Estimating life cycle carbon emissions of the global oil supply chain at a high-resolution using optimization in a network model

by

Yash Dixit

B.Tech., Indian Institute of Technology, Madras (2017)

Submitted to the Institute for Data, Systems, and Society

and

Department of Electrical Engineering and Computer Science

in partial fulfillment of the requirements for the degrees of

Master of Science in Technology and Policy

and

Master of Science in Electrical Engineering and Computer Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2021

Author_____

Technology and Policy Program
and
Electrical Engineering and Computer Science
May 13, 2021

Certified by_____

Steven Barrett
Professor of Aeronautics and Astronautics
Thesis Supervisor

Certified by_____

Arvind Satyanarayan
Assistant Professor of Computer Science
Thesis Reader

Accepted by_____

Noelle E. Selin
Director, Technology and Policy Program
Associate Professor, Institute for Data, Systems, and Society and
Department of Earth, Atmospheric and Planetary Sciences

Accepted by_____

Leslie A. Kolodziejski
Professor of Electrical Engineering and Computer Science
Chair, Department Committee on Graduate Students

Estimating life cycle carbon emissions of the global oil supply chain at a high-resolution using optimization in a network model

by

Yash Dixit

Submitted to the Institute for Data, Systems, and Society
and
Department of Electrical Engineering and Computer Science

on May 13, 2021, in partial fulfillment of the
requirements for the degrees of

Master of Science in Technology and Policy
and
Master of Science in Electrical Engineering and Computer Science

Abstract

Climate change, being the multi-faceted problem that it is, requires aggressive decarbonization across the entire life cycle. With the evolving energy mix, the oil industry is in a phase of adaptation. At present, petroleum fuels account for a third of the global primary energy supply. Future forecasts range across a spectrum from plateauing to decreasing supply, up to a 40 percent decrease from present levels [1]. Furthermore, certain applications such as aviation and petrochemicals have limited short-term, scalable alternatives. On this backdrop, there is an increasing push for better emissions reporting throughout the supply chain and regulatory mandates at making climate friendly choices. Notable examples include the Low Carbon Fuel Standard by the California Air Resources Board [2] and the Fuel Quality Directive by European regulators [3]. Existing literature is directionally aligned with these efforts, in that it points towards carbon accounting in the supply chain. However, studies are either limited to specific processes (e.g: crude oil extraction) and/or regions (e.g: North America). Furthermore, those with a wider scope including all phases of the supply chain, have a poor resolution whereby the carbon accounting is done at the level of countries and is thus unable to capture the complexities associated with oil trade. These inadequacies stem from poor availability of data and methodological challenges which fail to accurately portray the heterogeneity in life cycle emissions. The thesis quantifies this heterogeneity using a market-based approach that addresses the aforementioned limitations by estimating the life cycle carbon intensity of crude oil trades from sources (oil fields) to destinations (refineries). With a scope that includes crude extraction and transportation, the emission modeling is undertaken using high-fidelity commercial datasets, existing emission estimators and computational techniques based on optimization. The thesis concludes that globally, the carbon footprint variability ranges from 1.80 to 32.92 gCO_2/MJ with a volume weighted mean of 9.73 gCO_2/MJ. This variability coupled with supply forecasts up to 2050 from low-carbon scenarios amount to additional CO_2 savings of 2-5 GT.

Thesis Supervisor: Steven Barrett
Title: Professor of Aeronautics and Astronautics

Acknowledgments

I am deeply grateful to my advisor Steven Barrett for his support throughout the course of this work. Steven's leadership has been instrumental in not only making my research more intentional but also for facilitating my professional growth.

My work would not have been possible without the guidance of my research mentors Ray Speth and Mark Staples - their analyses, feedback and ideas have been pivotal to my contributions.

I would also like to thank Arvind Satyanarayan for introducing me to the world of visualization and helping me appreciate its amplifying power in communicating my research to the world.

I have been fortunate to be supported by my lab-mates at LAE. This opportunity of being able to learn from them and to collaborate with them has been one of the most intellectually enriching experiences of my life.

It has been a privilege to be a part of the TPP family. The entire TPP community has made these few years incredibly rewarding by being a source of inspiration and joy.

Lastly, I would not be here without the anchor of family and friends. I am who I am because of them and I owe everything I have to their sacrifices and unconditional support.

THIS PAGE INTENTIONALLY LEFT BLANK

Contents

List of Figures

List of Tables

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 1

Introduction

Petroleum-derived fuels currently account for approximately 31 percent of primary energy supply and are expected to continue to make up 22-27 percent of global primary energy by 2040 [1, 4]. At the same time, petroleum fuel combustion is responsible for approximately 34 percent of annual global greenhouse gas (GHG) emissions, with crude oil extraction, transportation, and refining operations adding a further 9 percent of GHG emissions [1]. Therefore, while petroleum will continue to play a significant role in the global energy mix, it is becoming increasingly important to identify opportunities to reduce GHG emissions at all stages of the transportation fuel life cycle.

Stakeholders are beginning to address the sector's emissions. Regulators have shown meaningful policy intent to incentivize low-carbon practices [2, 3], private investors are beginning to consider climate-related risk in oil investments and the industry as a whole is grappling with shareholder pressure to make climate-friendly choices in the ongoing energy transition [5]. Despite the intent, such efforts have struggled with methodological and data challenges. The global complexity of the supply chain results in loss of information along the life-cycle thereby hindering end-to-end visibility in the supply chain (i.e. from oil fields to refineries). Furthermore, regional heterogeneity leads to data gaps which in turn lead to emission estimates that are globally averaged.

This thesis presents a high-resolution life cycle CO_2 assessment that solves these aforementioned challenges using high-fidelity commercial datasets, bottom-up emission estimators and

computational techniques based on network modeling and optimization.

This section starts with an overview of the oil supply chain and introduces key terminologies that are used throughout the research. This is followed by the analysis of existing research literature that shows the methodological shortcomings of current studies which in turn points toward modeling solutions that the thesis presents.

1.1 Overview of the supply chain

The oil supply chain comprises of three key stages:

- *Upstream* - extraction of crude from oil fields
- *Midstream* - transportation of crude oil via pipelines, rail, trucks and tankers
- *Downstream* - refining of crude oil at refineries

Crude extraction typically entails drilling the well to extract hydrocarbons from the reservoirs, processing them, and in some cases uses enhanced recovery techniques that pump water or gases into underground cracks. In a few cases such as the Canadian oil sands, additional processes such as pyrolysis are undertaken to effectively extract the useful hydrocarbons from the reservoir [6, 7]. Thus, extraction operations are energy intensive and represent the first key source of CO_2 in the life-cycle.

