# Forecasting Case Study: ChampionX
## Using Machine Learning to Forecast Demand Despite Limited Data

**Executive Summary**

Using cutting-edge machine learning models and a wide array of forecasting techniques, SupChains delivered a forecasting model that helped ChampionX, an international chemical manufacturer, to **reduce their forecast error by 20%** (compared to a benchmark). These results were achieved despite the lack of historical data and the absence of demand drivers.

## Business Situation

Enjoying a dominating position in the chemical and oil industry, our client, ChampionX, manufactures and distributes a wide array of chemical agents, products, parts, and exploitation tools to other manufacturers. In 2021, ChampionX reported a total revenue of 3B$ doing business in over 60 countries with 7,000 employees.

Currently reviewing their inventory and forecasting practices, they have solicited our aid to increase their forecast accuracy while offering insights on the optimal aggregation level for forecasting demand. (SupChains also helped by training 25 planners worldwide on best practices for demand planning and inventory planning using two business simulation games.)

## Forecast Challenge

### Forecasting Hierarchies

To support its supply chain decisions, ChampionX needs to enjoy an accurate forecast at two different aggregation levels (or hierarchies):

1. [Aggregated] **Region x Chemical Agent** (~5,000 combinations) to support production planning.
2. [Detailed] **Sub-region x Product** (~10,000 combinations) to optimize inventory.

Using a specific granularity level to generate a forecast influences the accuracy on the various other levels. The end goal is to create an accurate forecast by Region x Chemical Agent and by Sub-Region x Product. So, the challenge for SupChains was finding which granularity level was the most interesting.

### Limited Historical Data

The client's ERP system only allows the conservation of historical sales for up to 3 years. This 36-month limitation is unfortunately only too common in supply chains and a significant drawback to achieving the best accuracy possible.[1]

---

[1] We greatly encourage all supply chain practitioners to keep as much historical data as possible and store it in a clean, consistent way. Applying data management best practices will show a return on investment in forecasting accuracy, as shown in our previous case studies (here and here).

Our client's market is the heavy industry, ChampionX being among the first links of complex global supply chains dealing in bulk volumes and greatly sensitive to the bullwhip effect. This behavior leads to erratic demand patterns that are especially difficult to interpret for forecasting models.

### Demand Drivers

Due to the nature of the chemical and oil industry, we couldn't leverage usual demand drivers such as pricing, marketing, or promotions to provide extra insights to our models.[2] In addition, the identified demand drivers that influence the market are macroeconomic and are unforecastable: We cannot use them to forecast future demand.

Moreover, due to the BOM complexity, we could not access historical inventory levels that could provide extra insights regarding shortages.

In summary, this dataset and forecasting setup is especially challenging for three main reasons:

- Multiple hierarchical levels
- (very) Limited historical data
- No demand drivers (promotions, prices, shortages)

## SupChains Solution

This forecasting challenge was one of the hardest we have had to face.

As such, to give ourselves the most chances of success, we have decided to make the best use of both statistical-based and machine-learning-based models. Optimizing each to the best extent, to finally compare their accuracy and bias to select the best solution.

### Forecasting Metrics

The dataset being erratic, we have chosen a simple combination of Median Absolute Error (MAE, or *forecast error* in the figures) and bias to assess the forecasting quality of our forecasts. This combination has the advantage of being simple to interpret while looking at accuracy and bias.[3]

### Data Cleaning (Product Transitions)

We identified product transitions throughout our intensive data-cleaning exercise. As illustrated in the figure below, when a new product replaces another one, our model can look at the historical sales of the former to the advantage of forecasting the new one.

