Executive Summary

This 17-page (+appendix) FracFocus analysis almost requires an executive summary. With more than 250 hyperlinks to database queries, this document is densely packed with numbers and calculations. At times, cleaning and curating the FracFocus data will appear daunting...

Data Highlights

1. The downloadable database on the FracFocus website has 60% of the disclosures. The only way to get the other 43,562 disclosures is to download and scrape every PDF. There are also duplicate disclosure records and duplicate sections within single disclosures that need to be removed.
2. We believe total water volume used in the fracs is misreported in ~7% of frac disclosures.
3. Using Colorado and a small sample of treatment summaries from the state site, we extrapolate that ~3% of frac jobs are not reported to FracFocus. Using the same CO dataset, proppant is misreported ~20% of the time.
4. ~11% of disclosures have a sum of total chemicals that add to less than 97% or more than 103%.
5. ~33% of all chemical rows have no defined purpose.
6. ~16% of ingredient rows are trade secrets, 48% of which list the actual ingredient name for the trade secret.
7. ~60% of unique CAS numbers used in FracFocus cannot be matched to the EPA database of 496K CAS numbers.
8. We cannot calculate total frac mass in 8% of the disclosures.
9. Using Texas RRC for comparison, ~6% of the geolocations reported in FracFocus are off by 1 mile or more.
10. Chesapeake misreports 60% of its chemical ingredients by adding leading zeros to the CAS numbers. These rows are not findable from the FracFocus website. Chesapeake marks 25% of their ingredients as secret.
11. On average, ~.5% of each frac is toxic chemicals.
12. Many chemicals have outlier rows. For example, 5 out of 87K methanol rows account for 29% of the total methanol used across 87K fracs.
13. ~49% of all chemical rows are marked as incomplete. ~32% of the 49% have a CAS number withheld (e.g. trade secret). 8% of frac headers are marked as suspicious (water problems).

Cleaning and curating is very achievable. We've mounted one of the largest efforts ever undertaken to clean and curate FracFocus data. We do more than 37MM lookups against the database. We append 9,843,607 data values, some corrections, some clarifications. We insert another 27,874,643 values that allow us to do calculations (see table on page 3). It's all processed by computers, we log every appended field and make it simple for any user to suggest/make a change to any of the 15,000+ rules. Just click on the field in question, view the log row and user has access to the exact spot in the rule sheet where changes can be made.

Cleansing and Curation Highlights

1. CAS number is fixed and updated in ~13% of the chemical rows.
2. ~48% of all trade secret (withheld) CAS numbers are solved using filled out ingredient names.
3. ~68% of all chemical rows get a corrected or preferred ingredient name change.
4. Purpose is fixed, updated and/or consolidated to standard names in 75% of the chemical rows. We fix 99% of the chemical rows where purpose is a comma-delimited list of multiple purposes.
5. Where purpose cannot be identified/fixed, we calculate the most probable purpose. ~9% of chemical rows have an unidentifiable purpose and we add calculated purpose to ~12% of the 9%.
6. Synonyms/Initials/Abbreviations for supplier are updated/consolidated for 32% of the rows.
7. We assign a key supplier to ~33% of the chemical rows, helping to identify which tradenames/chemicals are exclusive to a single supplier. This is used in pressure pumper identification.
8. We track 56 pressure pumping companies and identify which company pumped the frac in 50% of the disclosures. We use water, sand and exclusive chemicals to identify pumper. We estimate that Halliburton pumped ~18% of the fracs, Baker Hughes, ~9% and Schlumberger, ~8%.
9. 250 different ways to identify proppant, encompassing 160K rows, are consolidated into a single proppant quantity for each disclosure.

10. Combining the EPA AcTOR database with FracFocus, we track 381 carcinogenic chemicals representing 11.6 billion pounds, or .43% of the 2.6 trillion pounds of total frac chemicals. In addition, we track genetic, chronic, reproductive, developmental and hazardous toxicities for ~64% of chemicals used in frac jobs.
11. We detect 1.4MM EPA chemical name (synonyms) and add preferred names to frac database where applicable.
12. We have molecular weight, chemical formula and other chemistry for 64% of the chemicals used in fracs. We identify the density (lbs/gal) for 457 CAS numbers, representing 65% of the total chemical rows.
13. Outlier rows (extreme/misreported percentages) are suppressed before any detailed analysis is done.
14. Where mass is (optionally) entered, ~60% of all disclosures, each mass is compared to HF Fluid percentage and total water volume to determine accuracy.

There are calculated fields we add to the database (e.g. MassZscore833) that will not be discussed as part of this document. These fields (most normalized) are used to calculate the effectiveness of each chemical.

At the end of this document, after the recommendation section, there are 5 appendices: flowback analysis, working with FracFocus SQL data, chemical cost analysis and several pages of statistics, including tables and charts.

The Data

A quick note about the data used in this analysis. The FracFocus website offers a downloadable database with limitations. Of the 119,223 disclosures in the database, 43,871 (36%) are not present in the downloadable database. The database does have 119,223 headers, but not all headers have purposes and ingredients. To retrieve all disclosures, the PDFs need to be manually downloaded and scraped. We have this capability and all 119K disclosures, but decided to limit our analysis only to the 75,352 complete disclosure records, which one can use to validate this analysis.

To write this document, thousands of rules had to be developed to clean/curate the FracFocus database, so it can be properly analyzed. We log each data fix and calculated field we add. We currently do 38,872,627 lookups for appends (cleansing) and additions (curation) in the FracFocus database. All original data is left untouched. We add (n)fields to record the changes. Below are some descriptions and statistics on the rules we wrote and continue to add to…

The Rules

Normalized (n)CAS (2,316 rules):
As an example, there are 21,523 rows where CAS does not identify water. Frac Specialists may use 732-18-5. EES may use 773-18-2. Someone else may list it as 1732-18-5. Schlumberger may not provide a CAS at all and instead use NA. Still another may just list it as 0000-00-0. We correct for mistakes and enter the number 7732-18-5 in a new field called nCAS. The above 21,523 includes 11,886 that list water as CAS #007732-18-5. And this is just for a single ingredient, water. We do this for every chemical.

An additional 1,050 rules find CAS based on the ingredient name only. We load more than 17,500 ingredient synonyms (EPA sourced) for 1,050 CAS numbers to solve for CAS based on name only. We update 112,340 “confidential” CAS number chemicals with the correct CAS based on ingredient name alone. We solve for 216,400 withheld CAS numbers. We add a nCAS that is different from the reported CAS 344,951 times.

(n)Ingredients (2,485 rules based on CAS; 969 rules based on ingredient name):
Note: We always leave the original data intact. This is the reason we create n-fields. Most (if not all) the ongoing analysis will be done using n-fields. In the case of ingredients, we update 1,879,433 ingredient names with standard names. Some of this is splitting hairs. The EPA may have a preferred name and the operator may enter a synonym. But like everything else in this analysis, it’s nice to be able to retrieve all “water” from the database with a simple query.

Clean	
nCAS	344,951
nPurpose	2,084,690
nSupplier	908,659
nIngredients	1,879,433
nFluid	2,453,000
Calc_Purpose	2,172,874
Curate	
Frac_Type	4,913
Ingred833lbs	2,322,062
KeySupplier	925,141
MassAvg833	2,278,534
MassProb833	2,266,255
MassZscore833	2,278,509
nAdditive	383,999
nFluidAvg	2,277,065
nFluidprob	2,261,620
nFluidsigma	2,277,065
nFluidzscore	2,277,032
nIngredienttype	228,663
Pumper	74,823
xDensitylbgal	1,781,692
xIngred833volgal	2,322,062
xState	2,195,470
xSuspicious	1,364,242
Header	
Proppant833lbs	66,720
TotalFluidgal	69,138
TotalWaterperc	69,138
xhSuspicious	6,213
xTotal833lbs	69,138
xWater833lbs	75,149

(n)Purpose (1,038 rules):
A single purpose may have multiple ingredients. For example, an acid treatment may have a water component. Should we reclassify water as base carrier fluid? We have 18,780 examples where the ingredient is water, but has been classified as an acid. We have 2,803 different purpose names for ingredients that contain the term acid. And even if we don't reclassify chemicals, there are often multiple ways to state the same purpose. For example, we assign nPurpose to 327,892 chemicals where the operator decided to use a comma-delimited list. We add nPurpose to a total of 2,084,690 chemical rows. Some updates are just to consolidate purpose names. For example, the purpose Water would be changed to Base Carrier Fluid for consistency.

Calculated Purpose (1,805 rules):
<Here>, Marathon Oil decided not to list specific purposes and just grouped them all into a comma-delimited list. Marathon decided that all chemicals supplied by SLB only need one purpose line. Using the CAS number and/or the ingredient name, we backfill in the purpose for each line. For example, 14808-60-7 is a proppant. We calculate that Boronatrocalcite is mainly used as a crosslinker. If you disagree, the Boronatrocalcite rule can be updated within seconds.

What is the difference between nPurpose and Calc_Purpose? We identify 1,303 different purposes operators use to define proppant. We want to group these together as a single (n)purpose, Propping Agent. Calculating a purpose tries to determine if the operator entered the correct purpose (or left in blank) on the disclosure. We calculate purposes for 2,172,874 chemical rows, even though purpose may be correct in first place.

(n)Supplier (858 rules):
There are 18 ways Baker Hughes is listed as operator. Is the majority just Baker Hughes? Of course, but we still wanted to clean up those other 17 versions. There is a secondary reason for nSupplier. Remember we cut our teeth in pressure pumping, so being able to do a pumper market share analysis was extremely valuable. At the time, our calculations were looked at very closely. While supplier name fixes are secondary to other changes, we do it 908,659 times.

Key Supplier (376 rules):
Schlumberger has provided CAS #61789-77-3, 3,980 times in fracs. 98.3% of the time this CAS is used, SLB is the one providing it. We will list SLB as the "key supplier" of this chemical. Why? It helps us determine who pumped the job. If we cannot tell the pumper by sand or water supplier, we will use an "exclusive" ingredient (or set of) to help us take an educated guess that SLB pumped the frac. This allows for more accurate pumper market share calculations.

Pumper (286 rules):
We identified 56 different companies that offer hydraulic fracturing services. We include a field that searches for 285 spelling variations of the 56 pumping companies and mark 74,823 ingredients coming from 1 of these 56 as pumper. 8,051 of the 74,823 (10.7%) are spelling variations of the pumper's name. If we see that Baker provided 60% of all chemicals for a frac, we will mark the overall job as pressure pumped by Baker Hughes.

(n)IngredientType (208 rules):
We all know the two most important ingredients are water and proppant. But both can appear in a single disclosure multiple times. For example, here is a frac disclosure that lists water 11 times. One engineer will take the 90.38% (main) water row and use it with the 1,716,120-gallon water volume to determine overall liquid, then mass, in the frac. Another engineer will want to use all 11 lines of water (91.53%) before calculating overall volumes. Bottom line? We identify the "main" water line in every frac. Almost all submitters enter total water volume that matches the main water row.

Proppant is a bit more straight-forward. Engineers want total proppant, whether brown sand, white sand, SiO2, Silicon Dioxide, etc. We identify all proppant in a frac, so it can be used as a single number. We found 254 distinct ways operators label proppant. We mark 160,490 proppant rows and 68,173 main water rows (reminder – we are only using the downloadable database for this document).

