If Data is the New Oil, Nowcasting is the New Drilling Equipment

Florian Thaler
Co-Founder and CEO, OilX[1]

Juan Carlos Rodrigues
Oil Economist, OilX

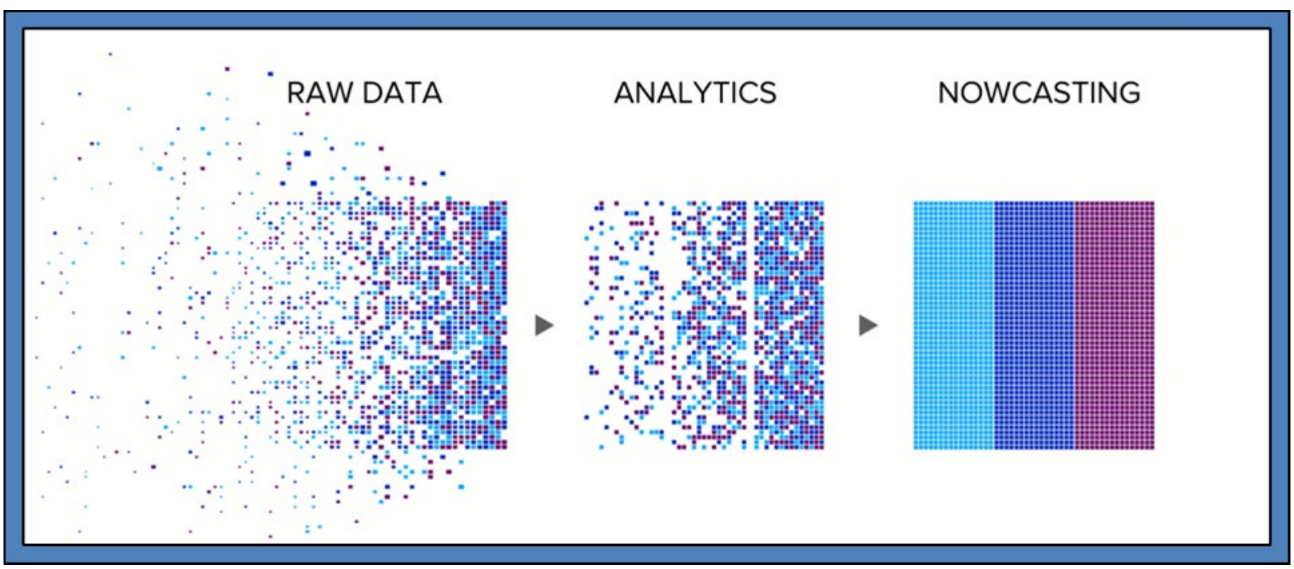
Bert Gilbert
Head of North American Business Development, OilX

Introduction

Traditional oil market analysis has often relied on applying classical statistical methods to historical data in order to identify possible patterns in the data that may have predictive power for relevant data points. However, this approach has a key limitation. The ever-changing structure of oil markets makes the detection of these relationships a moving target.

Because of the availability of new high-frequency data sets in oil markets, the phrase "data is the new oil" is commonly heard. Nevertheless, just as crude oil needs to be processed in order to be consumed, this new data lacks value without the proper refinement. Properly refined, these new data sets can be leveraged to generate something more reliable than a forecast, a "nowcast." See Figure 1.

Figure 1
Nowcasting Can Turn Big Unstructured Data into Valuable Market Insights



The rest of the article is divided into the following sections: first, a brief explanation of nowcasting is given, highlighting the benefits over forecasting. Secondly, a description on how nowcasting allows our company to generate a digital twin of the oil supply chain, expanding on the use of cargo tracking data to measure flows in real-time and geospatial imagery to measure stock changes. Finally, a conclusion is given on how nowcasting will continue to find its way in the energy markets.

What is Nowcasting?