---

[2] To see how our models deal with demand drivers such as promotions and pricing, see our previous case studies [here](#) (manufacturer with promotions only) and [here](#) (retailer with promotions and pricing).
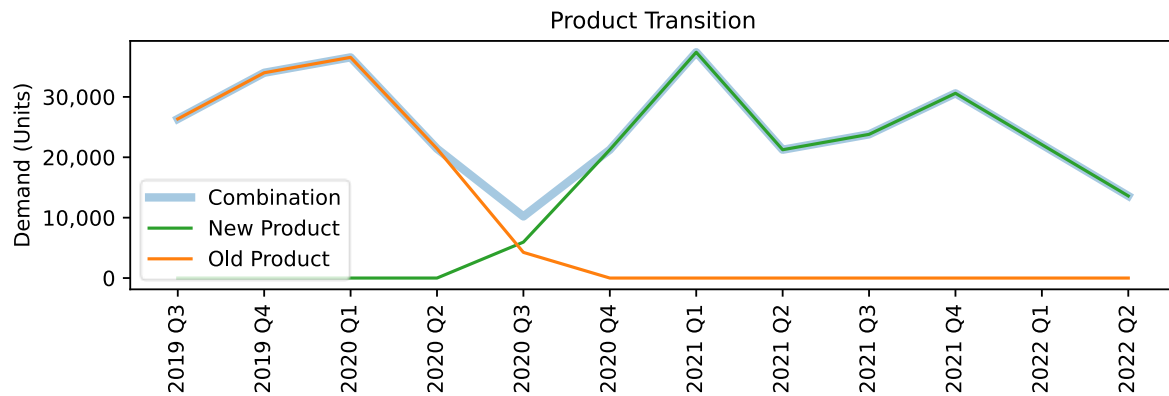
[3] More information about forecasting metrics [here](#).

*Figure 1 - Combination of the old and new versions of a similar product. The combined data will be used to forecast future sales.*

## Models and Technology

We developed two models to embrace these challenges and deliver an accurate forecast.

**Machine Learning**

The first one is a tree-based machine learning model (based on those introduced in *Data Science for Supply Chain Forecasting*). This model leverages the latest technology, such as CPU-multithreading and GPU computation capabilities, to **generate a forecast in 5 minutes** using any modern laptop.

**Statistical Models**

Our statistical models rely on three cutting-edge concepts:

1. Multiple Temporal Aggregation[4]
2. Advanced Seasonality Detection
3. Model Ensembling (multiple model aggregation)

and can deliver 10,000 forecasts in less than a minute.

## Results

As illustrated hereunder, our machine learning models delivered **an amazing 20% forecast added value compared to a benchmark** (a 6-month moving average).[5] This is more than twice the added value of our state-of-the-art statistical forecasting engine.

---

[4] https://researchportal.bath.ac.uk/en/publications/mapa-multiple-aggregation-prediction-algorithm

[5] We compute the added value as the % reduction of the scoring metric (the combination of MAE and Bias). More information about forecast value added here. For more info, see «*How do we select our benchmark ?*» and «*How do we test our models ?*» at the end of the document.
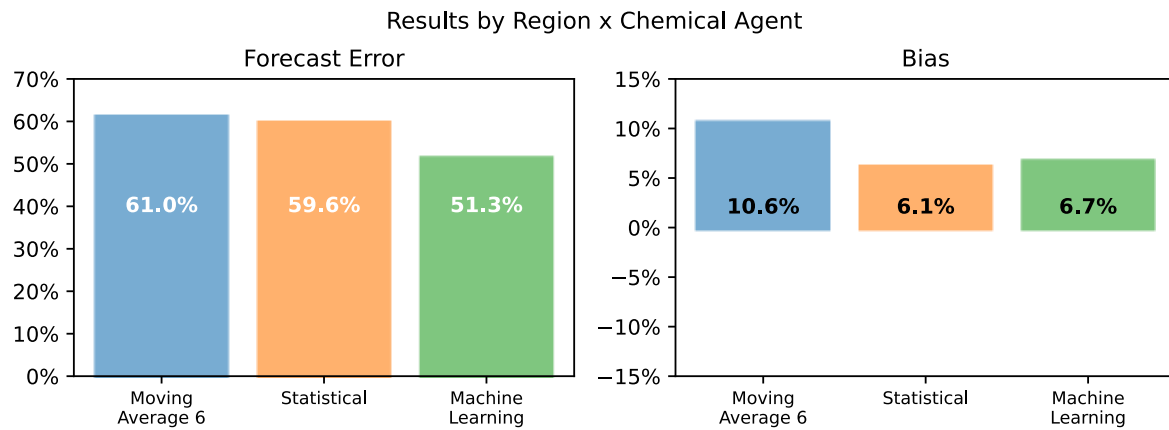
## Results by Region x Chemical Agent