The extracted crude is stabilized and blended to form "crude blends" that are marketed and sold to refineries. The formation of crude blends is a key input for the life-cycle assessment since the crude blends act as the identifying signature for oil barrels as they move in the supply chain (e.g: barrels of the blend "Arab Light" bought by the Jamnagar refinery in India).

The crude blends are transported to their destinations i.e. refineries via pipelines, rail, trucks and tankers depending on the producer and consumer countries. The transportation operations (midstream) thus represent the second source of CO_2 in the life-cycle.

Post transportation, the crude blends are refined to form petroleum products such as gasoline, jet fuel, etc. and this third stage (along with the transportation of refined products) is the penultimate piece of life-cycle CO_2 emissions (the last stage being combustion).

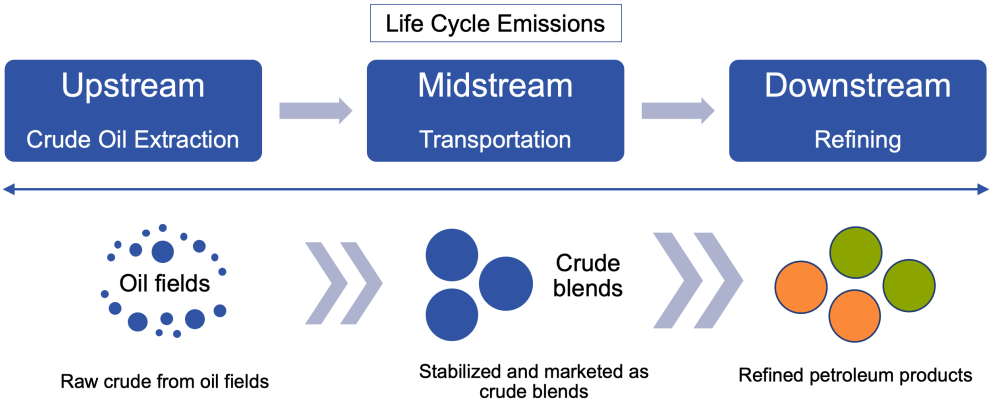

Figure 1-1: The oil supply chain segmented into three stages: Upstream, Midstream and Downstream

1.2 Analysis of the literature

The open-source Oil Production Greenhouse Gas Emissions Estimator (OPGEE) model has enabled a bottom-up estimation of carbon intensities associated with crude oil extraction [8]. This has led to the disaggregation of crude oil carbon intensities in different global markets nearly a decade ago [9, 10].

OPGEE, in conjunction with proprietary data about upstream operations has been used to estimate the global carbon intensity of crude extraction [9, 11]. Although the study generated emission estimates at the level of oil fields, it lacked the market data to link its findings with refineries. Consequently, industry and policymakers have been limited to either a field level picture or a country-aggregate picture. As this study provided the foundation for field-level emission estimation, this limitation continues to reflect in related future work that focuses on specific regions (e.g: China) [12]. A core missing component is the mapping of carbon intensities to the blend level which requires knowledge of the supply chain and a systematic approach towards estimating the blending process.

Furthermore, in both the above studies, transportation carbon intensity is set to a default baseline value based on models such as GREET (Greenhouse gases, Regulated Emissions, and Energy use in Transportation) [13]. Thus, in addition to the limitation of resolving emission estimates at the level of blends, there is a lack of high-resolution quantification of the heterogeneity in transportation emissions.

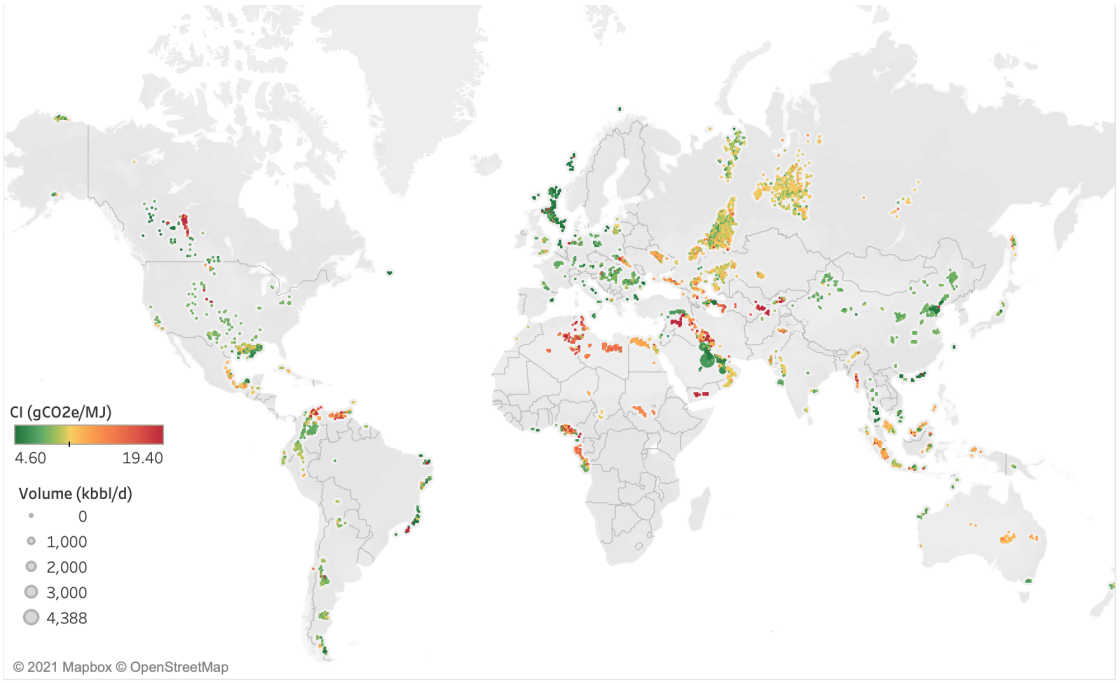

Figure 1-2: Global upstream carbon intensities - based on field-level results generated using OPGEE [10]

This heterogeneity is explored by Choquette et al. [14] using an emission estimator based on hydrodynamics to estimate CO_2 emissions associated with crude pipelines in Canada. Although the model is scalable, the study is limited to Canada due to data availability (pipeline locations, design specifications, etc.).

Literature in the domain of transportation life-cycle analysis has examined shipping as a sector and specifically the activity of crude oil shipping as a source of CO_2 emissions. Several studies [15–17] have estimated overall emissions associated with crude tanker activity. While these studies have global coverage, their granularity is limited to the level of trade lanes or regions (for example: Middle East to Asia). This prevents emission attribution to the different types of crude blends - for example, the carbon intensity of the "Middle East to Asia" shipping lane has lower resolution than the carbon intensity of the shipping route Basrah (Iraq) to Mumbai (India) handling the crude blend "Basrah Light".