(n)Fluid
The maximum ingredient concentration in HF fluid column should add to 100% in most cases (can be < 100%). 47,845 of 68,762 disclosures (69.5%) add to 100%. 1,336 add to 1%. 13,470 add to more than 100%, 4,700 more than 110%. 1,680 add to less than 10%. We try to correct those we can.

xState (2,227 rules):
We are leaving the cleaning (n)fields and starting to get into data curation. Here we lookup the state of each chemical at room temperature (gas, liquid or solid) from the EPA and other sources. We currently add state to 2,195,470 chemical rows. This number will increase over time as we track down more CAS states at room temperature.

xSuspicious (520 rules):
We wanted a way to mark any bad data in the disclosures. We mark 1,364,242 rows in the detail portion of the disclosures as suspicious. Many have more than one problem. Some may have a CAS listed as a trade secret or confidential. Errors can be a bad CAS, missing CAS, CAS entered as a word, leading zeros added to CAS. It can be purpose not disclosed or multiple purposes listed for a single chemical. This is basically a "don't shoot the messenger" field and we would rather mark everything that is not perfect in the database. We also identify 6,213 issues in the frac header, most are water related.

Other fields:
We have a few more boutique rules like identifying whether a frac is Oil, Propane, Nitrogen or Acidizing only. Then we have the calculation fields, massavg833, masssigma833, nfluidavg, nfluidsigma, xdensitylbgal, proppant833lbs, xtotal833lbs, xwater833lbs, nfluidzscore and others that are well beyond the scope of this document. But the above theme of '833" will be discussed below. As of 10/30/2016, we have 15,693 total rules in the Frac Engine Software.

GUTCHECK: In some of the screencaps, you may see data corrections where you disagree. For example, is KCL Water potassium chloride or water? We have zero skin in the game. Changes take seconds to make. New rules are easy to add. Within any row in the database a user can click the field, see the rule that applied the change and edit the rule.

Disclosures

How do we know if every frac job has been submitted to FracFocus? Colorado may help us answer that question (directionally). Colorado tracks treatment data on the state site. Colorado has recorded 89,638 treatments, 5,325 of which are identified as fracture stimulations. How many of these are in FracFocus? First thing to consider is Colorado only started requiring disclosure to FracFocus on 4/1/2012. Interestingly, all 5,325 meet this treatment label and we believe this label was added to coincide with Colorado's requirement to file FracFocus disclosures.

To analyze Colorado against FracFocus, we need build full 14 digit APIs from Colorado data. For example, treatment API# 05-123-39173 is 05-123-39173-02-00 in FracFocus because the treatment was applied to sidetrack #2.

API# 05-017-07793-0000 is an example where the operator (on 12/10/2015) reported "Fracture Stimulation" on the CO state site, but there is no corresponding FracFocus record. Here is an example where Noble Energy reported a sidetrack 02 treatment as a sidetrack 00 on FracFocus. Technically, this is not the API that was treated.

We have 203 (3.8%) CO treatment summaries, labeled Fracture Stimulation, that do not have a corresponding 14 digit FracFocus API. One could argue that Colorado labels both acidizing and frac jobs with the same treatment summary. If we require proppant to be present in the fracture stimulation, we still get 139 (2.6%). If we allow for 00 and 02 sidetracks to be the same API, we drop to 57.

Let's try a different tactic. We know that Colorado reports some basic FracFocus information.

If we do a query on the state site, we get 10,265 cases where Colorado has recorded FracFocus metadata. Interestingly, FracFocus has 11,507 disclosures (11,429 in our slightly dated database). The Colorado state site is off by 1,164 (11,429 – 10,265) disclosures. <Here> is a breakdown showing even back in 2014-15, not all FracFocus disclosures were being recorded on Colorado State site. This tells us CO is not automatically connecting to FracFocus site to retrieve metadata.

Colorado Treatment Summaries

Even more interesting… Colorado will list ingredient amounts on the state site both in fields and in the remark section. Once again, Colorado is going to let us "second source" (to borrow a phrase) the quality of data submitted to FracFocus. It's worth mentioning that the Colorado treatment section is not consistent. We have 2,075 FracFocus disclosures where Colorado has captured no treatment information on the state site. API # 05-123-38120 is an example where even the Colorado site mentions a FracFocus record, but the CO treatment summary has no data.

We have 5,364 FracFocus disclosures where Colorado also has treatment data. We will start with a few examples and walk-through them slowly. We are going to pick on a wellbore (05-067-07023) frac'd by BP America in June 2016. The FracFocus record (again, one that incorrectly separates purposes from ingredients) shows 162,817 gallons of water. Using elementary school math, we know how to compute the total volume in gallons…

	Water		*Total Volume*
Percents	80.17061%	=	100%
Gallons	162,817		X

We know the total volume of the entire frac is 203,088.14 gallons. We multiple that by density of water (8.33) and we get a total mass for the frac of 1,691,724.20 pounds. Turning to the treatment summary, we see 3,941 barrels of water were used. 3,941 barrels * 42 gallons/barrel gives us 165,522 gallons of water. Excellent. The treatment summary is only off the FracFocus record by 2,705 gallons of water (165,522 – 162,817). Please note, we are only looking at disclosures where the CO treatment date is exactly the same as the FracFocus date. Let's next check the proppant.

	Proppant		*Total Mass*
Percents	17.94982%	=	100%
Pounds	X		1,691,724.20

The total proppant is (x=) 303,661.45 pounds. This, too, is very close to the 298,000 reported on CO state site. Before we leave this example, let's go ahead and solve for HCL.

	HCL		*Total Mass*
Percents	0.84969%	=	100%
Pounds	X		1,691,724.20

We get 14,374.41 pounds. Here is where it gets interesting. Do we divide pounds by 8.33, the density of water, or more around 13.67, the density of HCL? If we divide by 8.33 we get 1,725.62 gallons and that divided by 42 gallons per barrel, gives us 41.09 barrels. Colorado site tells us 64 BBLs were used. FracFocus does not agree with CO treatment summary. Using a HCL density of 13.67 reduces the barrels to 25 on FracFocus compared to 64 on Colorado.

<This> is what the data looks like in our database. We make sure the frac job start date and CO treatment date are the same. We make sure the APIs from both sources are the same. From the FracFocus side, we bring in total water volume and the proppant calculated in pounds.

Second Source

This next part may cause some of you to totally dismiss this analysis. I get it. If there's a better way to do a checks and balances on operator disclosures, I am all ears. But we are fortunate to have a Colorado treatment summary table that may talk a slightly different language than FracFocus, but can give us a few (directional) insights.

First, we have 15,164 frac disclosures for Colorado. 4,827 of these have a match on (full) API to Colorado treatment summary. This number drops to 2,626 when we require the CO treatment date to match the FracFocus job start date. Of these treatment summaries, 1,196 have a water number, 1,976 have a proppant number and 878 have an acid number.

Proppant

We should be able to compare the proppant reported to FracFocus with the proppant reported on the Colorado Treatment Summary. For example, API# 05-005-07221 reported 10,698,380 pounds of sand to Colorado.

$$[\text{Water Volume} * 100 / \text{Water Percent}] * 8.33 * (\text{Proppant Percent} / 100)$$

Using the above formula, FracFocus reports 10,558,002 gallons of water and the water percent is 87.33470. Therefore, 10,558,002 * 100 / 87.33470 *8.33 * (10.1455 + 1.07111) / 100 = 11,295,398 pounds of proppant. 11,295,398 - 10,698,380 = 597,018 pounds that were reported on FracFocus and not Colorado state site. Here is just the opposite case where Colorado reported 3,751,551 pounds of sand used in the 11/12/2013 frac job and FracFocus reported only 1,344,562 pounds (2,917,367 * 100 / 94.55250 * 8.33 * 5.23140 / 100). We are off by over 2 million pounds of sand.

These are just two examples. What operators will do is include multiple treatments within 30 days into a single FracFocus disclosure. For example, API # 05-001-09529 has 3 formations that were stimulated.

co_full_api	treatmenttype	treatmentsummary	treatmentdate	totalproppantlbs
05001095290000	FRACTURE STIMULATION	Frac'd the Codell with 4...	2013-02-28	154100
05001095290000	FRACTURE STIMULATION	1 stage: Frac'd with 177...	2013-02-01	105060
05001095290000	FRACTURE STIMULATION	Frac'd the J-Sand with 5...	2013-02-28	154160

But the FracFocus disclosure only shows a single disclosure that started on 2/1 and ended 2/28. If we only compare the proppant from 2/1 treatment, it would be 105,060 pounds versus the 431,775 reported on the disclosure. Summing the 3 treatment rows gets is to within 18K pounds of the FracFocus number.

Of the 1,976 treatment summaries with proppant, 166 FracFocus disclosures underreport the proppant by more than 100,000 pounds. At 50,000 pounds, the number increases to 219. Overreported by 100,000 pounds is 265 and 50,000 pounds is 347. More detailed analysis needs to be done, but it looks like 431 (431/1976 = 21.8%) treatments have more than a 100,000 pound disagreement with FracFocus. Let's look at one more example…

API# 05-123-36245 reports 55,444,003 pounds of proppant was used in a 6/1/2013 stimulation. FracFocus reports 5,545,026 pounds of proppant (7,117,173 * 100 / 91.04444 * 8.33 * (8.50646 + 0.00893) / 100). Not only is this off by 49MM pounds, but if we take the FracFocus pounds and multiple by 10 (5,545,026 * 10 = 55,450,260), we are now within 6,300 pounds (rounding error) of the treatment summary. In other words, the operator underreported proppant by exactly 10x. In other words, s/he put the comma in the wrong place.

Filing the Disclosure

Schlumberger supplies a total of 255,019 ingredients. Of these, 230,908 (90.55%) have a comma in the (multi-) purpose. When we look at a Schlumberger supplied job for BHP Billiton Petroleum, Marathon Oil, Whiting Petroleum, and Chesapeake Operating, Inc., we see a trend. They are all reported incorrectly (purpose with comma and ingredients separated from purposes), but for different operators and in different states.

Did SLB, as part of its pressure pumping services, agree to submit the disclosure on behalf of each operator? Or did SLB supply the information to the operator and the operator submitted the disclosure exactly how it was supplied from SLB. I suspect SLB handled the disclosure paperwork with regulators. Did they also indemnify the operator and take full responsibility for the accuracy of the disclosure? And the above sample disclosures are from 2016, so we can't blame the early days of FracFocus. And Schlumberger recently updated their template. It still lists purposes separate from ingredients, but breaks out the purposes into different sections.

Schlumberger is not alone in this style of reporting. CWS (we rightfully assume this is Calfrac Well Services) has 88% of its ingredients reported with purposes separated from ingredients. The similarities are significant.

Halliburton has its own "signature" approach to disclosures. We estimate that HAL has pumped ~13,133 frac jobs. Of this number, 3,179 list most ingredients under the single purpose, "Hazardous and Non-Hazardous Ingredients." For example, in API# 04-030-03279 lists almost all the ingredients in a section marked as not having MSDS reports. Here is the MSDS for CAS# 1327-36-2. And note the 5% KCL Water row has been repeated in both sections, with different %HF Fluid amounts. Baker Hughes appears to list the ingredients with associated purpose on a regular basis. And note, I am not picking on old frac disclosures. API# 42-483-32460 was pumped one month ago.

Ingredients

Leading Zeros

Probably one of the most interesting and obvious alterations to the disclosures is the inclusion of leading zeros (mentioned at very top of this analysis) in CAS numbers. We have 82,183 supplier rows with leading zeros. 25,511 of those are supplied by Performance Technologies. The EPA ACToR (toxicology) database has 456,817 unique CAS numbers. Not one of them starts with a zero.

But a much more interesting leading zero query is the operator, not the supplier. Chesapeake Operating has 76,021 chemical rows (60.2% of all CHK rows) where CAS starts with zero. #2 on the list is Athlon Energy with 590 rows. Big deal? You decide. If we search on water at Chesapeake Operating, we get 6,319 disclosures. But there are another 3,067 (32.6%) that list water as 00732-18-5. One might think the leading zeros are coming from an (enterprise) application at CHK. But sometimes CHK enters water correctly, sometimes they don't.