Nowcasting is the prediction of the present, the very near future and the very recent past in economics and meteorology. The technique of nowcasting has been used in meteorology for a long-time. The term itself is a contraction of "now" and "forecasting" and refers to the utilization of readily available data sets to infer the current state of a variable. It is about predicting the present, the recent past and the near future. One can use this technique to estimate the global oil supply and demand in near real-time. Nowcasting models use unstructured data sets to make:

- *Direct measurements*: the target variable is directly observed (e.g., remote sensing via satellites of oil inventories or the digital twinning of the oil supply chain)

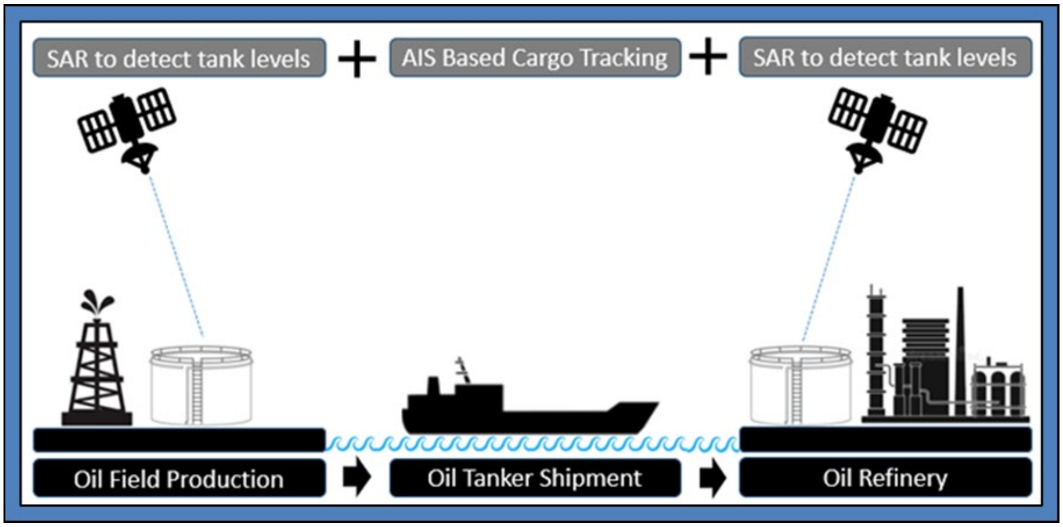
- *Short range predictions*: the target variable is not directly observed (e.g., Apple Mobility Data to infer U.S. gasoline demand)

As López de Prado and Lipton (2020) note, the advantages relative to forecasts are the following: direct measurements always hold true as they do not rely on a statistical lead/lag relationship. Short-range predictions are far more stable than long-range predictions.

A Digital Twin of the Oil Supply Chain

In aggregate the oil market functions like a giant bathtub: supply comes into the market from the faucet and exits through the drain, the demand. Too much supply relative to demand and the bathtub fills up; too little supply relative to demand and the bathtub drains. This flow sets the price. To go into a bit more detail, oil is produced and moved to some type of storage tank. Ultimately, the oil is moved to a user. The mode of transportation could be by truck, rail, water, or by pipeline. The oil is generally transported to a refinery that converts crude oil into a more valuable product such as gasoline. This is happening at a global scale. Figure 2 on the next page illustrates the flow of oil, including its tracking, from production through shipment to storage at a refinery.

Figure 2



Note: SAR stands for Synthetic Aperture Radar; and AIS stands for Automated Identification System.

Because of these continuous transit and storage phases, market participants with large assets through the supply chain, physical oil traders, have historically held an advantage in estimating if the bathtub is filling up or draining. However, in recent years this advantage has waned as new technologies are democratizing access to the otherwise rare data, allowing more players to understand the physical state of the oil market. Coupling these new data sources with other historical data sets through sophisticated machine learning algorithms can overcome the challenges of scale and allow a company to replicate digitally the global crude supply chain with nearly real-time observations.

At our company, there are two main alternative data sets that work as direct measurements, which allow us to come up with short-range predictions of the current status of the global crude markets: cargo tracking and geospatial imagery.

Cargo Tracking

The base data set involved in any effort to track the movement of oil by tanker is the Automated Identification System (AIS). AIS is a platform by which vessels communicate their location and other critical information such as current course and speed. In a way it could be compared to the Global Positioning System (GPS) system used in phones whenever we use a navigation application. AIS identification is required for many ships by the International Maritime Organization as well as other organizations such as the U.S. Coast Guard. While there are certainly exceptions and some vessels do not comply with this rule, AIS provides a massive amount of data on the movement of the world's vessels – many of which are transporting crude oil or refined products. This data is collected by receivers located around ports, but given that the range of an AIS signal is only about 50 nautical miles, satellite constellations are also used to collect AIS signals from vessels outside of the range of a terrestrial transponder. We leverage AIS data from both terrestrial and satellite sources and together with the maritime technology partner, Signal Ocean, we track over 3,200 vessels on a daily basis.