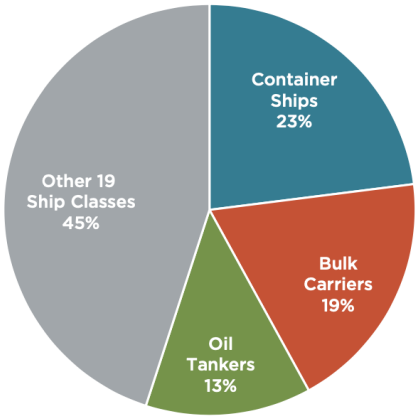

Figure 1-3: Global shipping emissions 2013-15 broken down by source [16]

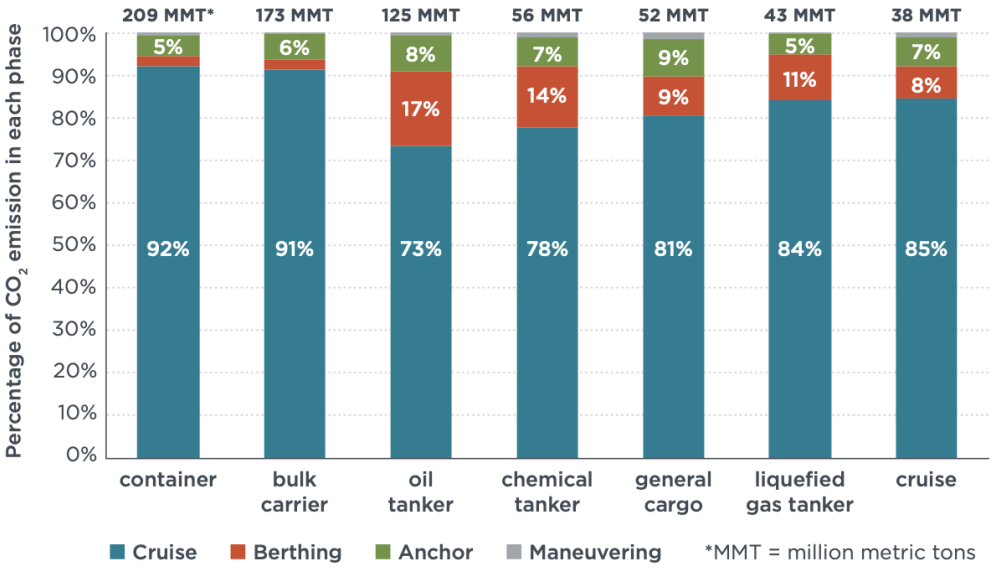

Figure 1-4: Global shipping emissions 2013-15 based on type of shipping activity [16]

The table below summarizes the analysis of four key studies in the literature:

Table 1.1: Studies with varying scopes estimating life cycle carbon emissions in the oil supply chain

Study	Scope	Comments
[10] Masnadi et al. "Global carbon intensity of crude oil production"	Global, Upstream	High-resolution analysis limited to crude oil production
[12] Masnadi et al. "Well-to-refinery emissions and net-energy analysis of China's crude-oil supply"	China, Upstream and Midstream	High-resolution analysis for the upstream, baseline defaults for the midstream
[14] Choquette-Levy et al. "COPTEM: A Model to Investigate the Factors Driving Crude Oil Pipeline Transportation Emissions"	Canada, Midstream	Bottom-up emission quantifier (based on flow hydrodynamics) for crude oil transportation
[18] Bergerson et al. "PRELIM: the Petroleum Refinery Life Cycle Inventory Model"	Global, Downstream	Representative analysis based on static refinery configurations

1.3 System boundary of the life-cycle analysis

The review of existing literature points towards two core areas of improvement in the assessment of global life-cycle carbon emissions :

1. High-resolution CO_2 assessment, down to the level of individual crude blends broken down into granular supply chain pathways
2. Estimating transportation emissions i.e. those associated with pipeline, shipping transport

Addressed with high-fidelity data discussed in Chapter 2, these aspects underpin decarbonization policy which shapes incentives for real-time carbon reporting and mitigation credits.

To that end, the system-boundary of this research is the upstream and midstream i.e. crude extraction and transportation. More importantly within this system boundary, the thesis preserves the high resolution and complexity of the supply chain by considering field level emissions and granular transportation pathways (e.g.: pipeline routes). As a consequence, this approach quantifies the significant heterogeneity in CO_2 emissions across the life cycle.

Limited by data, the study addresses pipeline and shipping transport while excluding trucking and rail.

1.4 Structure for the rest of this thesis

As detailed in this section, the thesis introduced the oil supply chain and the methodological shortcomings of current life cycle assessment, notably the role played by high fidelity data and modeling granularity at the level of oil-fields and individual transportation pathways.

Chapter 2 describes the data sources used in the research including but not limited to geospatial data and asset-level operations data. This is followed by the methods in Chapter 3 that describe how the data is used to represent the supply chain and estimate emissions.

After detailing the analyses, the thesis then highlights the heterogeneity in carbon emissions across the supply chain at different levels of aggregation (e.g.: field, blend, country, etc.) in Chapter 4. These results are used as the foundation for decarbonization policy opportunities described in Chapter 5.

The thesis ends with the conclusion that distils keys insights from the research and highlights directions for future work.

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 2

Data Sources

In order to achieve the well-to-refinery-gate coverage (i.e. fields to refineries) in the supply chain at a high resolution (i.e. at the level of oil fields and transportation pathways), the research uses high-fidelity data sources including, but not limited to:

- *Geospatial data*: location of supply chain assets such as oil fields, shipping terminals, pipelines, refineries
- *Crude trades data*: market data mapping crude blends to refineries
- *Shipping routes*: timestamped locations of crude tankers
- *Asset characteristics*: properties such as production volumes of oil fields, specifications of pipelines
- *Miscellaneous*: public datasets of ambient temperature, elevation

The study sources datasets from a range of categories - commercially available, publicly available and published literature. The commercial data sources include Wood Mackenzie, Kpler, GlobalData and IHS Markit [19–22], provided by external collaborators and classified as confidential.

To ensure integration across different modules, all data sources are 2015-based. Relying on sources with specific expertise ensures that the models get the most suitable data along different dimensions of the supply chain.

A high-level overview of the data sources segmented by source is given below:

- **Wood Mackenzie**: geospatial locations of upstream and downstream assets, properties of crude oil at the field level and blend level
- **GlobalData**: geospatial locations and specifications of midstream assets (e.g: pipeline locations and diameters)
- **Kpler**: crude trades data linking export terminals to import terminals
- **IHS Markit**: shipping tanker locations and vessel characteristics

2.1 Upstream - Crude oil production

The main component of the upstream data is the Wood Mackenzie crude production dataset. This includes the following features for all global oil fields:

- geolocation and asset name
- asset country
- production volumes
- properties of produced crude (density measured in API, sulfur content)

The upstream data is supplemented with the corresponding production carbon intensity data from Masnadi et al. [10]. As described in Chapter 1, in the section reviewing existing literature, the carbon intensity data is based on OPGEE (Oil Production Greenhouse Gas Emissions Estimator) [8]. This estimator uses a bottom-up approach to calculate the emissions associated with the extraction of crude oil. Sample oil fields in Norway with production volumes >50 k-barrels/day are shown in figure 2-1.