Format and Trade Secrets

CAS numbers follow a specific format, like APIs. Of the 496K EPA ACToR CAS numbers, not one breaks from this format. FracFocus has 560,117 (20.2%) rows that do not follow the guidelines. 442,110 (78.9%) of these have CAS as confidential or trade secret or proprietary (withheld). 56,465 unique disclosures have 1 or more withheld ingredient rows.

CAS Numbers

Operators have submitted 2,488 unique CAS numbers to the FracFocus database. Of the 2,759,313 ingredients in the database, we matched 2,433,395 to a normalized CAS number. There are 967 distinct nCAS numbers in the database, 204 of these are used less than 10 times. 2,488 – 967 means there are 1,521 CAS numbers used in FracFocus we cannot match with a rules engine. These 1,521 CAS numbers account for the remaining 325,915 (11.8%) rows (2,759,313 - 2,433,395). 285, 978 (87.7%) of these 325,915 are confidential, NA or trade secrets that we could not match to a nCAS.

No %HF Fluid

24,444 FracFocus disclosures list one or more ingredient with 0 or an empty %HF Fluid field. Here is an example frac where we think the %HF Fluid was entered in the wrong column. API# 30-025-42639 is a Chevron example from February of this year that has no %HF Fluid amounts listed. It's surprising that a disclosure submitted to FF 3.0 does not fail submission guidelines when all %HF Fluids are empty.

44,452 chemical rows have an additive amount (% of purpose section) but no %HF Fluid. We take an educated guess that ~23,225 of these are cases where the %HF Fluid is entered into the wrong column (as example APIs above show). This 44K is in addition to the 70,448 rows that just don't have any masses at all (either column). These ingredients are just plugged into the disclosure, but have no impact on the frac.

HF Fluid Sum not 100%
Starting with FracFocus 2.0, disclosure authors were alerted when the sum of %HFFluid was more than 3% away from 100%. We have 65,420 that fall within this +-3% range. 52,164 would be considered right at 100% (allowance for rounding error). This leaves 8,389 (11.3%) that fall outside the acceptable FracFocus ranges. These 8,389 disclosures account for 111,256 ingredients (HF sum out of range) where total pounds could not be calculated.

Other
>> We have cases where purpose sections are duplicated in their entirety. API# 03-145-11543 has 17 "Liquid Tracer" sections, all with the exact same ingredient information listed.

>> Looks like 1,035 FracFocus disclosures are duplicate records. API# 35-081-24198 has 10 disclosures for the same date. There is a slight difference in water reported on each of the 10 records. We cannot find a single difference with these two disclosures and they were filled a month ago. These are not all small, boutique operators either. Devon leads the list with 86 disclosures that have more than 1 duplicate.

>> We have 928,763 ingredient rows without a defined purpose. 927,052 (99.8%)of these purpose only rows have a %HFFluid number. We have 328,123 (35.3%) with a comma-delimited purpose. We deduce an nPurpose for 724,971 of the 928,763 and calculate the purpose for another 23,328 without nPurpose.

>> We have 10,911 chemical rows where the ingredient name contains "listed below" but an amount for %HFFluid. Let's look at 2 examples. The first is from Devon Energy in 2015.

Baker Hughes	Proppant				
		MSDS and Non-MSDS Ingredients Listed Below	N/A		20.45471
Baker Hughes	Acidizing				

Baker Hughes says see below for proppant ingredient, but lists 20.45471%. Below is lists Crystalline Silica (Quartz) also with a %HFFluid amount. The %HFFluid is fill out in both places. Of course, the HF Sum is off, in this case by 21.95%.

Here is another case from Encana in 2014 where the same behavior is present. In this case, the combination of listed below and actual ingredients equals 100%. In other words, all the %HFFluids are right, but the listed below is referring to chemicals that are not listed below.

Location

We have 1,105,286 geolocations for Texas wellbores. 33,394 FracFocus disclosures have Texas RRC geocoordinates. For example, API# 42-255-33674-00-00 has three frac disclosures. When we look up this API on the RRC website (both current and historical), we get a single record (screencap here) for a Marathon operated well named 4H. We see that in FracFocus alone, there are 3 different geocoordinates for this same wellbore. Only 1, Boris 4H (Marathon) appears to match up with RRC records. Boris 4H is a .02 mile difference between FracFocus and Texas RRC, well within a rounding error. The other two FracFocus records, with the same API, are 11 and 9 miles away from Boris 4H. And this API has only had one Texas RRC permit ever pulled, for horizontal drilling.

Searching our RRC database, we can only find 1 well close to either of the 2 other disclosures discussed above. In this case, we have a pretty close match to McAlister 6H, but this is a well run by Encana Oil & Gas. It's also a horizontally permitted well, under the API# 42-255-32546, which has no FracFocus disclosure.

Of the 33,258 APIs that both have FracFocus and RRC geolocation data, 560 are more than 5 miles away from each other. 2,148 (6.4%) are more than 1 mile from each other. Let's look at another example, API# 42-479-42638, which has a 30 mile gap between FracFocus and Texas RRC. FracFocus has the well located next to an apartment complex in downtown Laredo. Texas RRC Cobol tapes have the well northeast of Hidalgo and west of 142. The Texas RRC digital map viewer has the well in the exact same location, west of 142. The distance between the actual well and the apartment complex is 30 miles (as the crow flies). The majority of wells, 29,829 (89%), are within a quarter mile of the Texas RRC location, but this leaves 10% that would be considered inaccurate.

Trade Secrets

There are 444,110 trade secrets (16%) in the FracFocus database. We solve for 217,310 of these. Shocking is that we have enough information to make an educated guess on 48.9% of trade secrets.

Quaternary ammonium salt	Confidential	10.00000	0.00031
Fatty acids, tall oil	Confidential	10.00000	0.00031
1-(Benzyl)quinolinium chloride	15619-48-4	10.00000	0.00031
Ethoxylated amine	Confidential	5.00000	0.00016

Operators will list the CAS# as a trade secret or confidential, then tell us the exact ingredient name. We have 8,233 rows of Fatty acids, tall oil with a withheld CAS#. If we query this ingredient over on EPA AcTOR, we see the CAS for this ingredient is 61790-12-3. And these confidential ingredients are being used in very small amounts. Of the 444,110 trade secret rows, only 1,025 (0.23%) are more than 1% of a frac's total mass. Here's an example where water is a confidential ingredient.

Purpose	Ingredients	Chemical Abstract Service Number (CAS #)	Maximum Ingredient Concentration in Additive (% by mass)**	Maximum Ingredient Concentration in HF Fluid (% by mass)**
Fluid				
	Water	confidential	100.00000	83.56700

Operators, already under scrutiny for invoking the trade secret label, need to be careful showing common ingredients (name field) marked as trade secrets. This will lead regulators to start questioning whether the label is being overused. Only 76,238 of the trade secret ingredients have an ingredient name that is also secret and even some of these have the term secret plus the name of the ingredient in the ingredient field.

Chesapeake marked CAS secret 32,679 times out of 126,291 total ingredients (25.8%). This is worth saying a second time… CHK reports a secret ingredient 1 out of every 4 times. Of the 3,153 Chesapeake frac jobs, 2,850 (90.3%)have at least one trade secret ingredient. And 73 operators report a higher percentage than CHK. Here is a complete list of the 766 operators who use trade secrets.

Are Calculations Even Possible?

Note: We are about to leave our data cleansing and curating rules and start looking at calculated fields. We calculate the total pounds for every ingredient in the FracFocus database. But this is not easy. We must adjust for some bad behavior inside the disclosures. For example, let's look at nCAS# 533-74-4, Dazomet.

There are 5,212 rows of 533-74-4 in the FracFocus database. But we can only calculate total pounds for 3,950. Why? Let's look at API# 42-127-36365. Here we show the water pumped is 13,604,976 gallons but there is no main water row in the disclosure. In fact, the only water we see is 0.02500%. If we used this small percent to calculate total volume/pounds for the frac job, we would be looking at 54 billion gallons, which is ridiculous. And because Dazomet is listed as 85% of the entire frac job (Dazomet is a solid), it would get about 38 billion of those gallons assigned to it. Dazomet is usually < 1% in these wells, but even after we add 4000 rows of Dazomet together, the 85% may drop only to 75% or so. Do we believe the average amount of Dazomet used in a frac is 75%? Of course not. But we must eliminate bad disclosures (85% Dazomet) to have any shot of meaningful, actionable analysis (most < 1%).

Let's stay with this subject a little longer. API# 37-125-24920 shows Dazomet at 24%. Water is 100%. Once again, this does not compute. It looks like the operator switched the %Additive and %HF Fluid columns. Water is 95.46590% and Dazomet is 0.00650%. But we can't be sure. %Additive is supposed to add up to 100% in each purpose section.

Iron Control				
	Citric Acid	77-92-9	0.00050	55.00000
Corrosion Inhibitor				
	Isopropanol	67-63-0	0.00030	30.00000
	Dimethylformamide	68-12-2	0.00010	10.00000
	Organic Amine Resin Salt	Proprietary	0.00010	10.00000
Bactericide				
	Dazomet	533-74-4	0.00650	24.00000
	Sodium Hydroxide	1310-73-2	0.00110	4.00000
Acidizing for perf				

Could we write a rule to reverse the columns? Of course, but even after swapping columns, 24% + 4% NaOH does not equal 100% bactericide. Iron control does not add to 100%. Corrosion Inhibitor adds to 50%. We are better off marking the entire disclosure as suspicious and excluding it from our calculations.

There are also times where it's not Dazomet's fault. In API#42-329-34246, CAS# 533-74-4 is reported as .088%, which seem perfectly reasonable, although Biocide %Additive only sums to 24%? But if we look at the water section…

Trade Name	Supplier	Purpose	Ingredients	Chemical Abstract Service Number (CAS #)	Maximum Ingredient Concentration in Additive (% by mass)**	Maximum Ingredient Concentration in HF Fluid (% by mass)**	Comments
30/50 White	Santrol	Proppant					
			Silica, Crystalline Quartz	14808-60-7	100.00000	37.85238	None
Water	Operator	Carrier/Base Fluid					
			Water	7732-18-5	100.00000	18.20081	None
100 Mesh	Santrol	Proppant					
			Silica, Crystalline Quartz	14808-60-7	100.00000	10.34242	None
40/70 White	Santrol	Proppant					
			Silica, Crystalline Quartz	14808-60-7	100.00000	9.85443	None
15% HCL	Reagent Chemical	Acidizing					
			Water	7732-18-5	85.00000	7.39149	None
			Hydrogen Chloride	7647-01-0	15.00000	1.30438	None
30/50 China	Endeavor Energy	Proppant					
			Corundum	1302-74-5	60.00000	7.27024	None
ANE-102G	Impact Chemical	Non Emulsifier					

We see only 25.5923% total water (18.20081 + 7.39149). Yes, the total water volume is only 253,018 gallons, but the proppant percentages add up to 65.31947% (37.85238 + 10.34242 + 9.85443 + 7.27024). Is it realistic that 65% proppant and only 25% water was pumped into the hole? We mark this entire disclosure as suspicious and therefore cannot calculate Dazomet pounds.

