# Weather & products: summaries

**Agrochemical performance in the field**

Agrochemical products deal with all the environmental extremes that occurs in the field. Our client suspected high temperatures were inhibiting an herbicide, causing poor performance. A simple survey from growers confirmed their suspicions, but they needed hard numbers to improve their product.

Our client provided location information and end-user survey and we gathered the rest. Though temperature was important, our data scientists uncovered more questions about product performance in on farms in different states.

*Click for more on how we scoped this problem.*

**Digging deeper – providing straightforward answers from complex models**

With our eye for details, we noticed that product performance subtly differed across state lines. Higher temperatures clearly drove product failure in Nebraska, but the relationship was murkier in Pennsylvania.

We prioritize client time and comprehension. As the problem expanded in complexity, so did our models. We distilled nuanced results into simple highlights without neglecting the details. This helped us and our client to quickly zero in on a root cause they could act on.

*Click for more on how we report complexity with clarity.*

**Data engineering to uncover root cause**

Dense cloud cover blocks UV rays and impedes crop photosynthesis – both can impact product performance. Measuring cloudiness from raw imagery across many dates and sites required data mining and automation outside our client’s expertise.

We extracted cloud cover around each farm during product application. Combined with end-user surveys, this showed that the combination of cover and temperature was key to product performance in a humid east-coast state like Pennsylvania but not in dry and sunny western Nebraska.

*Click for more on how we used data engineering to extract cloud data*

# Subpage 1: Agrochemical performance in the field

**The problem: product failure**

Agrochemical products deal with all the environmental extremes that occurs in the field. High temperatures, cold winds, and record droughts prevent products from working as intended. Our client suspected high temperatures were inhibiting an herbicide, causing expensive wastage.

[image of frost damage]

The client sales team had collected simple survey responses from growers, recording if their product had worked or not. End-users were reporting more product failure in spring around days with extremely high maximum temperatures.

[salesperson surveying a farmer]

But was this really occurring in the field? And if so, at what temperatures? Our client's survey confirmed their suspicions, but they needed hard numbers to improve their product.

[flow chart of data being emailed or shipped]

**Our solution: enriching end-user surveys**

Surprisingly, that was all the data we needed from our client to start answering their question. The rest of the information we used was generated in-house by our environmental data scientists. Daily temperatures, rainfall, even cumulative degree days - all of these were leveraged to enrich the client's existing surveys.

[picture of someone working on a computer giving answers back to the salesperson. Lame but it gets the point across]

**Digging deeper**

We then used this information to pinpoint temperature thresholds where product failed, so our client knew when exactly their growers could spray to minimize waste. But how clear-cut these temperature thresholds were depended on state. We dug deeper to answer why.

(Links to other pages)

# Subpage 2

**Uncovering hidden complexities**

Our client’s initial hunch was that temperature alone drove product failure. This appeared broadly correct – the predictive power of temperature was good, but inconsistent. Our data scientists sliced through different customer segments in the end-user survey to find out why. What they found was that the impact of temperature on product performance depended on state.

[image of data segmentation/data scientist segmenting like a lege]

Breaking down the client’s question like this added tricky nuances to our results.

**Communication is key**

We value the time of our clients. Good product teams want to get into specifics, and the specifics are our strong suit. But “highlights” convey key results quickly without getting lost in the nitty gritty details.

[Picture of something techy like R code or reporting]

When reporting to clients, we like to walk through the models used to draw conclusions. Some model outputs are easy to interpret – like bar graphs or time series. But as problems become more complex so do the analyses and their outputs.

[pic of bar graph and trends]

One such tool is the binomial GLM, which used with yes/no data like our client provided. The outputs can be tricky to visualize, and tricker when different segments arise in the data such as geography. Our consultants worked closely with the client explain these outputs comprehensively, but clearly.

[picture of a consultant patiently explaining things to a stakeholder]

Here we illustrate how the binomial GLM worked using sample survey data, and how we distil things to a clear take-home message for stakeholders.

