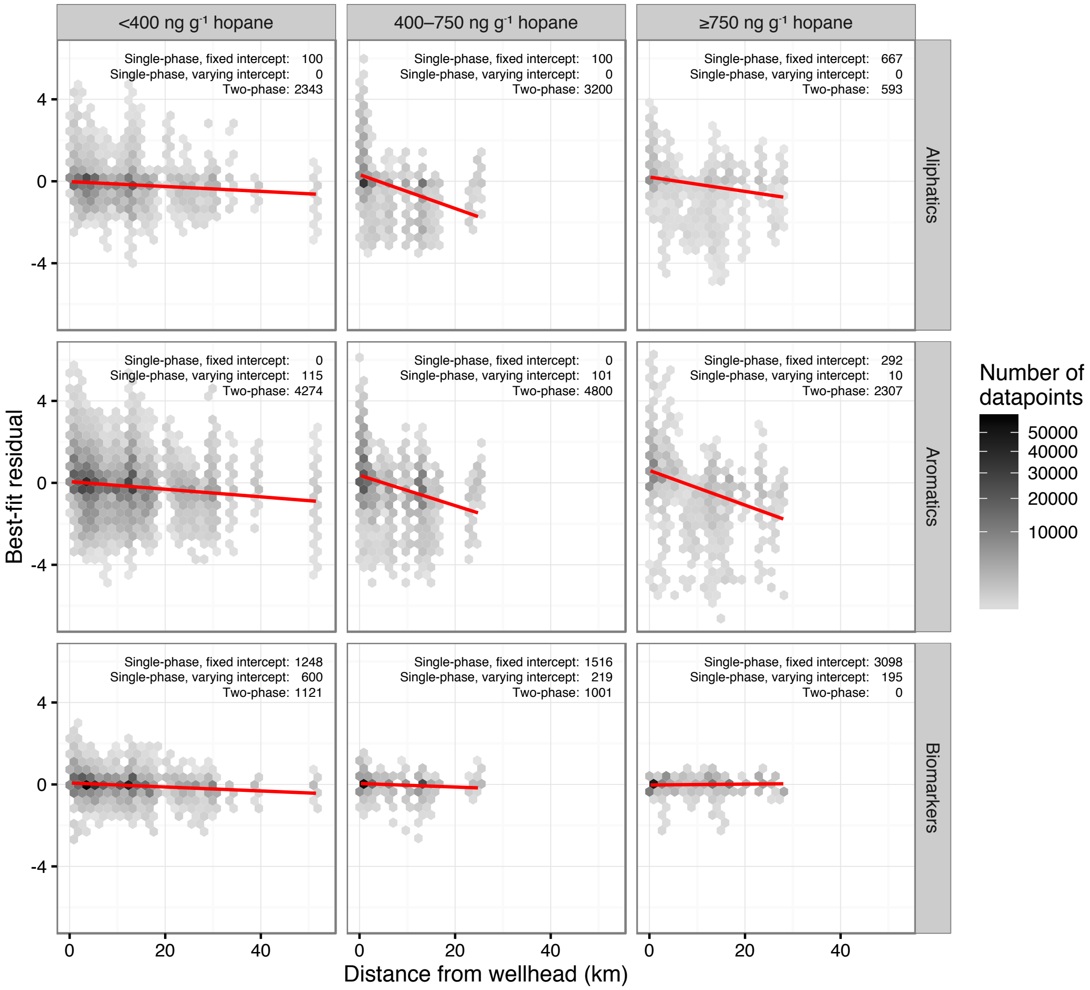

Figure S14: Relationships between model residuals and sample distance from the Macondo well. All residuals from pseudoreplicates with at least moderate support ($\Delta\text{BIC} \geq 2$) for the best-fit model are plotted against the distance of the sampling site from the wellhead. Datapoints are too numerous to plot individually; the color of each hexagonal cell indicates the count of datapoints in that region. The numbers of pseudoreplicates best fit by the three models are shown at the upper right of each panel. Red lines show least-squares linear regression fits.