Here's an example where we pumped 1.1MM gallons of water, which accounted for 2.00138% of the total frac. Here is API# 34-013-20936, from May 2016, which appears to have a reasonable amount of Dazomet and water is 98.12687%. So, what could be the problem…

Purpose	Ingredients	Chemical Abstract Service Number (CAS #)	Maximum Ingredient Concentration in Additive (% by mass)**	Maximum Ingredient Concentration in HF Fluid (% by mass)**
Carrier				
	Water	7732-18-5	100.00000	98.12687
proppant				
	Silica, Quartz	14808-60-7	100.00000	10.63860
proppant				
	Silica, Quartz	14808-60-7	100.00000	1.60582
Acid				
	Water	7732-18-5	63.00000	0.94476
	Hydrogen Chloride	7647-01-0	37.00000	0.55486
Friction Reducer				

The first 3 ingredients alone add to 110.3713%? The entire %HFluid adds to 112.0371%. How do all the ingredients pumped into any wellbore add to 112%? We have additional cases where Dazomet is listed as an ingredient, with no %Additive or %HFluid recorded…

propoxylated				
Tetrahydro-3,5-dimethyl-1,3,5-thiadiazine-2-thione	533-74-4			

d water, and/or recycled water

As stated at the top of this section, we have over 1200 cases where Dazomet total pounds cannot be calculated, 647 of which are cases where Dazomet is listed on the disclosure, but we have no %HFFluid listed.

In total (across all fracs), we have 6,082 disclosures (8.3%) that lack the necessary information to compute overall total pounds and therefore pounds for each ingredient. This 8.3% impacts 295,198 of the 2,759,313 total ingredients,

meaning 10.7% of the disclosure ingredients do not have an accurate mass number. Is all hope lost? No. Remember, we still have 3,950 Dazomet ingredient lines that we can use for analysis, which should be enough.

Outliers and Ntile

One might ask how we can possibly track down and factor out all the bad disclosures and bad ingredients in FracFocus. Analysis helps a lot. When we see the (overall) average Dazomet is 75% per well, we know this is not possible. We track down the 85% culprit + a few >20% rows and see that water is underreported. We write a rule, re-process all 15K rules (which we have done 50+ times) and start the analysis again. Over time, we tend to tease out many of the problems with disclosures.

Most of our rules remove extreme behavior (38MM gallons of Dazomet!), but there will be cases where everything on the disclosure looks good (can be calculated) but one of the ingredients is significantly off. Let's look at a very simple example:

	All Water Rows	Main Water Rows
Total Mass	2,460,947,936,535	2,371,081,314,592
Row Count	220,982	66,589
Average Mass Per API*	33,528,357	32,304,000

Note: The row count is total number of ingredients, but the average mass is the average mass per frac job that contains water (73,339 disclosures). The average amount of water used in a frac job is ~4 million gallons.

There are 220,982 mentions of 7732-18-5 in the database. But we know there are wide variations in these water amounts. Let's look at API# 42-269-32872. We have two water rows, one at 81.92505% and the other at 0.03077%. But these water rows have two different functions. One is the main water, the other a small part of clay control. Beyond the 66,589 main water rows above, we see another 154,393 water rows (220,982 – 66,589), but these 154K only account for only 89,866,621,943 pounds of water (2,460,947,936,535 - 2,371,081,314,592) or 3.65%. Water is a simple example.

Methanol Example

The total mass for methanol is 376,170,800 pounds. But five ingredient rows, out of 87,993 methanol rows, account for 29.4% of the total methanol used. Note that we have 99,148 total methanol rows, but 11,155 of these have no %HFFluid reported on the disclosure or the disclosure is bad (see above section) meaning we could not calculate total pounds.

54,574 frac disclosures have at least 1 methanol row. Of these, we can calculate total pounds of methanol in 49,179 of the 54,574 (see previous section). Five of these 49,179 disclosures have more than 25% methanol in the frac. In API# 42-329-39052, we see 10,449,150 gallons of water was used and this accounted for 59.55691% of the total frac.

Trade Name	Supplier	Purpose	Ingredients	Chemical Abstract Service Number (CAS #)	Maximum Ingredient Concentration in Additive (% by mass)**	Maximum Ingredient Concentration in HF Fluid (% by mass)**	Comments
ZeroWash Tracer	ProTechnics	Diagnostics					
			Water (major)	7732-18-5	70.00000	59.55691	
			Methanol (major)	67-56-1	30.00000	25.46698	
			Ceramic Proppant	proprietary	14.00000	13.29844	Regulatory Compliance (713) 328-2320
			Dipropylene glycol methyl ether	34590-94-8	1.00000	0.83955	

With another 25.46698% being methanol, that means this single frac pumped 37 million pounds, or 4,468,134.66 gallons of methanol into this single wellbore (~500 tanker trucks). My guess? The 10MM water volume number, reported in header, probably includes both water and methanol.

These five, 25%+ rows don't change the average %HFFluid by more than 1/10th of a percent. But the average mass of methanol per frac (where methanol was used) changes dramatically. If we include all rows, the average mass is 7,649 pounds per frac. If we knock out those 5 rows (out of 87,993), our average drops to 5,392 pounds per frac, or 29% less

methanol per frac. Therefore, we conclude that to do a proper analysis of frac chemicals, we need to remove the outliers first. Don't get me wrong, these 5 fracs with 25% methanol are worth studying (individually and very) closely.

Ntile

We need to cut off the extremes, but can't do it with standard deviation because the data is not normalized. Instead, we will turn to a function called ntile that breaks the data into *n* equal groups. If we use ntile(1000), we break each nCAS into 1000 groups. Group #1000 would represent the top .1%. Looking at this for methanol, we see that the 1000th group has only 87 rows, but a total mass of 142,517,045 pounds. This group includes those 5, 25% disclosures that represent 29% of the entire methanol usage. The next step is to apply a filter to knock out the 1st group (lowest .1%) and the 1000th group (highest .1%) – fd.tentile >= 2 and fd.tentile <= 1000.

	Ingredient Count	Total Mass
All 67-56-1	87,993	376,170,800
1/10th removed	87,818	233,653,667
Differences	175	142,517,133

By knocking out the bottom and top 1/10th, we reduce the ingredient rows only by 175 (0.20%), but the mass by 142,517,133 pounds (37.9%). This is quite extraordinary that .2% of the rows can represent 37.9% of an ingredient's overall mass. Could we have just removed the 5 rows? Yes, but every ingredient has different extremes and we need to do this at scale.

Here's another reason to knock out extremes…

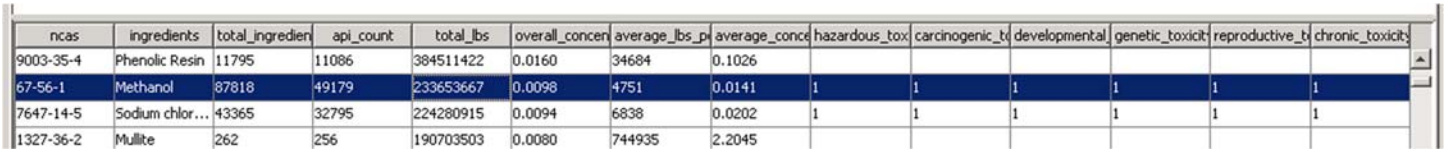
Water	Encana	Carrier/Base Fluid					
			Water	7732-18-5	100.00000	60.98526	None
Sand (Proppant) 40/70	Frac Specialist	Proppant					
			Silica Substrate	14808-60-7	100.00000	21.07571	None
Sand (Proppant) 100	Frac Specialist	Proppant					
			Silica Substrate	7647-01-0	100.00000	17.17873	None
Hydrochloric Acid (15%)	Independence Oilfield Chemicals	Acidizing					
			Water	7732-18-5	90.00000	0.27829	None
			Hydrochloric Acid	7647-01-0	15.00000	0.04638	None
LFC G4	Independence Oilfield	Guar Slurry					

In API# 42-227-38448, we see a 17% row for proppant. The CAS number for that proppant row is Hydrochloric Acid. Is that row sand or acid? We may never know, but the (main) water percent is only 60% and I can't see much more than 20% sand in the frac. If 17% row is acid, that means they pumped down 15,780,458 HCL pounds or 1,894,412.75 gallons. That's ~210 tanker trucks of acid showing up at this single wellbore. Whether legit or not, eliminating this row as an extreme is important for reliable analysis.

Toxicity

EPA runs a database of 456K CAS numbers called ACToR (Aggregated Computational Toxicology Resource) that tracks toxicity data. There are 2,543 unique CAS numbers in the FracFocus database. As we showed above, 967 of these were identifiable (nCAS). 818 of the 967 have a match between FracFocus and the ACToR database. These 818 account for 1,784,633 of the 2,759,313 (64.6%) total ingredient rows in FracFocus.

To understand toxicity, we need to do some calculations and combine them into a query. This is why the above section is so important. For example, we have seen extreme behavior in Methanol and Dazomet that would call the entire toxicity and other analyses into question.



ncas	ingredients	total_ingredien	api_count	total_lbs	overall_concen	average_lbs_p	average_conce	hazardous_tox	carcinogenic_t	developmental	genetic_toxicit	reproductive_t	chronic_toxicity
9003-35-4	Phenolic Resin	11795	11086	384511422	0.0160	34684	0.1026						
67-56-1	Methanol	87818	49179	233653667	0.0098	4751	0.0141	1	1	1	1	1	1
7647-14-5	Sodium chlor...	43365	32795	224280915	0.0094	6838	0.0202	1	1	1	1	1	1
1327-36-2	Mullite	262	256	190703503	0.0080	744935	2.2045						

The total ingredient rows for methanol is 87,818. Remember this count has removed the top and bottom 1/10th percent (87,993 – 87818 = 175). Ingredients can be listed more than once in a disclosure. We have 47,179 frac disclosures that include the chemical methanol. If we look at the total pounds of methanol divided by this 47,179, we see a methanol to water ratio of .0098%.

Toxic	Gallons	0.0098%	0.0141%
100ml	0.026417	269.5633	187.356
30 ml	0.007925	80.86898	56.20681

There is some debate about how much Methanol is toxic (30-100 ml), but it looks like one would need to drink between 50 and 270 gallons of frac water / methanol mix to reach dangerous levels. But fracs are not just water and methanol. There can be a mix of toxic chemicals within that frac that together make the frac water more dangerous. This would require a very intense analysis, but directionally, one would have to consume many gallons of frac solution.

	Pounds		Without Sand	
Total	2,680,627,130,000			
Carcinogenic	275,867,959,000	10.291%	11,621,556,200	0.434%
Chronic	275,830,342,000	10.290%	11,583,886,300	0.432%
Developmental	274,885,968,000	10.255%	10,639,299,600	0.397%
Genetic	272,512,680,000	10.166%	8,265,893,400	0.308%
Hazardous	280,110,629,000	10.449%	15,864,340,500	0.592%
Reproductive	9,510,779,900	0.355%	9,510,779,900	0.355%

Let's look at some general themes. Sand is toxic, per the EPA. Yes, the sand we walk on at the beach can lead to silicosis and lung cancer (breathing it in a mine setting). Therefore, we can't generalize about overall toxicity, because each chemical has very different levels required to be toxic. With that said, we estimate about .5% of frac fluid mixtures are toxic chemicals, with wide ranges of toxicity levels.

The Role of 8.33

When a disclosure is submitted, the author has two options: 1) Enter the HF Fluid percentages directly or 2) enter the mass number for each chemical and the website will calculate %HFFluid for you.

In 46,843 disclosures (62.2%), submitters enter the actual mass of every single chemical. This, in turn, is used to calculate the percentages. Add up the masses and divide each chemical by that total to get %HFFluid. 20,825 have only percentages (no mass) entered in the disclosures. The rest only have some of the masses recorded and its usually water.

API# 04-030-49274 shows water volume at 79,254 gallons. 79,254 * 8.33 = 660,186 pounds of water. Using the water percentage of 96.23778, we calculate the total mass of the frac as 685,994.44 pounds (660,186 * 100 / 96.23778). The operator reports a water mass of 6,675,276 pounds. This divided by 8.33 (density) of water = 801,353.66 gallons of water. Was the well treated with 79,254 gallons of water or 801,353 gallons of water? If we look closely at the rest of the data, my guess is the operator accidentally entered 79,254 instead of 793,54_, which would be within a rounding error of our 801K gallons. The comma was put in the wrong place. Note how easy it would be to implement a check within FracFocus. The number 6,675,276 you entered for water mass does not equal (or come close to) the 79,254 gallons you entered in the FracFocus header.