As mentioned above, although AIS provides a significant amount of information, the picture provided by AIS data is often incomplete. For instance, some vessels may not have an active transponder, either intentionally or caused by a malfunction. Moreover, AIS is not an entirely automated system, and some data such as the depth of a vessel in the water (draft) and destination are updated manually and sometimes prone to inaccuracies or input mistakes.

This inherent imperfection in the data causes challenges for those tracking the vessel's cargo. Therefore, just as crude oil needs to be refined to be consumed, in order to leverage this data it is necessary to translate a series of vessel locations into a history of standard actions that may be understandable by a computer algorithm.

Our methodology to purify this data is the following: firstly, we map every vessel's location to a large layer of infrastructure allowing us to generate a history of potential loadings and discharges. Next, we compare this with other vessels in order to identify possible ship-to-ship transfers, something that is especially relevant to understand U.S. crude exports.

At this point, new challenges arise from the complexity and granularity in the movement of cargoes, specifically from docks that load multiple types of cargo such as those that load refined products. To sort out these intricacies we combine additional data sets such as import manifests and Signal Ocean's deep database of shipping chartering details and fixtures. These data sets provide critical context for the AIS data and can tell which products a vessel was chartered to carry, the dates of that charter, the involved parties, and the origin and destination of the vessel.

By combining these different layers of data, we can accurately understand active cargo movements as well as make inferences about future movements based on historical actions. Additionally, we can predict which crude grades load from which docks and infer possible destinations.

To sort out these intricacies we seek to add additional context. For this, we rely on U.S. import manifests reported daily, port reports and Signal's fixture database, which contains multiple sources of fixture data. By assigning the origin, destination, quantity, and grade to each vessel movement we are then able to provide a real-time view of how much oil has been exported by a specific port. From a port level, we can roll up to a country level and understand flows between countries or regions. We then compare our estimates to national statistics to test and fine-tune our model. Finally, on top of all this, an additional layer of data is added: country specific customs data provided on a lagged monthly basis. By doing this regularly we are able to validate our results and also ensure that we are reacting and adjusting for changes in infrastructure.

Oil Inventories

The second alternative data set that has become widely used in the oil market is geospatial imagery. Satellite imagery was once only available to the government or large corporations, but this data is becoming increasingly available and on a more frequent basis. The eye in the sky is most useful for tracking oil in storage. Examples of satellite imagery for oil in storage are provided in the next two pages in Figures 3 through 5.

Figure 3
Tank Terminal: High-Resolution, Low Frequency Optical Image on the Left and Low-Resolution, High-Frequency Image on the Right

Since a large percentage of crude oil is stored in tanks with floating roofs, we can use the depth of the tank's roof to estimate the amount of oil being stored in the tanks. For estimates of inventory in floating roof tanks, two types of satellite imagery are generally used: optical and SAR (Synthetic Aperture Radar) images.

Figure 4
Tank Terminal: Optical Image on the Left and SAR Image on the Right

Optical satellite imagery is similar to traditional pictures, only taken from space. However, standard optical cameras cannot penetrate dark skies or dense clouds. Therefore, the use of SAR technology becomes essential. Unlike optical images, SAR signals are not weather nor light dependent.

We leverage the imagery provided by the Sentinel constellation of the European Space Agency (ESA) whose satellites complete a scan of Earth every six days. Given the vast number of tank farms and the dispersion of their location, machine learning algorithms are needed to accurately translate both images into an estimated level of storage.