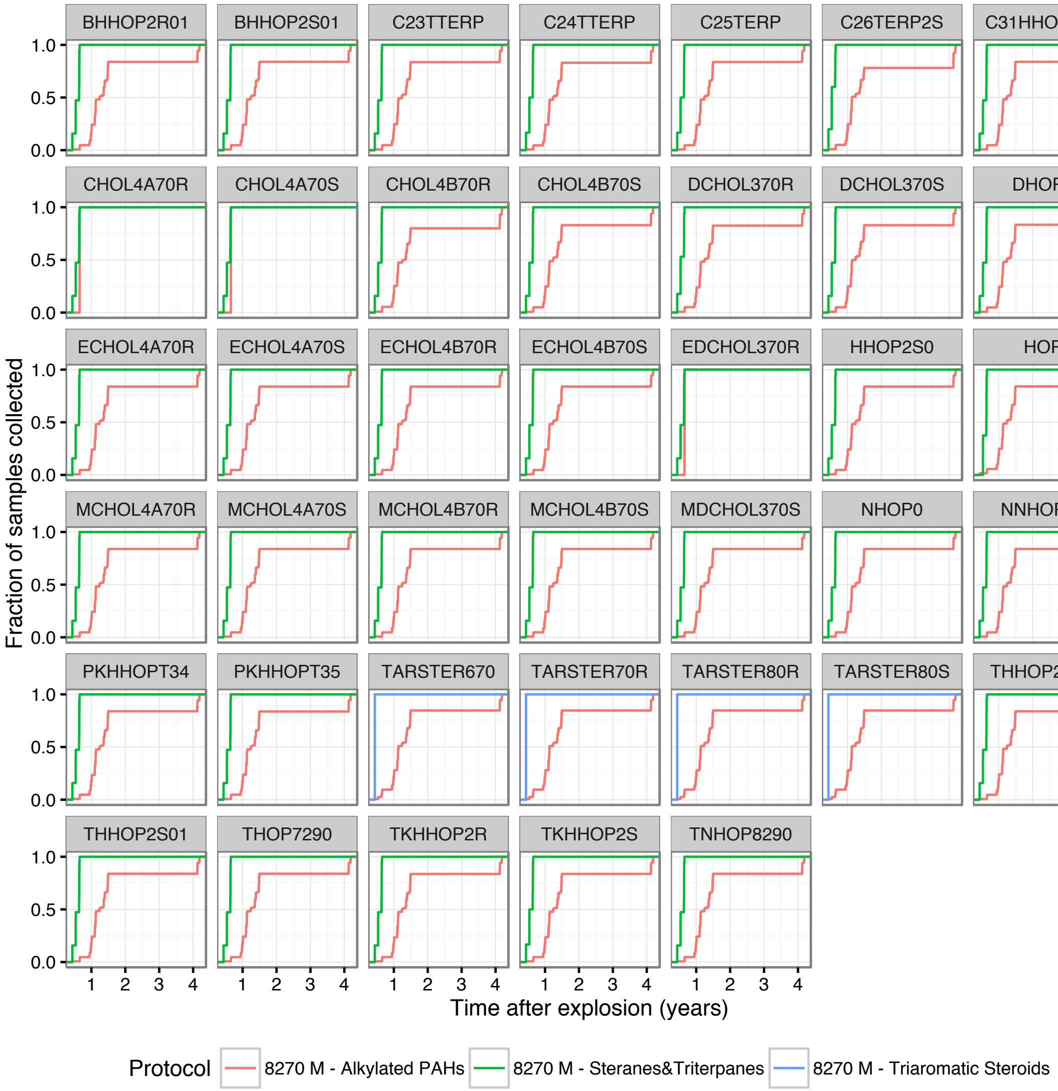

Figure S15: Empirical cumulative distribution functions showing the stage of the response effort at which biomarker concentrations were measured with different lab protocols. EPA protocols “8270 M - Steranes & Triterpanes” and “8270 M - Triaromatic Steroids” were used almost exclusively until 238 d post-explosion; thereafter, only “8270 M - Alkylated PAHs” was used. The compounds shown here are the only compounds for which the dataset analyzed derived from multiple protocols.