There are 52,268 distinct frac disclosures where water is entered as a mass (some disclosures include only a water mass row, others have masses for all ingredients). 6,800 (13%) of these under or over report water by 25,000 gallons (208,250 pounds). We have 3,753 (7.18%) cases where the water in the header is lower than the water reported in the mass field. Remember, we are not comparing a state site to FracFocus. This is one part of FracFocus (the header's total water volume) disagreeing with the ingredient part of the database (hand entered, mass ingredient).

Some may think we need to add all water from each disclosure, instead of main water row, and compare it to the gallons reported in the header. In other words, when an operator enters 1,000,000 gallons of water, is this just the largest percent of water row or all the water rows, no matter how small, added together? When we redo the analysis using the sum of all water rows in each frac, the number with a different greater than 25,000 gallons jumps to 14,324. Therefore, we believe most submitters are only using the main water row volume when reporting gallons in the header.

Harvard

FracFocus version 2.0, launched in June 2013, introduced checks on data integrity before disclosures were accepted. Harvard, who had published a report in 2013, was invited to participate in FracFocus 3.0 upgrades. Harvard published an updated report in January 2016. One of the upgrades (launched in March 2015) was a check on the CAS number. Since the 3.0 launch…

1. There have been 10,901 invalid CAS numbers starting with a zero. 682 of these are 00-00-0.
2. 865 start with NFIDB? For example, NFIDB:FDP-S1050-12.
3. 220 list only a single dash (all by itself) as the CAS number.

We assume Trade Secrets will be listed using that term or the term confidential or perhaps proprietary…

1. Listed below was used 10,282 times. This comes from the separation of purposes from ingredients issue.
2. 29,315 are listed as NA, N/A or Not Assigned.
3. 1,007 are listed as CAS Not Listed.

Therefore, we believe the statement Harvard made in November 2015, "… issue an error message and require amendment of any CAS number that fails the check-digit verification…" is incorrect.

Going further back to the launch of FracFocus 2.0, authors were alerted to invalid CAS numbers and given tools to improve data quality. Since, 6/1/2013, there have been 160,336 ingredients that do not comply with CAS or Trade Secret standards. The same document that talks about FracFocus data quality also discusses Texas Rule 29 and penalties for non-compliance.

Technologists have a radically different view of the world. It doesn't matter if there are 886 bad entries or 1 or 30,000. We look for the possibility that any cell can be submitted with bad data. If it can happen once, we need to look for it always, across the entire database, and attempt to correct it. And it's the aggregation of these smaller scale bad data entries that add up to a database that cannot be used for detailed chemical analysis, without a thorough cleaning.

Today, there are two types of FracFocus analysis. Harvard, EPA and others point out the problems and recommend future enhancements (for newly submitted disclosures). The other is to cleanse, curate, add value to the data, and take the next step to analyze each chemical's role in the frac. This allows us to come up with a best practice frac recipe. It's the imperfections in the data and the 37MM fixes/enhancements that create first mover, competitive advantage. If the data was perfect, this document would not exist and technologists (data engineers) would not be needed.

The Analysis

We can approach this cleaned, curated frac data into two different ways – deductive and inductive analysis. I talk about the difference between these two approaches in another document called Machine Learning Primer. The document you are reading will assume you understand the basics of machine learning.

Machine Learning

There are two parts to our machine model, features and outcomes. Let's start with outcome or label as it's called by data scientists. We need to know if a well is under (label = 0) or overperforming (label = 1). We could have more than 2 labels, but it makes the model more complex and less effective at predicting outcomes. We prefer a binary model.

Label

Our model will start by taking the second month of production after the well has started producing. In other words, if we have production in month 1, 2 and 3, we will use month 2. If we have no production in month 1 and 2 and production in month 3 and 4, we will use month 4. We then need to take lateral length into account. This is not always reported, but we can get a pretty good estimate by looking at the difference between surface and bottom hole coordinates. At this point, we have production per lateral foot.

Unfortunately, this data will not be normalized. For this, we recommend a power transformation. We use Box-Cox. Once the production per lateral foot is normalized, we can assign a 1 (winner) or 0 (loser) value to each well. This feels very simplistic. If you have a more sophisticated way of judging frac production, we are totally up for it. but we need to apply your rules/logic at scale. This often knocks out a lot of good ideas.

Features

Next up is features. For this side of the model, we will look at each chemical used in the frac. Can we use all 900+ chemicals for the model? 70K rows of data (one for each disclosure) and 900 columns (one for each nCAS). Unfortunately, no. If a chemical is only used in 30K (of 70K) disclosures, this leaves 40K empty rows. Machine learning does not like empty fields. We can set NULL values to zero (chemical was not used), but we would not want 90 rows with (chemical used) values and 69K+ values with 0 (not used). We might set a minimum requirement that at least 5,000 fracs must use the chemical (= 82 nCAS numbers)? We can adjust this threshold after the models start running.

Next we need to assign a value to each chemical. This could be as simple as 1, chemical was used or 0, chemical was not used. Instead we recommend using a 5-point scale to judge the concentration of each chemical used in the frac. To get this number, we will look at the ratio of each chemical's mass in a frac to the total mass of the entire frac minus proppant. This will give us the concentration percentage of that chemical in the frac. We eliminate proppant because it's usually a significant amount of the total mass and it can vary depending on the type of proppant used. Almost all other chemicals, except water are used in very small amounts. Proppant variety can throw off our analysis.

Once we have a concentration for each chemical in each frac disclosure, we will normalize the results. Again, we will use a power transformation. If a chemical in a disclosure is very close to the average across all fracs, it will be given a value of 3. If it's above average, we will assign a 4 and significantly above average will get a 5. Same with 1 and 2. 0 will be the absence of the chemical in the frac.

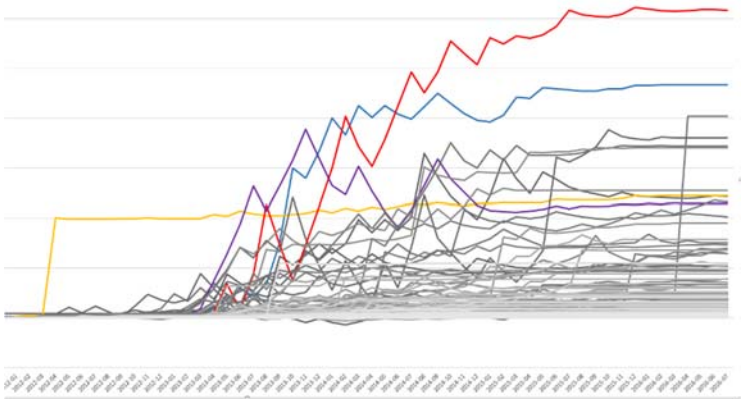
Models

Ultimately, we will have all 70K disclosures as rows (more when we use the entire PDF database), nCAS as columns and a value of 0 through 5 in each field. We will have an additional column called label where we assign production success/failure to each API in the database. Note, we (our team) will be using 119K disclosures (rows) versus the 70K in this analysis. Then it's a matter of running multiple models and seeing which model comes closest to predicting a successful set (recipe) of frac chemicals.

If this model works and we have enough data, an endgame would be for operators/pumpers to enter a recipe into the final model and it will predict if the recipe is good (above average production expected) or bad. It will also give us a degree of certainty in that prediction.

Deductive Algorithms

Where machine learning is a pretty radical step for many (let the computer write the program for us), conventional software engineering should also tell us a lot about the success of frac chemicals. Let's look at a simple example…

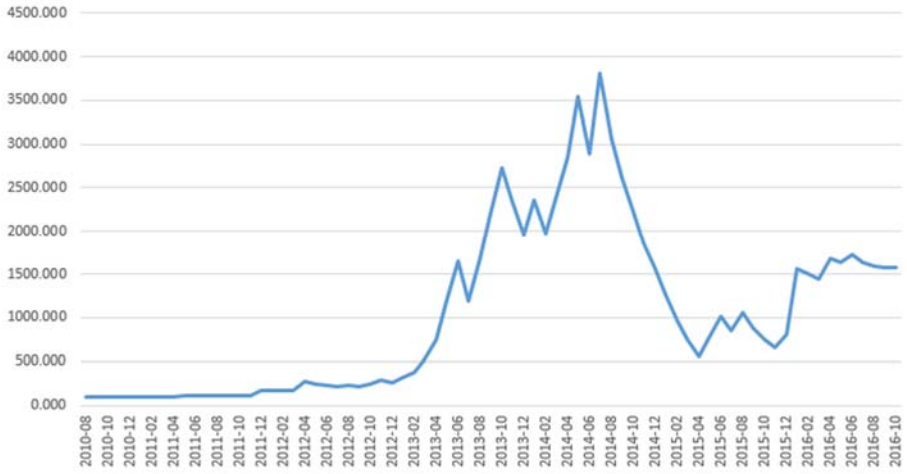


Larger version of chart at right <here>. We take all operators (155) who have used methanol in fracs for more than 20 months (noncontiguous is ok). We sum the methanol, by operator, by month and divide it by the count of operator's fracs in that same month. Again, this is a very simple example.

Each operator is set to 100% in July 2010. Month after month, we plot whether each operator increased or decreased their use of methanol. Looks like operators tried it around beginning of 2011 (left side of chart). Most just used the same amount, if any, throughout the next 5 years. Linn Energy was the first to try larger amounts of methanol (yellow line). And it's easy to track down the data behind the huge bump in the yellow line for further review.

api	jobstartdate	ncas	ingredients	maxingrconchffluid	totalwatervolume	totalwaterperc
42483333810000	2012-04-11 00:00:00	67-56-1	Methanol	3.2303999999999999	5670813	92.00242
42483333820000	2012-04-15 00:00:00	67-56-1	Methanol	3.21515	5766087	92.16171
42483333800000	2012-04-05 00:00:00	67-56-1	Methanol	3.1786400000000001	5385373	92.15415
42329378390000	2012-04-11 00:00:00	67-56-1	Methanol	0.079869999999999997	1005060	89.60511

In April 2012, Linn Energy tried an experiment with 3 fracs, using 3%+ methanol and high water volumes and low proppant percentages. API# 42-483-33381 shows 1 of the 3 examples. The next logic step would be to look at production for these 3, but it does look like Linn immediately leveled off after this experiment and went back to using very low percentages. However, Apache (red line) and Devon Energy (blue) found methanol and never looked back. They have consistently increased use of methanol in fracs for 3 years and leveled off only recently. This is something we would want to study.



We can also study overall methanol usage using the same approach, but summed (above chart) across all operators. Again, this is not the average methanol per month. This is how methanol usage changed over time. We assume 100% methanol in 7/2010, then look to see if methanol increased or decreased month over month. Above we can see that methanol use increased up until mid-2014 and dropped off significantly, almost like it grew out of favor. To contrast the above (cumulative, chart, here is what the data looks like if we plot the raw data.

The Crawling Algorithm

Our 15K rules that cleanse, curate and calculate over FracFocus are examples of conventional software development. These rules are loaded on a server and we go line by line through raw data making changes and additions. We log every change, so an admin, with a single click can find the rule and change it without a developer's help. We have to know what's interesting about the frac data, in advance, to write code on what to look for and seek out. Is Linn, with a 3% experiment, more interesting than Apache and Devon leveling off their use of Methanol?