Figure 5
Cushing, Oklahoma Tank Storage: Automated Optical Tank Detection

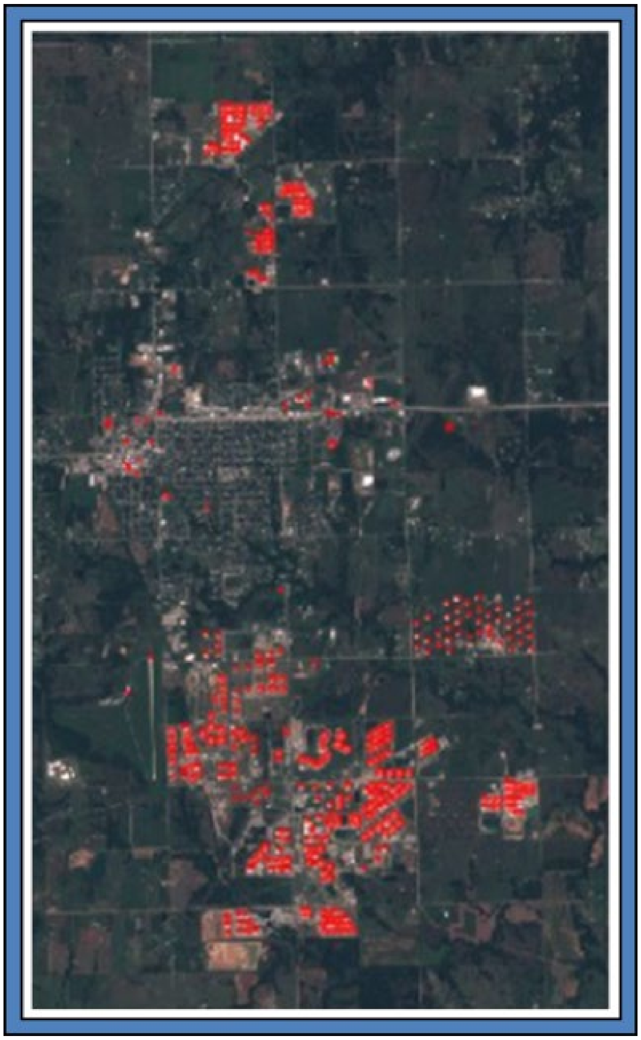
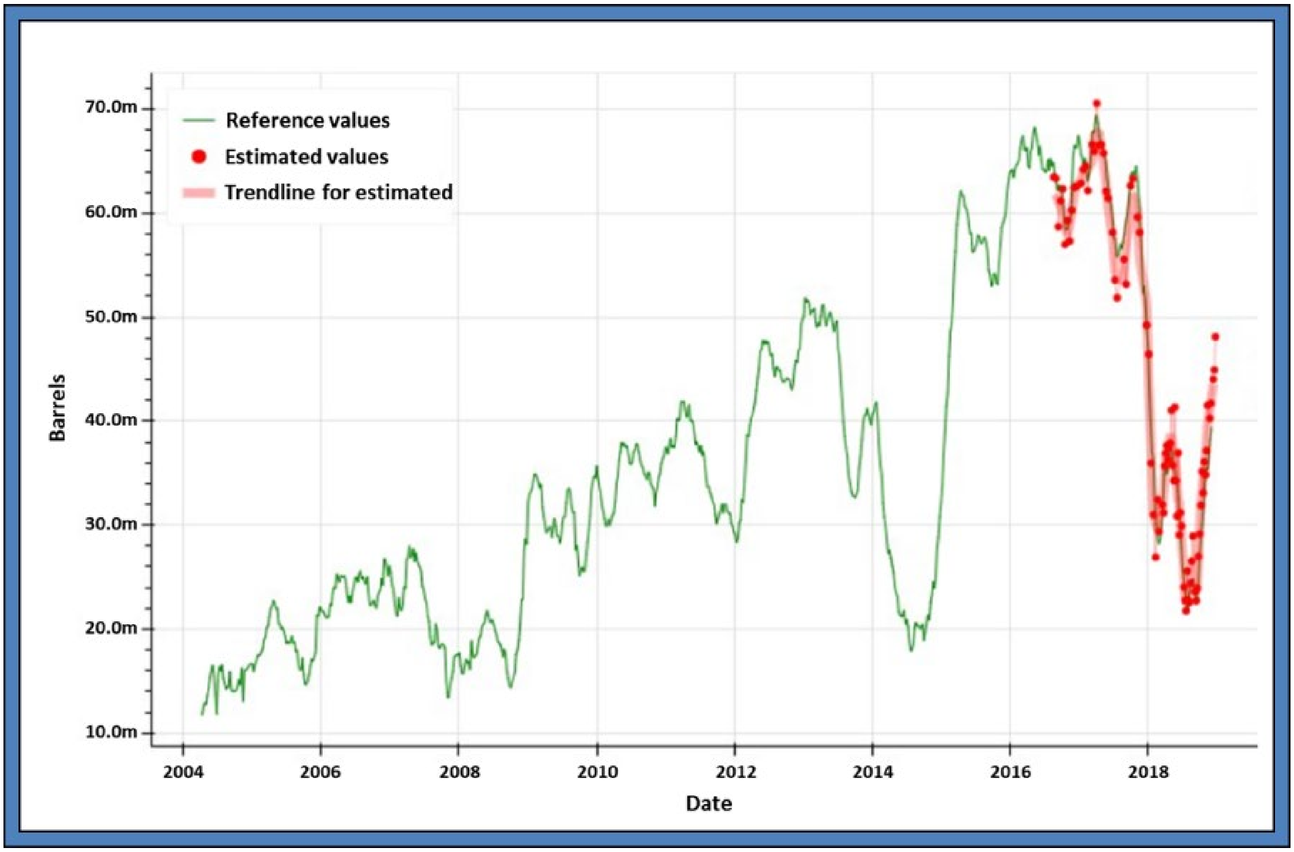