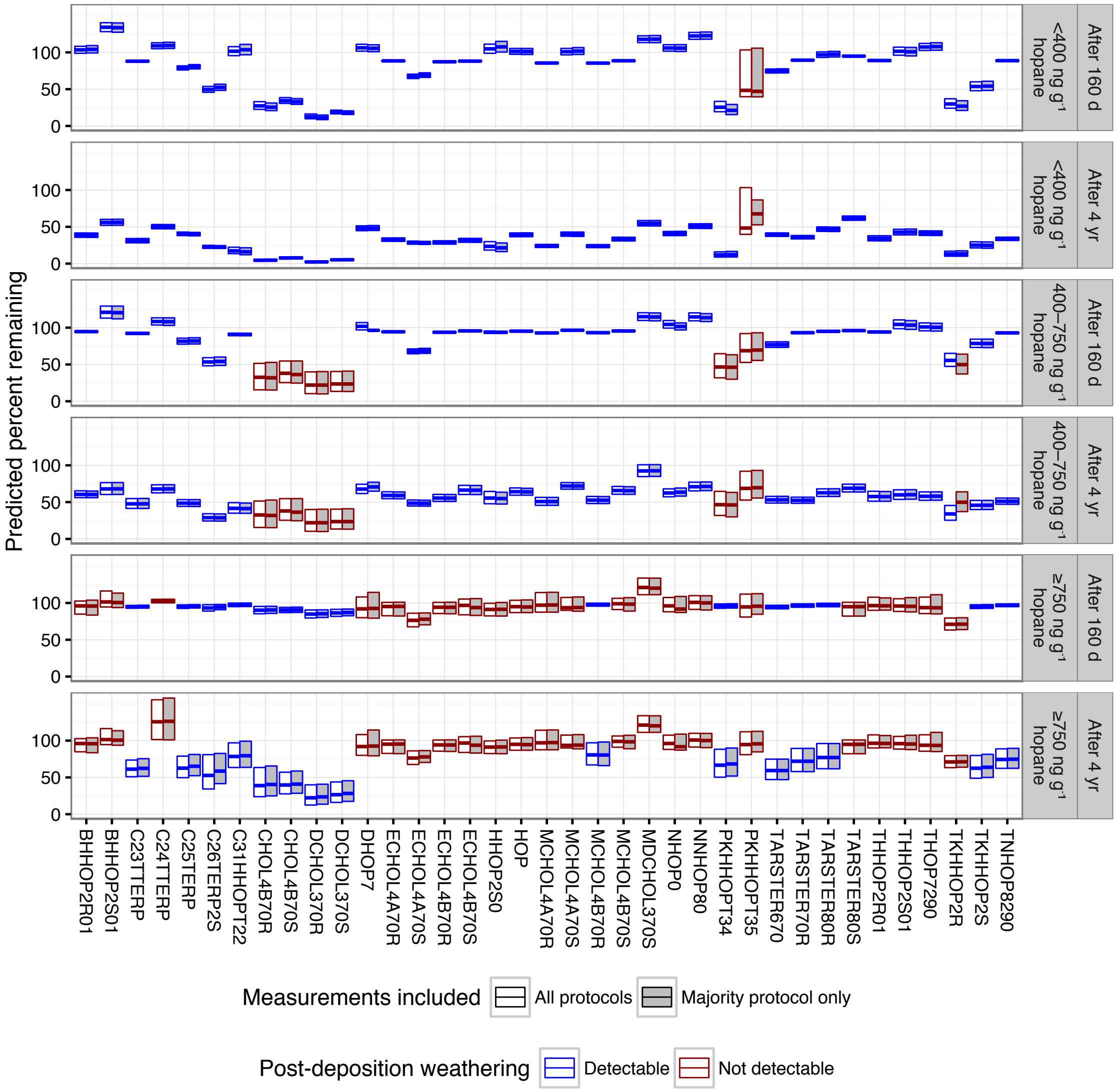

Figure S16: Comparison of biodegradation analysis results when the working dataset includes data from all lab protocols (open symbols, at left in each pair) or only data from the most commonly used protocol (gray symbols, at right in each pair), showing predicted percent remaining at 160 days and 4 years post-explosion for each contamination bin. As in Figs. 3–6 and S7–S10, symbols are outlined in blue wherever post-deposition weathering was detectable and in dark red where it was not.