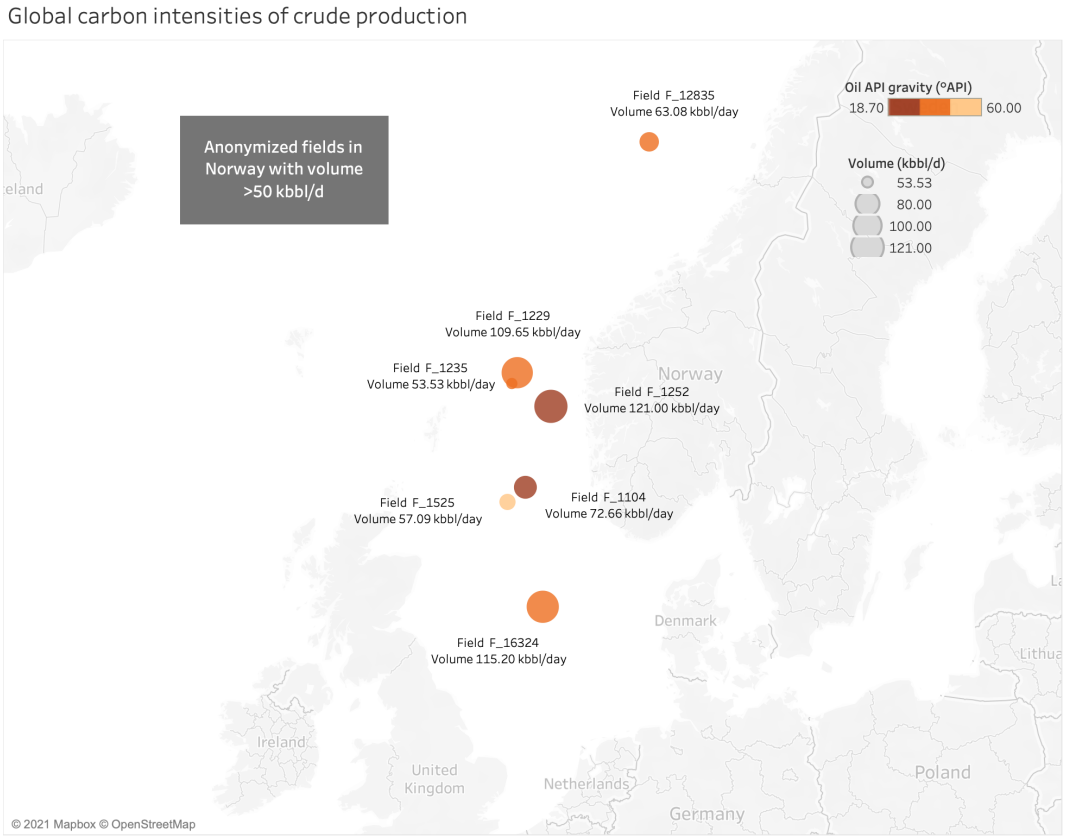

Figure 2-1: Oil fields in Norway with production >50 k-barrels/day (kbbl/d)

2.2 Market trades

The proprietary market trades data provided by Wood Mackenzie [19] has three key components:

1. Specifications of global marketed crude blends
2. Demand data from refineries consuming the blends
3. Geolocation data and asset identification of refineries

The specifications of global crude blends are key, particularly in conjunction with the upstream

data. This feature is used in the methodology described in Chapter 3 to estimate how blends are formed from source oil fields. Figures 2-2 and 2-3 show the specifications for a sample set of oil fields and blends in Iraq.

Field_ID	Country	Oil API gravity (°API)	2015 Oil and lease condensates (kbbl/d)
Field_7598	Iraq	34.00	782.01
Field_7601	Iraq	34.00	577.72
Field_9862	Iraq	28.45	384.00
Field_5310	Iraq	24.00	206.00
Field_9856	Iraq	24.00	188.80
Field_9859	Iraq	24.00	188.20
Field_3590	Iraq	28.00	179.05
Field_10444	Iraq	35.00	171.17
Field_177	Iraq	25.00	144.20
Field_10441	Iraq	35.00	140.83

Figure 2-2: Sample of the oil fields data with key features - Top 10 oil fields in Iraq by volume

Crude Stream	Source Country	API [-]	Sulfur content (wt%)	Sum Volume 2015 (b/d) values
Basrah Blend	Iraq	31.10	2.62	58336
Basrah Heavy	Iraq	23.70	4.12	344964
Basrah Light	Iraq	28.89	3.19	2278095
East Baghdad	Iraq	23.00	0.00	8573
Kirkuk Blend	Iraq	34.20	2.22	427847

Figure 2-3: Sample of the crude blends data with key features - blends from Iraq

The downstream data i.e. mapping of crude blends to refineries is later used to create end-

to-end traceability in the supply chain (linking oil fields to refineries through information about crude blends). The geolocation data of refineries is coupled with geolocation data of other assets such as fields, pipelines, shipping terminals to model the supply chain as will be described in Chapter 3. Sample refineries in India with locations and production volumes are shown in figure 2-4.