*Figure 2 At the aggregated level, machine learning provides a 19% FVA compared to the benchmark (8% of FVA for the statistical model).*

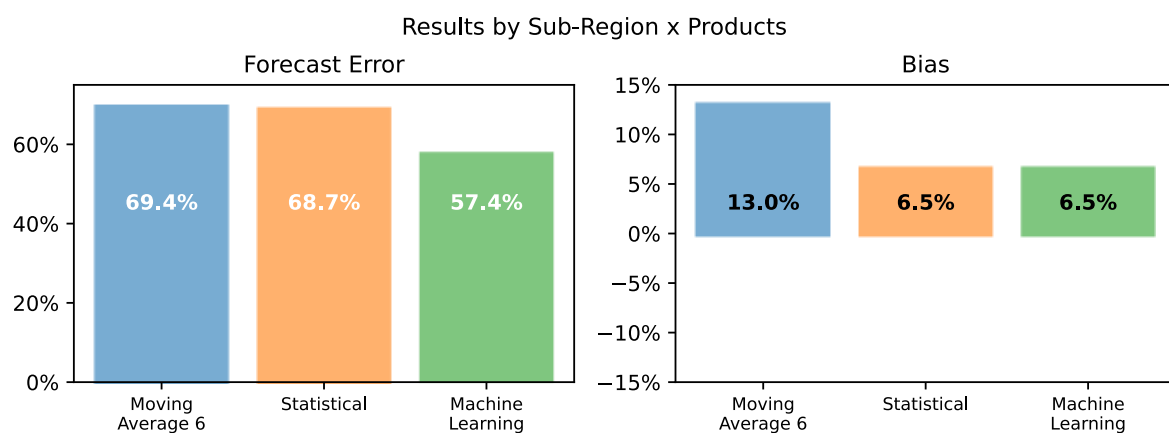## Results by Sub-Region x Products



*Figure 3 At the detailed level, machine learning provides a 22% FVA compared to the benchmark (9% of FVA for the statistical model).*

## Results over the Forecasting Horizon

As shown in the figure below, our machine-learning model delivered **accurate forecasts over the 6-month horizon**. The spread between the model and the benchmark even widens over time! Our model only loses around 1% accuracy per month, whereas the benchmark loses about 2%.

## Results by Region x Chemical Agent
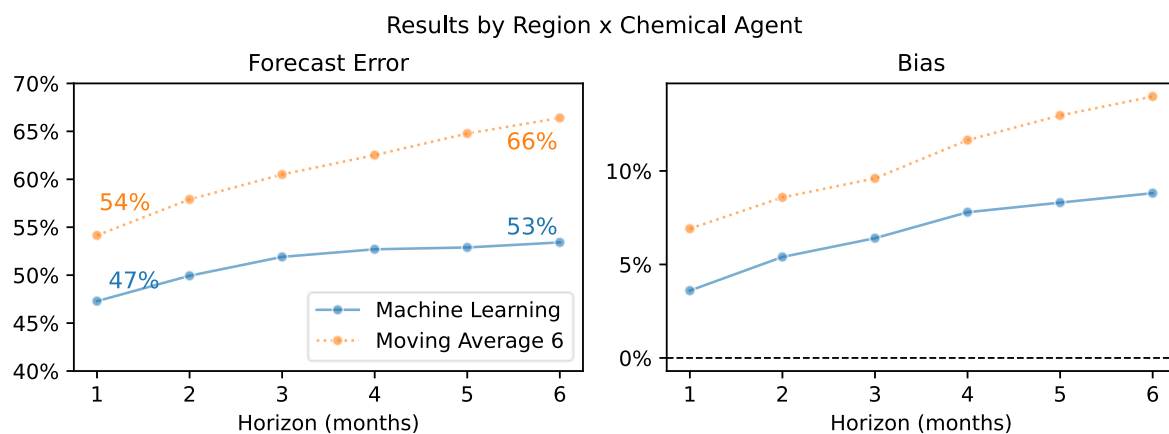
## Top-down vs. Bottum-up

Our two machine learning models delivered great added value on both granularities. Still, we want to use a unified forecast (*one-number forecast*) for both granularities. In other words, the aggregated forecast (Region and Chemical Agent) should reconcile with the detailed forecast (Sub-Region and Product).