Figure 6
Cushing Inventories: Time Series Modeling vs. EIA Data



Note: The U.S. Energy Information Administration (EIA) data is represented by the green line while the estimated values from our algorithms are depicted with the red dots and solid pink line.

Our machine learning algorithms identify the position of a tank's roof relative to the rest of the tank effectively allowing us to estimate the amount of oil stored inside every tank. Figure 6 above provides an example of our algorithms' estimated values for Cushing, Oklahoma storage relative to U.S. Energy Information Administration data.

Field expertise is crucial in order to point the algorithm to the right tank farms. Our oil experts have access to a global network of oil logistics experts to assist with model calibrations. For instance, some tank farms in the Middle East are partially buried into the ground for extra security and additional calibrations are needed to enhance the accuracy of the algorithm. Furthermore, in densely urbanized areas, some industrial tanks holding other liquids can be mistakenly classified as crude tanks causing incorrect readings.

Where Do We Go From Here?

Oil cargoes and oil inventories are just two examples of how oil market participants are leveraging the vast amount of data now being generated to create a digital view of the world. By employing a "nowcasting" methodology in conjunction with this data, we are able to achieve a far better and accurate view of what the current state of the market is. This provides an opportunity for efficiency gains across the industry.

Finally, we are continuously looking for new and different sources of data to further our understanding and provide insights into areas that previously relied on traditional forecasting. As we build out our understanding of different pieces of the oil market, we can gain a better understanding of what is happening in the whole. We can create a picture of the market, and this allows us to understand aspects for which data might not exist.

Endnote

1 Established in 2018, OilX combines the latest in advanced data science frameworks with extensive oil analytics knowledge to create the first real-time supply-demand balance tool. Headquartered in London, OilX has more than 15 years of oil analytics experience across oil majors, investment banks and hedge funds at the highest level. For more information, please visit www.oilx.co.

Reference

López de Prado, M. and A. Lipton, 2020, "Three Quant Lessons from COVID-19 (Presentation Slides)," March 27. Accessed via website: https://ssrn.com/abstract=3562025 on October 9, 2020.

Author Biographies

FLORIAN THALER
Co-Founder and CEO, OilX

Mr. Florian Thaler, co-founder and CEO of OilX, brings together a wealth of experience analyzing the oil markets from different perspectives at an oil major, investment bank and hedge fund (Shell, Citi, and Och-Ziff respectively). He was also a member of the 100 Future Energy Leaders of the World Energy Council and is currently serving as a board member at Bloc-X and Enerjen Capital.

JUAN CARLOS RODRIGUEZ
Oil Economist, OilX

Mr. Juan Carlos Rodriguez was the Head of the FX & Commodities Desk at the Central Bank of Mexico where besides investing the reserves portfolio worth over $175 billion, he led the execution of the world's biggest oil hedging program. He now works as an Oil Economist at OilX where his focus is on generating market research and combining fundamental and systematic trading strategies.

BERT GILBERT
Head of North American Business Development, OilX

Mr. Bert Gilbert was the Americas Oil Market Specialist at Bloomberg for five years and later ran the oil focused research for Bloomberg New Energy Finance. Prior to these roles, he held a business development role selling Bloomberg licenses to oil market participants in Calgary and Houston. Prior to joining OilX, he most recently worked on developing a vessel tracking platform for Enverus.