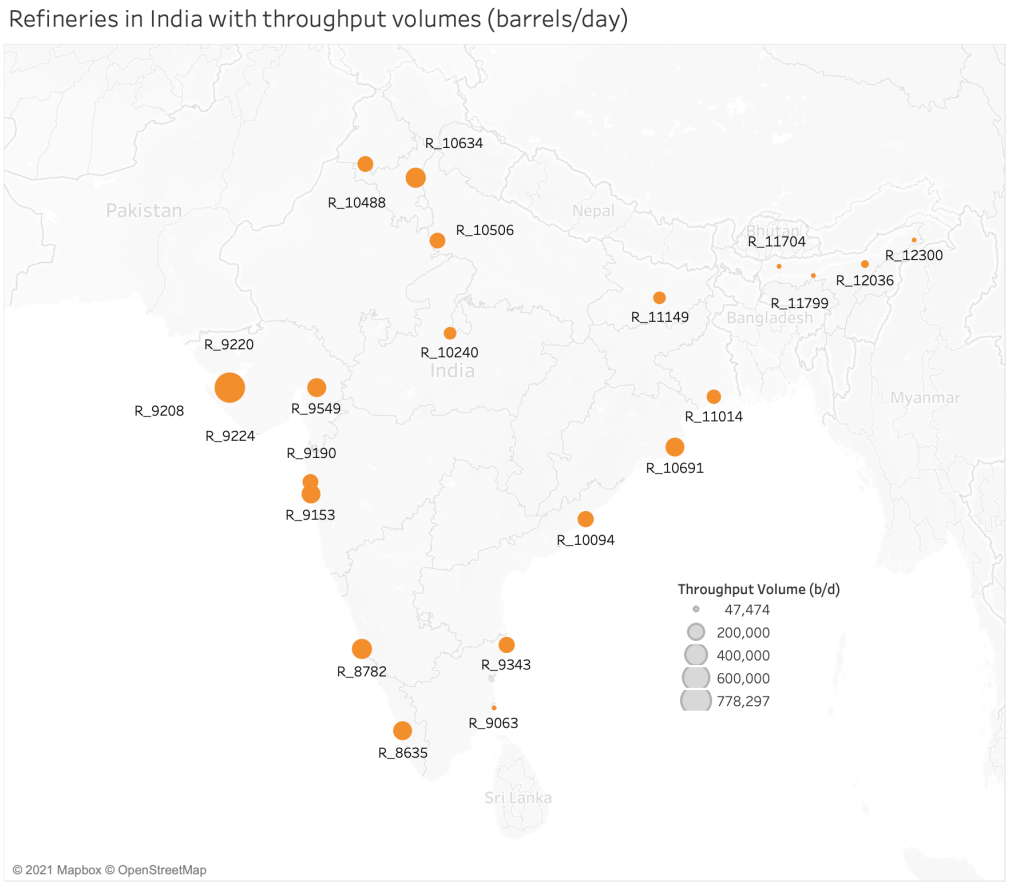

Figure 2-4: Sample of the refineries dataset - with cumulative throughput volumes

2.3 Supply chain infrastructure

The rest of the supply chain data i.e. entities excluding oil fields and refineries are sourced from GlobalData [21]. This includes locations of shipping terminals, locations of pipelines and asset characteristics of pipelines such as diameter, length etc. High-fidelity estimates of these characteristics at the level of individual pipeline segments, are of key importance in the bottom-up estimation of emissions associated with pipeline transportation. Geolocations of pipeline systems in Europe are shown in figure 2-5.

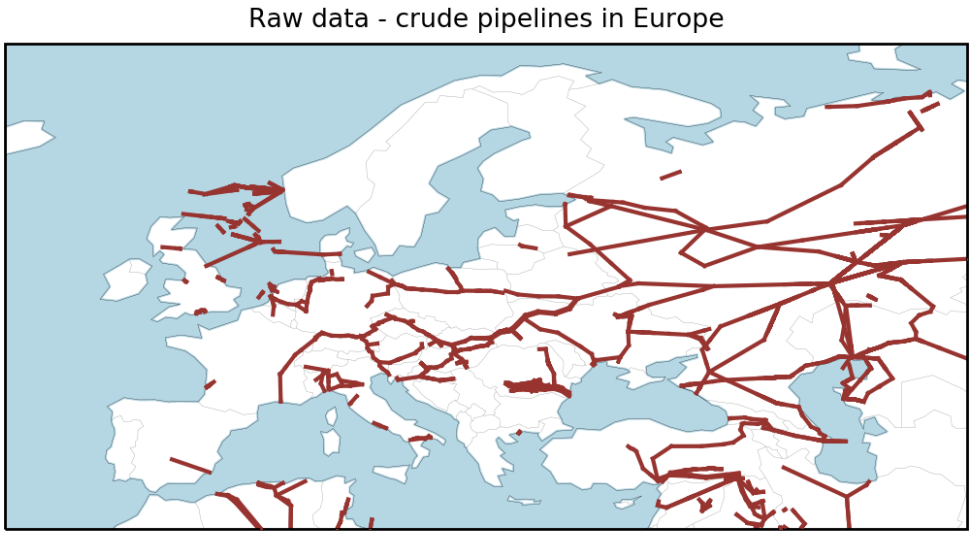

Figure 2-5: Sample of the pipelines dataset - raw data of pipeline segments in Europe

In order to merge the infrastructure data with the trades data, the dataset provided by Kpler [20] is used. This includes linkages from export terminals to import terminals depending on the existence of trade relationships between them.

The export-import terminal pairing lacks information about the trade fulfilment i.e. the shipping mechanism by which the trade takes place. This gap is filled using the IHS Markit data [22] that provides crude tanker locations along with relevant features such as tanker engine configurations, speed, etc. necessary to estimate shipping emissions.

2.4 Geographical attributes

To construct physics-based emission models on top of the supply chain model, ancillary features such as ambient temperature and elevation are needed. For instance, elevation change across a pipeline is a relevant input guiding transport emissions.

For land and ocean temperature, the study uses the NASA MODIS Land/Ocean Surface Temperature and Emissivity data. The data is retrieved at 1 km pixels by the generalized split-window algorithm and at 6 km grids by the day/night algorithm [23].

In addition, elevation data is obtained at a resolution of 1 arc degree from two sources - the 2000 Shuttle Radar Topography Mission and NASA SRTM, which gives the elevation or altitude at any given geolocation [24].

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 3

Methods

The methods based on data sources described in the previous chapter, are designed to generate CO_2 estimates that are truly actionable from a policy perspective. To that end, the research adopts a methodology that preserves the pathway-level resolution in the supply chain (i.e. granular routes from oil fields to refineries). In addition, the methods are modular in order to have the flexibility of ingesting newer data streams whereby depending on data augmentation, parts of the emission estimates can be enhanced selectively.

First, motivated by the complexity and global heterogeneity of the supply chain, a network-based approach is used to model the location, specifications and trade relationships of assets and infrastructure. Second, a "blend estimation" algorithm is designed to predict how crude blends are formed from oil fields. This uses the properties of the global network in conjunction with a multi-objective optimization approach based on automatic differentiation and unsupervised learning. The output from this algorithm generates estimates of blend-level upstream carbon intensities and a high-resolution mapping of crude barrels from sources (i.e. oil fields) to destinations (i.e. refineries). Third, the source-to-destination mapping of barrels serves as the input for the barrel tracking algorithm based on shortest-paths in the global supply chain network. And fourth, results from the tracking algorithm are fed into mode-specific bottom-up physics-based models to estimate emissions associated with transportation of crude via pipelines and shipping tankers.