**A case study: leveraging simple yes/no surveys**

A client has provided us with 49 growers reporting success or failure of a product. We’ve enriched the survey with additional environmental data performed a statistical analysis. It looks like very cold temperatures negatively impact product performance.

[bar graph of how many yeses and how many nos]

But how can we know for sure? And when exactly is it too cold and when is it ok to use the product? We can just count the number of yes and nos. But that doesn’t tell us how confident we are, or where the temperature cut-offs lie.

**Solving the problem with binomial GLMs**

In Nebraska, the results are clear. The client can advise their end-users to save their herbicide at high temperatures as the risk-adjusted cost of *failing* to spray is very low. We’re very confident it would just be a waste.

[Image of a binomial glm ggplot output with perfect data that would give a p-value of 0.0001]

On the left side are product failures, on the right are product successes.

[Show a ggplot figure highlighting the curve]

This area is where we figure out what the thresholds are. The product starts to fail at xxF (35.6 C), and by xxF (37C) the product no longer succeeded.

On the other hand, we’re much less confident about how temperature impacts the product for growers in Pennsylvania.

[ggplot figure with a weak result]

With a confidence (measured with p-value) of 0.2, there’s a 20% risk that higher temperatures *don’t* affect the product. In other words, advising end-users to save their herbicide at high temperatures might lead to pest damage that could’ve been avoided.

This kind of uncertainty is ok – uncertainty is critical to know and allows clients to price risk, like the risk of telling end-users the wrong thing to do. If the cost of pest damage are high, then it might make sense to spray just in case. But if the cost of pest damage is low, the risk-adjusted cost of wasting herbicide might be higher.

[interaction of risk-adjusted cost of acting on a false-positive across a range of p-values]

**Distilling things down to simple insights**

Here we would suggest that our client provide confident advice to end-users in states like Nebraska with clear patterns in the data. Above xxF, don’t waste the product.

For end-users in Pennsylvania, their advice should be transparent but qualified. They wouldn’t want to bet the farm on it (no pun intended). A result like this justifies a closer look for the product team or maybe warning salespeople that there is a small risk of high temperatures messing up the product.

[image of a farmer weighing up a tricky decision]

Our client wanted to know *exactly* when their Pennsylvanian customers should spray, so we dug even deeper and found the answer in cloud cover.

(Links to other pages)

# Subpage 3: Cloud cover data engineering

**Why cloud cover mattered**

A key difference between growing in the moist East Coast and dry Midwest is cloud cover. Dense cloud cover blocks UV rays, but also starves fast-growing crop plants of light for photosynthesis. Both factors can affect agrochemical products.

Leveraging location information and end-user surveys provided by the client, we took care of the rest with our data mining expertise.

[stock image of sun shining through clouds onto a farm]

**What makes cloud cover tricky to measure?**

Cloudiness is easy to define but hard to measure with remote sensing. LandSat 8 is a satellite photography program that hosts images taken of the Earths surface. We can’t just download the photos – these are millions of high-resolution images.

To explore how cloud cover impacted product outcomes in different states, we needed to estimate cloud cover on farms for the entire week prior to product application.

[Landsat image]

**How we extracted cloud cover data**

First we had to limit measurements to each farm. Using GIS tools, we mapped the locations of farms and drew 5km circles around the properties. We then downloaded imagery for the week leading up to product application. With automation we scaled this process efficiently to gather data for every grower’s farm.

[Image of a cloud cover photo with a grid on it and a number]

We then calculated which proportion of the photograph was covered with thick clouds (% cover) using R. We took the average values for the week prior and *viola* – user surveys were enriched with weekly % cloud cover, specific to each and every farm that used our client’s herbicide.

[a short table showing % cover and product survey results]

**How we improved the end-user experience**

Cloud cover turned out to be key. While temperature explained product failure in drier Nebraska, we uncovered a more nuanced interplay of temperature and cloud cover in wetter Pennsylvania. With these insights our client could advise growers with different agricultural systems how to get the most out of their product.

(Links to other pages)