We recommend a technology called crawling, walking through the data, looking for deviant behavior (as the crawler goes). We might have a chemical crawler that keeps track of each chemical, its relationship to other chemicals, and its normal role in a well frac. When the crawler sees something abnormal in the data pattern, it marks the event and compares it to production and other outcomes of the wellbore. If the outcomes changed significantly with the input behavior, we want to bring this data relationship to the surface (excuse pun).

We might have crawlers for water, location (radial), proppant type, proppant levels, operators, suppliers, etc. We could see 4-5 chemical crawlers. One might look at the relationship between chemicals, one might look at levels within an operator or a frac, one might look at level changes over time and another might look at the chemical's overall relationship to production.

Recommendation

Throughout this analysis, I almost sound like an environmentalist, someone pointing out all the flaws, in detail, with frac disclosures. I do this for competitive advantage as a technologist (data engineer). If the data were perfect, all would have equal access to this benefit. Instead, if we can correct (most of) the inaccurate/incomplete data, it puts us in very, very small company… some would say a company of one. It allows us to start asking and answering questions like, "Is this chemical we are putting down the wellbore going to make any difference at all?" This hard-fought battle with the FracFocus data will also allow us to come up with a best practice frac recipe that is a baseline for completing any well.

Some might ask, "Is the data clean enough?" After all, I point to a lot of problems and fixes in the disclosure database. While we might have 5,000 rows of Dazomet data and only 4,000 that are accurate enough to be analyzed, we feel this is more than adequate. And what's the alternative? Do no analysis at all. Any analysis that gives us insights into chemical effectiveness puts us ahead of where we were yesterday.

We recommend operators and pressure pumpers evaluate every single chemical, combination/mix and quantity going into the hole. What's the value of each chemical versus it's toxicity? What's the perfect blend of sand and water with respect to well mechanics, geology and location? This will be the first-time chemicals can be evaluated in the aggregate and at scale (across all disclosures).

Appendices

Appendix A: Flowback Expectations

We got a question about flowback after frac chemicals have been pumped into the well.

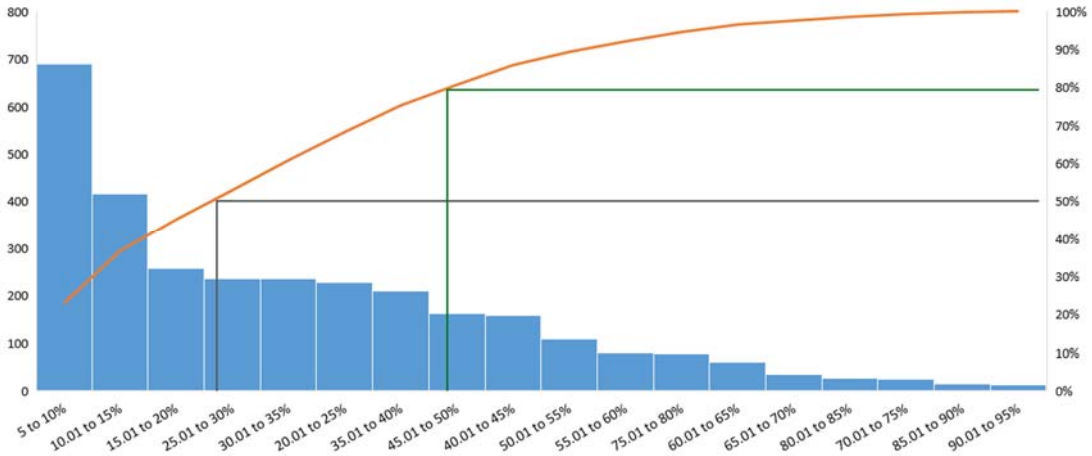
```
select api, treatmentdate, totalfluidbbls,
round(totalproppantlbs/12.76/42,2) as proppant_bbls,
round(totalfluidbbls + totalproppantlbs/12.76/42,2) as totalpumpedvolumebbls,
totalflowbackvolumebbls,
round(totalflowbackvolumebbls/(totalfluidbbls + totalproppantlbs/12.76/42)*100,2) as percent_flowback
```

from coformationtreatment where totalflowbackvolumebbls > 0 and treatmenttype like 'fracture%'

We can use Colorado Treatment Summaries, which report flowback after a fracture stimulation. We have total fluids in barrels. We convert proppant to barrels by taking pounds of proppant and dividing by a density of 12.76 to get gallons of water. Then we take the gallons of proppant and divide by 42 to get barrels of proppant.

Next, we add total fluid barrels to total proppant barrels. We then divide total flowback by the sum of fluid barrels + total proppant barrels to get the percent of all fluids that flowed back to the surface after the frac job.

The 4,134 fracture stimulation results in Colorado tell us that 50% (grey line) of all wells flowback < 25% of total fluids pumped into the wellbore. By 45% flowback, we have covered almost 80% (green line) of the wells.



The average flowback for the 3,008 wells analyzed (flowbacks between 5% and 95%) is 27.88%.

Appendix B: Importing the SQL Data

The query to connect the three imported tables after downloading SQL BAK file has recently changed.

```
Select * from FracFocusRegistry.dbo.RegistryUpload ru
join FracFocusRegistry.[dbo].[RegistryUploadIngredients] ri on ru.pkey=ri.pKeyDisclosure
left join FracFocusRegistry.dbo.RegistryUploadPurpose rp on ri.pKeyPurpose = rp.pKey
where (ri.pKeyPurpose is null or ru.pkey = ri.pKeyDisclosure)
```

Note a left join has been added. This allows for headers to jump over the purpose table and connect directly with ingredient rows. While there are ingredients with blanks purposes, left joins affect 34,689 ingredient rows with no connection to the purpose table. Interestingly, this issue came about after FracFocus 3.0 launch.

For example, API# 47-049-02406 has one ingredient connected to purpose table (water). Everything below the green line has no purpose in the database. API# 05-123-41038 is another example. This impacts 969 disclosures (1.3%), 97% of which are from 2016.

Appendix C: FracFocus Statistics

While most of this document is about cleaning and curating data for analysis, here are some general statistics.

Top Ingredients

TOX	Ingredient Name	Count	Mass (Pounds)
	water	251,834	2,460,954,065,967
	quartz-alpha (sio2)	96,589	169,299,086,737
	silicon dioxide	56,958	91,228,745,702
*	silica, amorphous, fumed, cryst.-free	5,891	7,323,818,790
*	sulfur dioxide	20,537	4,581,434,681
*	hydrochloric acid	63,890	3,992,984,932
*	alumina	6,402	2,230,556,164
	guar gum	43,379	1,570,358,942
	msds and non-msds ingredients listed below	10,723	952,071,678
*	sodium chloride	53,162	410,236,312
	distillates, petroleum, hydrotreated light	19,269	360,755,372
	kerosene, low odor	35,151	328,921,092
	white mineral oil, petroleum	13,851	260,609,864
	phenolic resin	10,158	251,990,453
*	propargyl alcohol	104,430	226,945,253
	petroleum	13,684	215,965,982
*	choline chloride	13,630	205,908,881
*	sodium hydroxide	37,803	167,591,691
*	potassium chloride	5,609	156,560,829
	calcium dichloride dihydrate	5,979	153,458,112
*	ethylene oxide	44,577	126,415,564
*	isopropanol	53,161	100,963,438
*	triethylene glycol	26,086	95,107,052
*	methenamine	9,941	92,122,396
*	ulexite	6,440	86,210,181
	proprietary	27,172	84,579,280
	trade secret	6,376	83,296,021
*	iron(iii) oxide	6,299	79,710,131
*	ammonium acetate	9,213	70,697,107
*	vinyl acetate	27,937	70,544,200
	ammonium salt	7,351	66,279,813
*	potassium hydroxide	27,817	64,076,624
*	glutaraldehyde	20,946	61,096,523
	sodium; acrylic acid; prop-2-enamide; prop-2-enoate	9,780	58,235,009
	borate salts	4,464	57,456,095
*	ammonium chloride	25,726	55,044,482
	distillates (petroleum), hydrotreated light paraffinic	4,927	52,445,204
*	sodium chlorite	12,224	49,530,760
*	ethylene glycol	5,575	46,703,210
*	diammonium peroxydisulfate	30,952	46,141,161
	2-propenoic acid, polymer with 2-propenamide, sodium salt	4,835	44,800,509
*	methanol	22,472	42,907,292
	formaldehyde, polymer with 4-(1,1-dimethylethyl)phenol, 2-methyloxirane and oxirane	13,333	40,607,592
	solvent naphtha, petroleum, heavy arom.	12,366	36,838,626
*	titanium dioxide	5,762	33,673,665
*	2-butoxyethanol	21,307	33,607,448
*	potassium metaborate	10,307	32,403,071
*	glycerol	9,714	31,879,897
*	sorbitan, mono-(9z)-9-octadecenoate	10,674	28,757,872
	ethoxylated c12-16 alcohols	8,766	28,301,425

This is a list of the Top 50 chemicals listed in the FracFocus database, ordered by mass (pounds) descending.

The rows highlighted in yellow are not chemicals. They are 10,723 rows where chemical name is “msds…” that add up to 952MM pounds of chemicals. Even though designed to be a PDF header/separator, operators have written in a mass number for 2,740 rows out of the 10,723.

We aggregate proppants for detailed analysis, but have left them as separate chemicals here.

Proprietary and trade secret are ranked as the 26th and 27th largest ingredient. These 27,172 + 6,376 are FracFocus rows where we could not solve for ingredient by looking at ingredient name.

The asterisk in column 1 shows which of the Top 50 chemicals are marked as carcinogenic by the EPA. 29 of the 50 are marked. SiO2 (row 2) is also carcinogenic but only at high quantities and if breathed in to lungs over time (like in a mine).

Top Suppliers
Halliburton supplies the most chemicals, but Schlumberger provides a much higher mass than HAL and BHI combined.

When we remove water, SLB mass drops to 28,439,948,967 lbs and HAL takes the top spot with 50,648,435,113 lbs.

Baker Hughes supplies a lot less chemicals than the other 2 main pumpers.

"Operator" is almost always used to describe water provided locally, instead of by a supplier.

N/A is listed 13,233 times. There is no reason not to list who supplied the chemical. That's not the trade secret part.