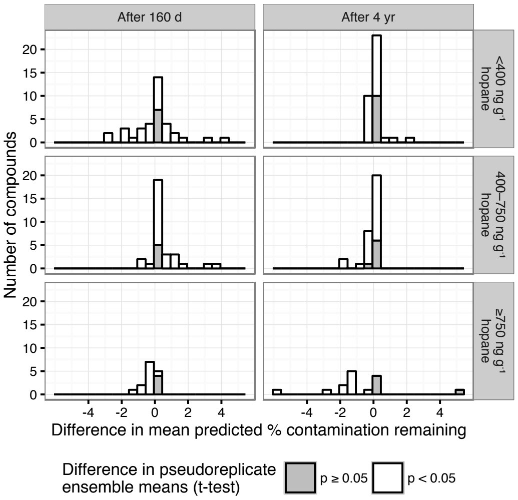

Figure S17: Distribution of differences in biodegradation predictions when the working dataset includes data from all lab protocols or only data from the most commonly used protocol. T-tests were used to compare the predicted percent remaining at 160 days and 4 years post-explosion for each compound-contamination bin set of pseudoreplicates under the two conditions. While a majority of comparisons showed significant differences ($p < 0.05$; white bars), these differences were consistently very small.

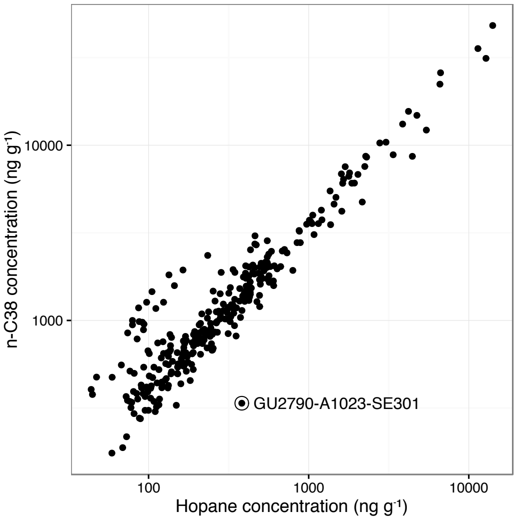

Figure S18: n-C38 concentration vs. hopane concentration among surficial sediment samples identified by MDI as contaminated with Macondo oil. The circled point, representing a sample collected at a much greater distance from the wellhead than the next farthest sample, was excluded from further analysis.

References

(*1*) Wang, Z., Yang, C., Fingas, M., Hollebone, B., Yim, U. H., and Oh, J. R., In *Oil Spill Environmental Forensics*, Wang, Z., and Stout, S. A., Eds.; Elsevier: 2007, pp 73–146.

(*2*) Aeppli, C., Nelson, R. K., Radović, J. R., Carmichael, C. A., Valentine, D. L., and Reddy, C. M., (2014). Recalcitrance and degradation of petroleum biomarkers upon abiotic and biotic natural weathering of *Deepwater Horizon* oil. *Environ Sci Technol 48*, 6726–6734.

(*3*) Valentine, D. L., Fisher, G. B., Bagby, S. C., Nelson, R. K., Reddy, C. M., Sylva, S. P., et al. (2014). Fallout plume of submerged oil from *Deepwater Horizon. Proc Natl Acad Sci 111*, 15906–15911.

(*4*) Aeppli, C., Reddy, C. M., Nelson, R. K., Kellermann, M. Y., and Valentine, D. L., (2013). Recurrent oil sheens at the *Deepwater Horizon* disaster site fingerprinted with synthetic hydrocarbon drilling fluids. *Environ Sci Technol 47*, 8211–8219.

(*5*) Stout, S. A., and Payne, J. R., (2016). Macondo oil in deep-sea sediments: Part 1—Sub-sea weathering of oil deposited on the seafloor. *Marine Pollution Bulletin 111*, 365–380.

(*6*) Hazen, T. C., Dubinsky, E. A., DeSantis, T. Z., Andersen, G. L., Piceno, Y. M., Singh, N., et al. (2010). Deep-sea oil plume enriches indigenous oil-degrading bacteria. *Science 330*, 204–208.