The overview of these methods is summarized in table 3.1:

Table 3.1: Overview of methods

Index	What	Why	How
1	Modeling the supply chain	Efficient choice of data-structure for subsequent high-res carbon intensity estimation	Network-based approach - assets as nodes, transportation modes as edges
2	Estimating how oil fields combine to form marketed crude blends (Blend estimation algorithm)	Link sources (oil fields) and destinations (refineries) using shared info about crude blends	Multi-objective gradient-based optimization with an initialization algorithm
3	Tracking barrels from source oil fields to destination refineries	Generate visibility into how barrels move along supply chain pathways	Shortest-paths from oil fields to respective refineries based on crude blend data
4	Estimating carbon intensities along pathways	Quantify carbon emissions granularly and highlight decarbonization potential	Mode-specific emission estimation models (pipeline and shipping) + results from tracking

3.1 Network construction

The global crude supply chain is modeled as a network comprising of nodes and edges. The modeling choice of a network is ideal to represent the myriad pathways that convey crude oil from source oil fields to destination refineries. This not only enables the estimation of transportation carbon intensity, but also future climate-oriented supply chain extensions such as CO_2-optimized rerouting. Thus, the network seeks to further research contributions both in this study as well as for related work such as techno-economic analysis and supply chain optimization.

3.1.1 Creation of nodes

Supply chain assets from all aforementioned data sources are consolidated and are assumed to be point geospatial objects. After categorization into the five classes of “fields”, “terminals”, “shipping ports”, "pipeline stations” and “refineries”, these objects are encoded as nodes in the supply chain graph. For subsequent use, the nodes carry other useful attributes such as precise geocoordinates, asset name and country information.

“Terminals” and “shipping ports” are treated separately due to the different sources of data; suitable edge construction is performed to minimize asset duplication. In order to focus on the crude supply chain, terminals which do not handle crude (but handle other petroleum products) are excluded as mentioned in the commodity type attribute of the dataset.

The node types, counts and corresponding data sources are summarized in table 3.2.

Table 3.2: Summary of nodes in the network

Node Type	Asset Category	Node Count
p	Pipeline Station (GlobalData)	10681
t	Terminal (GlobalData)	607
f	Oil Field (WM)	10460
r	Refinery (WM)	746
kt	Shipping Port (Kpler)	310

3.1.2 Creation of edges

The nodes in the network are connected by three types of edges - mode edges, concurrent edges and heuristic edges.

Mode edges

Mode edges are constructed using the datasets, specifically using pipeline geometries and shipping routes. These edges are assumed to be known precisely from data and are created between the corresponding nodes. In the case of a typical pipeline system which is defined in the dataset as a list of coordinates that correspond to nodes in the network, the edges are constructed between consecutive nodes to represent the system.

For example:
Edges in a pipeline defined by the coordinates $[p_1(x_1, y_1), p_2(x_2, y_2), ..., p_n(x_n, y_n)]$ correspond to connections between points $k, k+1$ where $1 <= k <= n-1$

The limitation of this chosen graphical representation is that supply chain assets (oil fields, refineries, terminals) are assumed to be point entities, whereas in reality they are spatial entities. This limitation constrains the relationship between fields and pipelines in that, pipelines are typically observed to be constructed through the perimeter of oil fields which is hard to capture in the chosen representation.

To manage this limitation, sub-segments of pipelines are created at junctions where field nodes intersect pipeline segments and thus allow fields to have proximate and efficient pipeline access. As a consequence, the network ends up having artificial pipeline junctions and a greater number of constituent pipeline segments. Relevant data features such as pipeline diameter and length are encoded as edge attributes.

The other important mode edge captures shipping routes by linking shipping ports based on the dataset of export-import trades. Data features such as route mileage and tanker type are encoded as edge attributes.

Mode edges represent the majority of the transportation lanes in the supply chain. The others enable a completion of the network by resolving issues of asset duplication, geolocation errors etc.

Concurrent edges

Concurrent edges are created between nodes which approximately have the same geolocation. A tolerance of 1 km is to construct edges which fulfil this criterion. This category of edges is salient in cases where asset classes like fields and refineries are key junctions in pipeline systems.

Heuristic edges

Heuristic edges are added to make the network representation more realistic. Partially aiming to address the aforementioned limitation of point entities, these edges are constructed between the field, refinery nodes and the rest. These heuristic edges use a thresholding criteria of 10

km and 50 km in a hierarchical manner i.e. if a particular field node lacks connectivity, it is connected to nodes within 10 km and if it still lacks connectivity the process is repeated with a 50 km threshold. These thresholds are chosen based on how well connected the network appears in terms of allowing for supply chain pathways from fields to destination refineries. More than 90 percent of the nodes are observed to secure connections after implementation of the three types of edges. The remaining are either:

(A) Zero crude volume entities (defunct refineries, closed oil fields)

(B) Terminals without any assets near them and without any shipping trades associated with them (likely down to misclassifications in the dataset)

(C) Coastal refineries in importing countries which have the capabilities of shipping ports without having an explicit shipping port node near them (these cases are seen in South Asia and Iberia)

Of the above, categories A and B do not contribute to the goal of tracking supply chain pathways. In order to capture reasonable connectivity for cases in category C, we create artificial terminal locations coinciding with the refinery locations such that the corresponding terminal-refinery pairs are connected. Category C can be seen typically in importing countries that do not have extensive pipeline coverage (examples include Spain, Portugal, Japan).

The network design and the different edge types are summarized below:

- *Mode edges* - connections capturing the different modes of transportation in the supply chain
- *Concurrent edges* - connections managing data redundancy and mutual inconsistencies across datasets
- *Heuristic edges* - rule-based connections to make the network more realistic

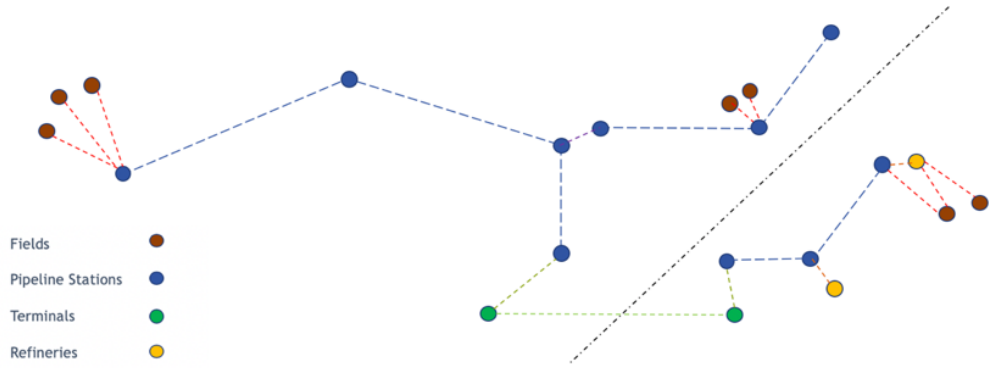

Figure 3-1: Schematic diagram of the network representing the supply chain

3.1.3 Network attributes

The network representation is augmented by encoding salient data features as node and edge attributes.