To select the best model to use (between the aggregated and the detailed model), we ran a second batch of tests on both aggregation levels using top-down and bottom-up techniques, as highlighted in the figure below. Moreover, we added a third *combined* forecast by averaging the forecasts of the initial two models.
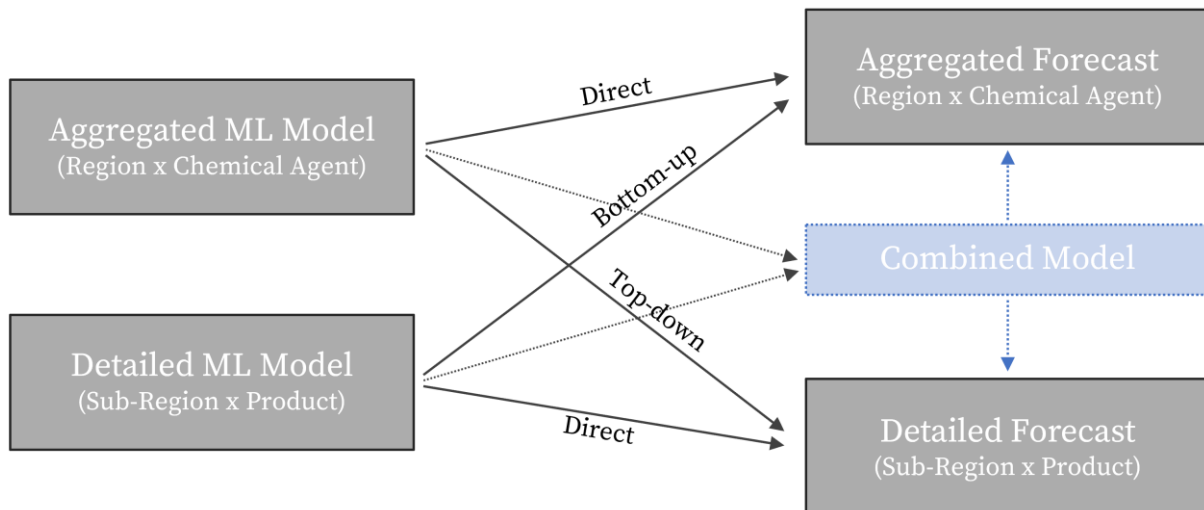


*Figure 4 Top-down and bottom-up aggregations to generate forecasts on different aggregation levels.*

We used 6-month rolling-horizon forecasts (the first one with only 24 months of history!) to generate our second batch of tests. For our machine learning models, such a low amount of historical data is similar to fighting with one hand attached to your back. Nevertheless, it also successfully delivered 20% added value compared to the benchmark.
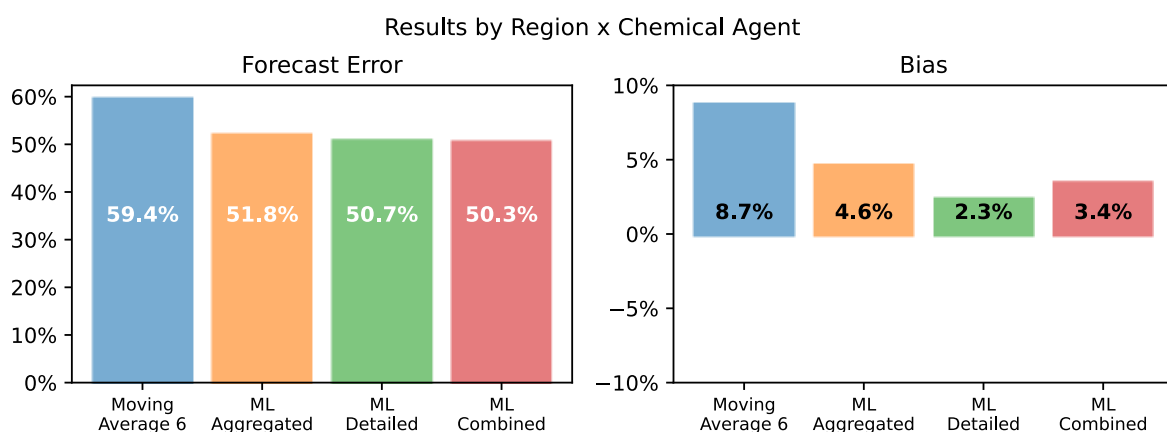