Supplier	Count	Mass (Pounds)
halliburton	246,760	61,447,253,667
schlumberger	217,065	210,227,788,906
baker hughes	168,048	15,939,980,648
chemplex	113,038	738,010,226
frac tech services	102,671	33,609,117,203
nabors completion	87,891	7,608,813,233
weatherford	54,197	11,708,998,227
performance technologies	47,566	16,841,237,272
pioneer natural resources	37,483	87,124,476,530
operator	34,140	907,486,836,759
cws	33,977	6,147,031,492
cesi chemical	31,320	38,230,567
economy polymers	30,226	203,303,650
nalco	25,132	71,548,152
trican	24,406	6,123,228,406
universal well services	24,111	8,353,630,566
c&j energy	23,672	8,764,384,663
ask chemical	22,758	53,069,540
sanjel	20,595	5,306,799,912
frac specialist	19,267	462,234,615
multi chem	18,261	108,132,948
independence	17,744	132,004,883
liberty oilfield services	14,802	4,662,013,110
n/a	13,233	74,850,171
archer	12,021	33,602,188,783
rockpile energy	11,941	1,801,572,604
southwestern energy	11,394	5,424,982,023
reagent chemical	11,017	1,124,410,567
wst	10,947	36,830,189
u.s. well services, llc	10,485	5,205,599,337
calfrac well services	10,181	934,551,534
basic energy	9,973	336,346,912
santrol	8,825	6,674,042,315
cudd energy service	8,662	3,732,342,362
catalyst chemical	8,499	15,525,305
pfp technology	8,430	43,944,433
energy & environmental services	7,471	45,963,510
roywell services, inc.	6,657	644,490,772
tes	6,649	869,186,692
consolidated oil well services	6,483	401,397,833
united petroleum products	6,399	475,288,288
nabors	6,213	4,131,730,806
reef services	6,037	106,438,868
x-chem	6,020	51,562,686
advanced stimulation	5,570	458,992,784
petroplex	5,569	83,123,429
eog resources	5,103	186,909,271,205
multiple suppliers	5,076	2,294,474,191
dow chemical company	5,026	11,042,406
other chemicals	5,002	181,927,217
colossal chemical	4,779	22,569,211

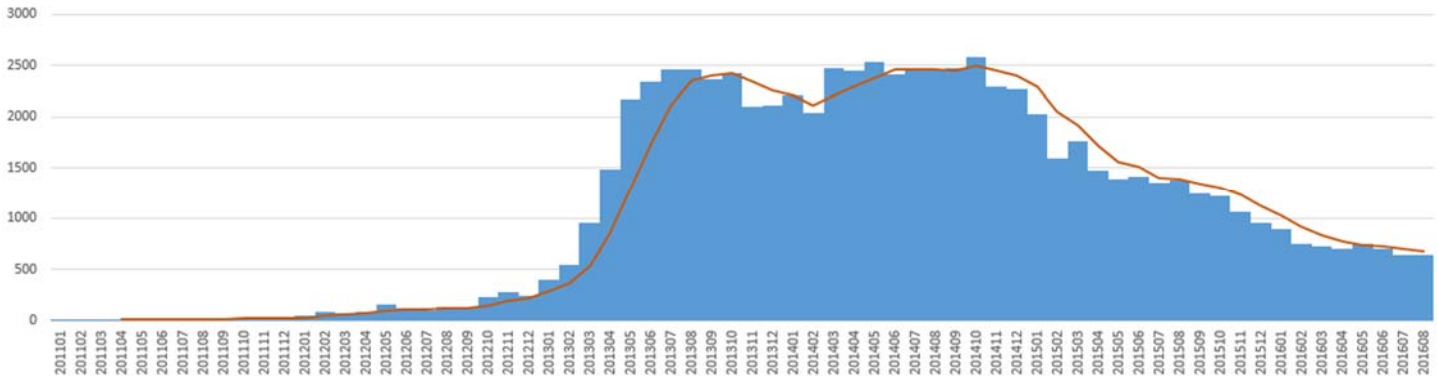
Top Operators

Highest numbers in each column are highlighted green. “Go to” is the supplier most hired by the operator.

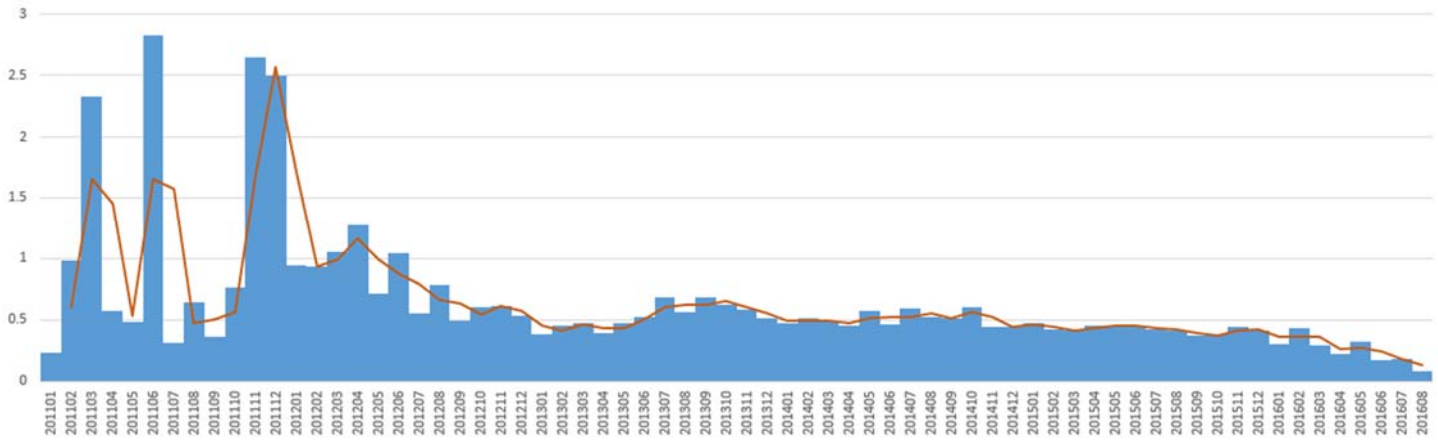
Operators	Count	Mass (Pounds)	Go to Supplier	2016 Fracs	2016 (%)
Anadarko Petroleum Corporation	3,616	166,829,124,065	Nabors Completion	264	7.30%
Chesapeake Operating, Inc.	3,162	156,055,812,023	Performance Technologies	267	8.44%
XTO Energy/ExxonMobil	2,626	107,682,160,724	Chemplex	250	9.52%
EOG Resources, Inc.	2,453	192,307,957,889	Chemplex	295	12.03%
Devon Energy Production Company L. P.	2,321	107,252,378,259	Baker Hughes	68	2.93%
Apache Corporation	2,161	79,105,465,631	Baker Hughes	84	3.89%
Occidental Oil and Gas	1,905	36,545,178,047	Nabors Completion	170	8.92%
Pioneer Natural Resources	1,745	101,622,820,789	Pioneer Natural Resources	149	8.54%
ConocoPhillips Company/Burlington Resources	1,728	50,038,074,657	Halliburton	107	6.19%
Aera Energy LLC	1,720	1,494,918,494	Baker Hughes	138	8.02%
Southwestern Energy	1,539	79,820,805,048	Southwestern Energy	17	1.10%
SandRidge Energy	1,499	20,368,408,555	Frac Specialist	26	1.73%
Marathon Oil	1,355	48,781,084,002	Schlumberger	137	10.11%
Encana Oil & Gas (USA) Inc.	1,304	61,704,470,664	Baker Hughes	89	6.83%
Chevron USA Inc.	1,195	28,818,269,497	Baker Hughes	69	5.77%
Whiting Petroleum	1,154	33,191,128,752	Schlumberger	57	4.94%
BHP Billiton Petroleum	1,106	60,257,342,694	Schlumberger	39	3.53%
Newfield Exploration	1,098	30,785,138,374	Nabors Completion	83	7.56%
Noble Energy, Inc.	1,083	67,187,998,499	Halliburton	109	10.06%
Continental Resources, Inc	1,003	58,713,610,118	Schlumberger	67	6.68%
COG Operating LLC	861	38,803,702,710	Baker Hughes	84	9.76%
Hess Corporation	856	20,852,448,055	CWS	83	9.70%
WPX Energy	840	16,594,980,637	Halliburton	47	5.60%
EP Energy	813	24,714,876,566	Frac Tech Services	95	11.69%
QEP Energy Company	785	20,589,337,181	Halliburton	78	9.94%
Range Resources Corporation	663	34,544,325,187	Frac Tech Services	85	12.82%
Energen Resources Corporation	627	25,100,814,888	Schlumberger	53	8.45%
BP America Production Company	619	10,457,138,369	Halliburton	100	16.16%
EQT Production	619	53,751,372,964	Frac Tech Services	77	12.44%
Linn Energy, LLC	590	4,707,039,939	Baker Hughes	5	0.85%
SM Energy	540	28,912,213,155	Sanjel	94	17.41%
PDC Energy	531	18,364,495,880	Halliburton	126	23.73%
Laredo Petroleum, Inc.	495	29,421,933,084	Frac Tech Services	34	6.87%
Murphy Exploration and Production USA	495	41,110,421,679	Frac Tech Services	19	3.84%
Antero Resources Corporation	491	49,404,113,006	U.S. Well Services, LLC	50	10.18%
Parsley Energy Operations, LLC	463	18,280,605,089	Economy Polymers	59	12.74%
Ultra Resources	462	3,917,348,178	Halliburton	80	17.32%
Lewis Energy Group	449	17,023,217,146	Magnablend	33	7.35%
Oasis Petroleum	448	16,838,786,273	WST	23	5.13%
Cabot Oil & Gas Corp	445	36,756,319,190	Halliburton	56	12.58%
Midstates Petroleum Company	420	6,390,744,075	Halliburton	36	8.57%
Cimarex Energy Co.	417	29,851,026,042	Petroplex	46	11.03%
EnerVest, Ltd.	408	12,336,057,850	Halliburton	64	15.69%
CrownQuest Operating, LLC	399	7,929,459,282	Economy Polymers	13	3.26%
Citation Oil and Gas Corp.	391	664,932,571	CESI Chemical	1	0.26%
Endeavor Energy Resources	380	7,473,133,147	Advanced Stimulation	22	5.79%
Carrizo Oil & Gas, Inc.	371	24,291,302,874	Halliburton	65	17.52%
Athlon Energy Operating LLC	345	6,960,283,258	Frac Specialist	0	0.00%
MEWBOURNE OIL COMPANY	340	6,938,059,717	Baker Hughes	34	10.00%
Crescent Point Energy US Corp	329	2,817,765,330	Halliburton	8	2.43%

Graphs

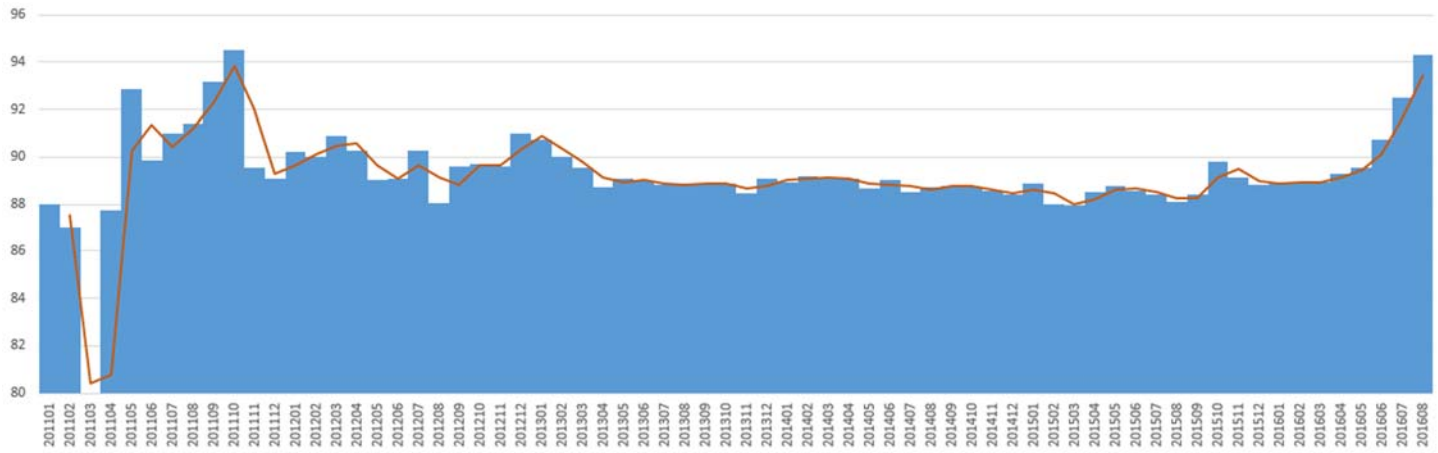
FracFocus disclosures per month, orange line is moving average