Key node attributes include latitude, longitude, node-type (field, pipeline station, shipping terminal, refinery), asset name, asset country. In addition, in order to facilitate subsequent emission estimation, physical attributes such as annual average ambient temperature and elevation are also included. These attributes are inferred from the data sources described in Chapter 2.

Universal edge attributes include edge types and distances. Type-specific attributes include diameter, length, elevation change for pipeline edges and shipping route distance, vessel type for shipping edges.

The network attributes are summarized in table 3.3:

Table 3.3: Attributes of the supply chain network

Node attributes	**Edge attributes**
Location (latitude, longitude)	Edge type (pipeline, shipping, other)
Asset name, asset country, asset type	Shipping distance, vessel type
Average ambient temperature, elevation	Pipeline distance, diameter, elevation change

The completed network data structure is used in the subsequent modules of the methodology.

3.2 Blend estimation algorithm

The oil supply chain entails transformation of the commodity flowing through the network – crudes from oil fields are transformed into marketable crude blends which eventually get refined. Consequently, a high resolution life cycle analysis needs to have visibility about the variety of global crudes as they are transformed and transported from the respective origins to destinations. This process is summarized in figure 3-2.

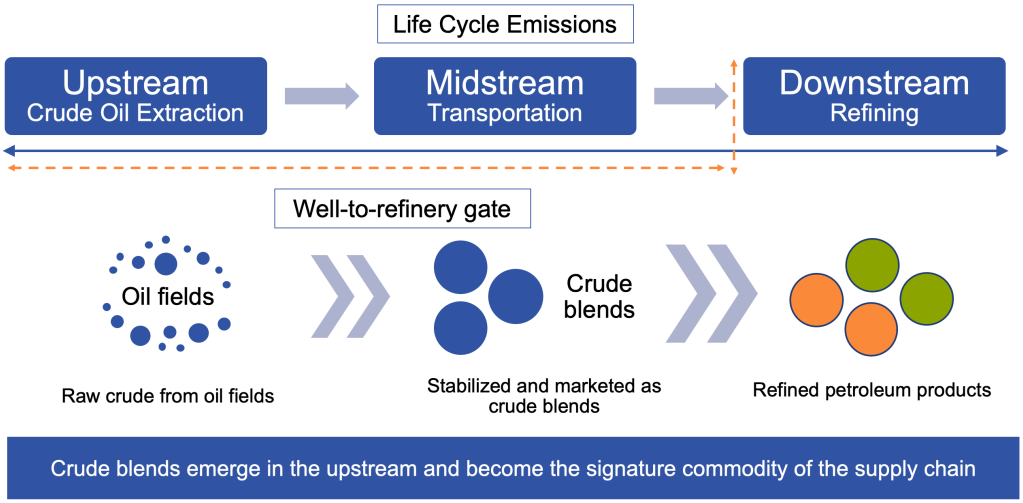

Figure 3-2: Contextualizing crude blends within the supply chain

Thus, mapping crude volume from fields onto crude blends and then mapping crude blends onto refineries has two salient implications:

First, it enables the carbon footprint estimate of crude trade from origin to destination; this lays the foundation for actionable decarbonization policy because identifying crude trade with carbon footprints is fundamental for regulators. Second, it leads to a distinct carbon intensity accounting for the upstream (i.e. crude extraction) and midstream (i.e. transportation) segments of the supply chain rather than relying on default baseline values.

To that end, the blend estimation algorithm predicts how marketable crude assays form from source oil fields; i.e. for a given oil producing country, the algorithm estimates the relationship between oil fields and crude blends. This facilitates both, quantification of the upstream carbon footprint of crude assays and prediction of pathways from sources to destinations.

3.2.1 Objective and formulation

The blend estimation algorithm is framed as an optimization problem, in particular, a set of country-specific multi-objective optimization problems i.e. independent problems for every oil producing country.

For a given oil producing country, the goal is to estimate the relationship between oil fields and crude blends. This relationship is represented in the configuration matrix which encodes the fraction of volumes from all oil fields contributing to make all crude blends in the given country.

The country-specific configuration matrix is a matrix of $F \times B$ dimensions where F is the total number of oil fields and B is the total number of crude blends. The value corresponding to the ith row and jth column is the fraction of crude volume from field i that contributes to blend j.

Thus, under this framing, the process of blend estimation is the estimation of the configuration matrix for each oil producing country. As described in the chapter overview, this estimation problem is posed as an optimization problem that minimizes a cost function described below.

In terms of the notation, the goal is to estimate Θ - the Configuration Matrix where

- rows of Θ correspond to fields
- columns of Θ correspond to blends
- shape of $\Theta = F \times B$ were F = Number of fields, B = Number of blends in the country

$$\Theta = \begin{bmatrix} p_{11} & p_{12} & \cdots \\ \vdots & \ddots & \\ p_{F1} & & p_{FB} \end{bmatrix}$$

where p_{ij} is the fraction of crude volume from field i contributing to blend j

With the decision matrix defined, the cost function of the optimization problem seeks to account for the nuances of the supply chain. Specifically, given a configuration matrix, the goal is to quantify the cost associated with it such that the cost reflects the feasibility of the encoded relationship between fields and blends in the supply chain. Given the complexity that guides blend formation, the sub-components of the cost function are distilled down to four factors such that they capture real-world features shaping blend formation.

1. Distance - C_d
2. Connectivity - C_c
3. Volume Error - C_v
4. API Error - C_a

The notation that guides the formulation of these components is shown in table 3.4.

Table 3.4: Notation guiding the computation of the cost function

Variable	**Description**	**Dimensions**
V_F	Volume vector of oil fields	$F \times 1$
V_B	Volume vector of crude blends	$B \times 1$
A_F	API vector of oil fields	$F \times 1$
A_B	API vector of crude blends	$B \times 1$
D_F	Distance matrix of oil fields (where D_{F_ij} is the distance between the ith and jth oil field in the given country)	$F \times F$
P_F	Boolean connectivity matrix of oil fields (where P_{F_ij} is the boolean for the existence of a path from the the ith and jth oil field in the given country)	$F \times F$
Θ	Configuration Matrix	$F \times B$

These variables are then used to compute C_d, C_c, C_v, C_a as shown in table 3.5 [1] [2].

[1] o is the Hadamard product
[2] $||_1^1$ is the first norm

Table 3.5: Cost components in the optimization problem

Sub-component	What	How
C_d	To what extent do proximate oil fields make up the same blends	$\|\|(\Theta\Theta^T)o(D_F)\|\|_1^1$
C_c	To what extent are oil fields contributing to the same blends connected in the supply chain network	$\|\|(\Theta\Theta^T)o(P_F)\|\|_1^1$
C_v	How closely does the distribution of crude from oil fields manage to approximate the blend volumes	$\|\|((V_F^T\Theta)^T - V_B)\|\|_1^1$
C_a	How closely does the distribution of crude from oil fields manage to approximate the blend APIs	$\|\|((A_F^T\Theta)^T - A_B)\|\|_1^1$

C_d and C_c, as described table 3.5 respectively quantify the extent to which proximate and well-connected fields make up the same blends. $(\Theta\Theta^T)_{ij}$ indicates the co-blending of crude from field-i and field-j. Coupled with (D_F) and (P_F) through element-wise multiplication, it thus measures the cost associated with distance and connectivity respectively.