*Figure 5 At the aggregated level, the combined forecast displays an FVA of 19.4% (1.6% short of the bottom-up forecast).*
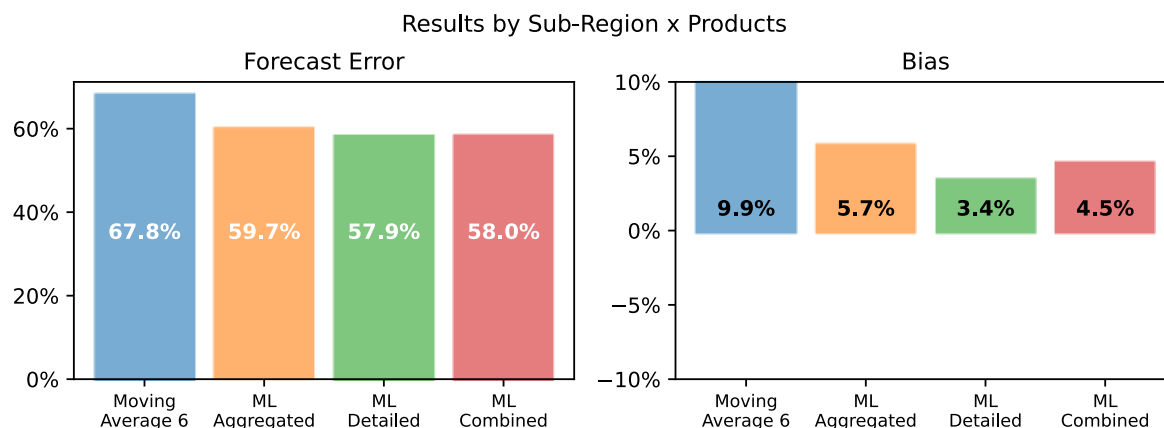
## Results by Sub-Region x Products



*Figure 6 At the detailed level, the combined forecast displays an FVA of 21.0% (1.1% short of the bottom-up forecast).*

As displayed in the two figures above, the top-down forecast (made at a detailed level) provides slightly more accurate predictions than the bottom-up and combined forecasts.

Nevertheless, we decided to use the combined model as a final model. Indeed, we have good reasons to think that this combination will provide better, more insightful results over time than a single model top-down or bottom-up model.

## Project Timeline

Three weeks of work were needed to gather the correct pricing, product transitions and construct a clean dataset that could be fed into the various models. Two more weeks were spent creating the models and testing features and ideas. A final week was dedicated to analyzing the models' performance and making the final report.

**Do you have a similar case?**
**Let's get in touch** to see how much value advanced forecasting can deliver to your supply chain.

**How do we select our benchmark?**
For multiple reasons, we do not recommend using external industry benchmarks when assessing forecasting accuracy (more info here). Instead, we compare our forecasting models against a moving average (and so should you!). In the past, simple naïve forecasts used to be considered suitable benchmarks. But they are too easy to beat. So instead, we look for the best moving average by trying out different horizons (3, 6, 12, and 24 months).

**How do we test our models?**
We follow data science best practices: our models are tuned, trained, optimized, and selected using a specific *training* dataset. Then, once we have chosen our favorite model, we assess it against a *test* set (that is entirely unknown to the model). By doing so, we ensure that the results obtained on the test set are as close as possible to real-life results.