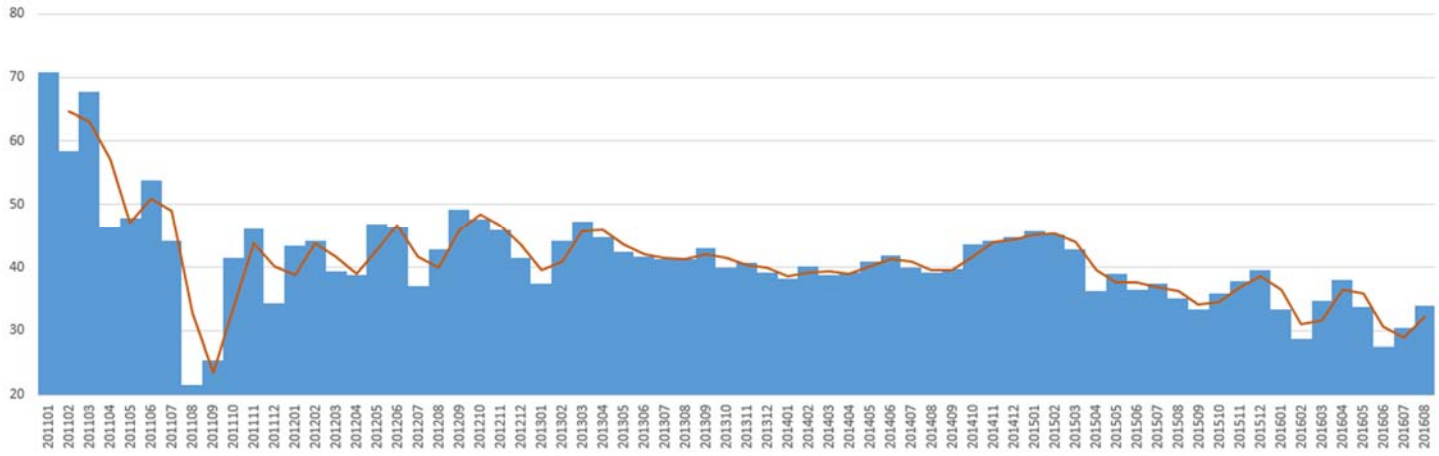
Percent toxic (carcinogenic) chemicals to overall mass. Operators tried larger amounts of toxic chemicals before 2011 and have since reduced the percentage of total frac significantly. We are even seeing further reductions in 2016. Note: This chart does not include SiO2 (proppant), which can be toxic at very high levels.



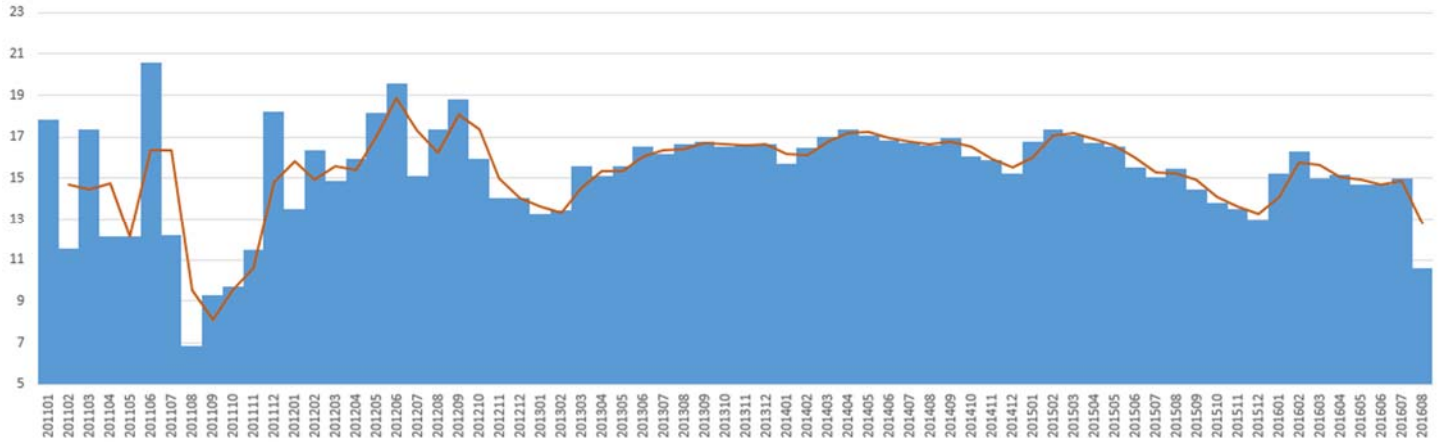
Water mass as a percent of total mass. In 3/2011, water was 73.83% and does not show below (outlier). Water percentage appears to be trending up in 2016 after staying consistent for 2-3 years.



Percent suspicious ingredients, excluding withheld secrets, as percentage of total ingredients used each month. Can see that FracFocus 3.0 improved reporting accuracy, but bad rows still have issues 25-30% of the time.



Trade secrets (+ confidential and proprietary) as a percentage of total ingredients continues to hover around 15%.



Appendix D: Cost Data

Using Alibaba and metric ton pricing, we built a chemical price list for the top 150 chemicals used in fracs. These prices can be updated very easily as better data becomes available. We subtract water from our analysis and used $100 as the average price for a ton of sand. Sand varies from ~50-$150, depending on quality. We also ran the analysis at $50 per ton of sand too.

We believe ~ $19,651,298,271.32 in non-water chemicals have been pumped into the 73,352 frac disclosures in the database. This $20 billion covers ~98% of the chemical mass in all the fracs and 66% of the CAS numbers. We also had to exclude 1,338 disclosures that either had a frac record on the exact same job start date (same API) or were complete duplicates. We expect that when the entire database (119K) disclosures are included, our costs will increase by $11.7B to just over $31B. We believe these costs estimates are low (average cost per frac = $299K), but wanted to get an idea of which operators spent more per API on chemicals.

	Chemical Cost
Texas	$ 10,287,448,911.23
Oklahoma	$ 1,342,012,940.34
North Dakota	$ 2,158,192,378.75
Colorado	$ 1,333,619,073.18
Pennsylvania	$ 1,509,188,824.40
Utah	$ 108,666,388.80
Wyoming	$ 286,511,090.23
New Mexico	$ 429,529,609.33
California	$ 43,165,311.16
Arkansas	$ 367,561,814.82
Ohio	$ 790,046,053.93
West Virginia	$ 546,398,321.24
Louisiana	$ 358,481,117.36
Kansas	$ 18,438,013.57
Montana	$ 70,467,278.15
Virginia	$ 1,571,144.85

Texas spends more on chemicals that all other states combined.

Here are the largest operators, by chemical cost, not including water:

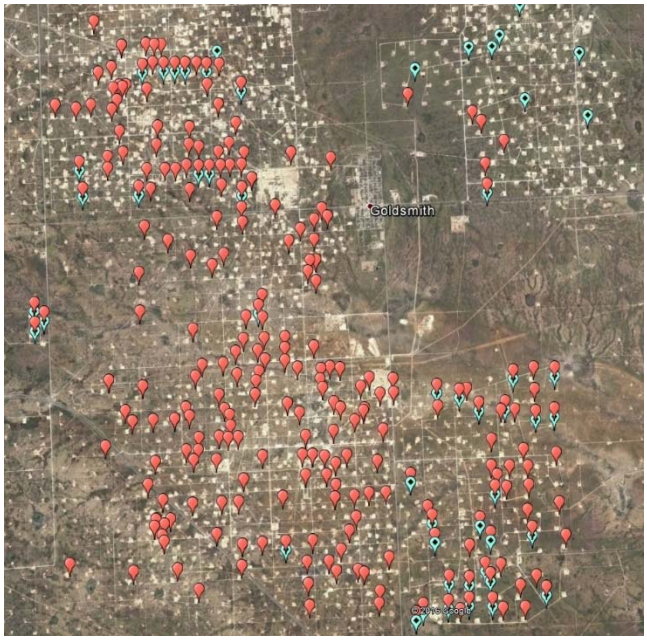
	Totals	Frac Count	Avg per Frac	Texas Only	% TX to Total
Chesapeake Operating, Inc.	$ 1,563,837,107	3162	$ 494,572	$ 801,059,925	51.22%
EOG Resources, Inc.	$ 1,202,786,063	2453	$ 490,333	$ 961,006,558	79.90%
Pioneer Natural Resources	$ 876,412,784	1745	$ 502,242	$ 875,173,397	99.86%
Anadarko Petroleum Corporation	$ 770,732,188	3616	$ 213,145	$ 442,662,686	57.43%
XTO Energy/ExxonMobil	$ 645,024,028	2626	$ 245,630	$ 203,138,906	31.49%
ConocoPhillips Company/Burlington Resources	$ 591,211,085	1728	$ 342,136	$ 383,923,881	64.94%
Marathon Oil	$ 505,828,654	1355	$ 373,305	$ 412,851,929	81.62%
Devon Energy Production Company L. P.	$ 480,540,682	2321	$ 207,040	$ 191,506,085	39.85%
Southwestern Energy	$ 454,199,860	1539	$ 295,127		
Noble Energy, Inc.	$ 438,117,770	1083	$ 404,541		
Apache Corporation	$ 433,965,209	2161	$ 200,817	$ 369,417,636	85.13%
BHP Billiton Petroleum	$ 420,218,295	1106	$ 379,944	$ 382,578,717	91.04%
Continental Resources, Inc	$ 411,232,641	1003	$ 410,003	$ 2,525,783	0.61%
Encana Oil & Gas (USA) Inc.	$ 322,010,113	1304	$ 246,940	$ 171,310,223	53.20%
COG Operating LLC	$ 321,589,010	861	$ 373,506	$ 262,297,162	81.56%
Newfield Exploration	$ 291,383,717	1098	$ 265,377	$ 34,596,204	11.87%
Antero Resources Corporation	$ 281,855,808	491	$ 574,044		
Occidental Oil and Gas	$ 275,608,095	1905	$ 144,676	$ 194,637,829	70.62%
Murphy Exploration and Production USA	$ 273,154,774	495	$ 551,828	$ 273,154,774	100.00%
EQT Production	$ 265,653,010	619	$ 429,165	$ 2,845,787	1.07%
Cabot Oil & Gas Corp	$ 260,825,317	445	$ 586,124	$ 120,765,570	46.30%
Whiting Petroleum	$ 246,636,146	1154	$ 213,723	$ 1,090,765	0.44%
Lewis Energy Group	$ 235,281,167	449	$ 524,012	$ 235,281,167	100.00%
SM Energy	$ 229,059,740	540	$ 424,185	$ 179,579,202	78.40%
QEP Energy Company	$ 202,355,026	785	$ 257,777	$ 46,600,304	23.03%

For a more complete analysis of cost data, click here. And again, any and all costs can be updated and the analysis re-computed within a couple minutes.

Appendix E: No Perfect Recipe (Yet)

One might think 10 or more ingredients dominate every frac. This is not the case at all. In fact, there is very little agreement between operators on what chemicals to use beyond water and sand. For example, Potassium Metaborate, is used in 10,028 (13.7%) frac disclosures (10,307 total, 279 have more than one BKO2 row). It appears to be randomly used? Same with Ulexite (8.5%) and Methenamine (12.5%).

When we map a small geography (right), we see that methenamine is used in random areas (blue markers) and most of the area wells do not have it (red).



Perhaps the use of methenamine is linked to an operator. It's not. There are 362 different operators who use methenamine. Occidental Oil is the number one user of methenamine at 772 uses, but this is less than 50% of Occidental frac jobs. And even within a confined area (Pecos, Texas) there is a strong mix of wells that use methenamine and do not use it.

Is methenamine the favorite of a specific pressure pumper? HAL, BHI and SLB have never used methenamine in any frac. Same with Potassium Metaborate and Ulexite. We conclude Occidental, i.e. no operator nor any pumper has found a "magic" combination of chemicals they use consistently.

This entire document comes down to one idea. We are far from consensus on what chemicals to put into a frac. There is wide disagreement within an area, an operator, a pumper. The number of chemicals in an average frac is 32. There are only 4 chemicals that are in more than 50% of the 73K fracs. Four! There is tremendous room for innovation.