Regarding C_v and C_a - the formulation quantifies the difference between estimated and actual crude blend volume/API ($[V_F^T\Theta]^T$ and $[A_F^T\Theta]^T$ are the estimated crude blend volumes and APIs respectively).

To ensure all cost terms are comparable, every sub-component is scaled using an estimate of its magnitude. This estimate is the sample mean of 100 sub-components computed by randomly sampling the configuration matrix subject to the constraint of every row summing to one (i.e. fractions of crude from an oil field allocated across all crude blends add to one).

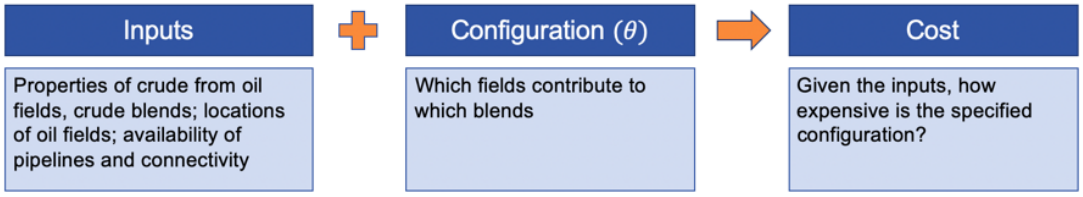

Figure 3-3: Formulating the cost function from the inputs and configuration matrix

The net cost associated with a configuration matrix is a linear combination of the sub-components multiplied by non-negative weights. This weighted summation makes the objective function "multi objective" wherein the weights drive the relative importance of the respective sub-components.

$$Cost(C) = \sum_{i=(d,c,v,a)} w_i C_i \quad \text{where}$$

$$w_i > 0 \quad \text{and} \quad \sum_{i=(d,c,v,a)} w_i = 1$$

The objective is to find the optimal configuration matrix that minimizes the cost function, conditioned on the four weights.

$$\Theta^* = argmin_\Theta \left[Cost(C) = f(inputs, weights, \Theta) \right]$$

3.2.2 Overview of the optimization approach

The optimization problem is solved using a gradient-based technique coupled with an initialization algorithm.

The gradient-based technique uses autodifferentiation [25], a core component of training deep learning models. This approach uses gradient descent coupled with the concept of momentum [26] which is prominent in speeding up the training of deep neural networks.

The initialization algorithm acts as a bridge between real-world supply chain attributes and

the configuration matrix. It ingests information that is not captured by the cost function such as similarity between crude blend names and basin and/or oil field names. Furthermore, it includes unsupervised learning and a genetic algorithm to avoid issues of local minima traps encountered in gradient descent.

3.2.3 Weights associated with the multi-objective cost function

The four weights associated with the cost function dictate the relative importance of the corresponding four sub-costs.

The weights are set to 0.25 i.e. all are equal to assign sufficient importance to the sub-costs associated with volume, API, distance and connectivity. The weights are varied to quantify uncertainties shown in Chapter 4.

The choice of weights represents a direction for further improvement. With better and more data about the specifics of supply chains in different countries, there is significant room to better set the weights to reflect the physical realities of crude blending. This aspect is addressed in more detail in Chapter 6.

3.2.4 Initialization

The initialization algorithm is comprised of multiple modules executed in series. For a given field-to-blend variable Θ_{ij}, the initialized value $\Theta_{ij_{init}}$ is determined as follows:

$$\Theta_{ij_{init}} = \prod_{k \in Modules} \Theta_{ij_k}$$

This enables a successive weighting of the configuration matrix that effects a collectively informed initialization for gradient descent.

Similarity scores based on entity names

Nomenclature of marketed crude blends is known to be region specific and occasionally, the same is observed in the case of oil fields. Furthermore, in cases where the blend originates from a specific cluster of fields which reside in the same basin, there are similarities in the way

the blends and the fields are named. Taking advantage of this information, this module uses name similarity scores to bias the initialization of the configuration matrix.

Similarity scores are calculated using the difflib library in Python which measures the degree to which two sequences are similar. The pseudocode is shown below:

---

function name similarity score(field_name_i, blend_name_j, threshold, match_factor)

```
// Example blend names: West Texas Intermediate, Arab Light, Kirkuk Blend

list_of_processed_blend_substrings = process_and_split(blend_name_j)
 // Blend name typically includes strings such as "blend","crude" which
 don't carry useful information.

for si in list_of_processed_blend_substrings
 degree_of_similarity = similarity(field_name_i, si)
 list_of_similarities.append(degree_of_similarity)

max_similarity = max(list_of_similarities)

if max_similarity > threshold:
 Θ_ij = match_factor  x  Θ_ij
```

end

Sample results from the name similarity module are shown in table 3.6.

Table 3.6: Blend, field pairs with high name similarity

Blend Name	Field Name	Country	Similarity Score
Bozhong	Bozhong 19-4 (11/19D)	China	1
Kirkuk Blend	Kirkuk (Avanah Dome)	Iraq	1

Genetic algorithm

The configuration matrix has large dimensions which are variable (100s of oil fields and 10s of crude blends). This translates to a high number of parameters in the configuration matrix which makes the gradient descent susceptible to local minima traps [27].

This limitation hinders the gradient-descent from converging in a stable, robust manner. To mitigate this, a genetic algorithm, which is a non-gradient optimization method, is used in the initialization process [28].

The parameters of the genetic algorithm are the parent population size, offspring population size, fitness function, the chromosomes and the number of iterations. Starting with the parent population, the offspring population is generated by crossing over chromosomes i.e. swapping the two halves of equisized field-to-blend arrays in the configuration matrix. Candidates in this population are mutated by randomly varying one row in the matrix (corresponding to one field) and are selected for fitness by passing through the cost function.

Table 3.7: Parameters guiding the genetic algorithm

Parameters in the genetic algorithm	**Value**
Parent Population size	20
Fitness function	1 – Cost Function
Offspring Population size	10
Chromosomes	Two equisized field-to-blend arrays
Iterations	5000

Figure 3-4 shows the decrease in the cost function (or the generation of fitter samples of the configuration matrix) over 5000 iterations of the genetic algorithm. The decrease has a non-trivial component of noise which provides further evidence for limiting the use of this module solely in the initialization phase.

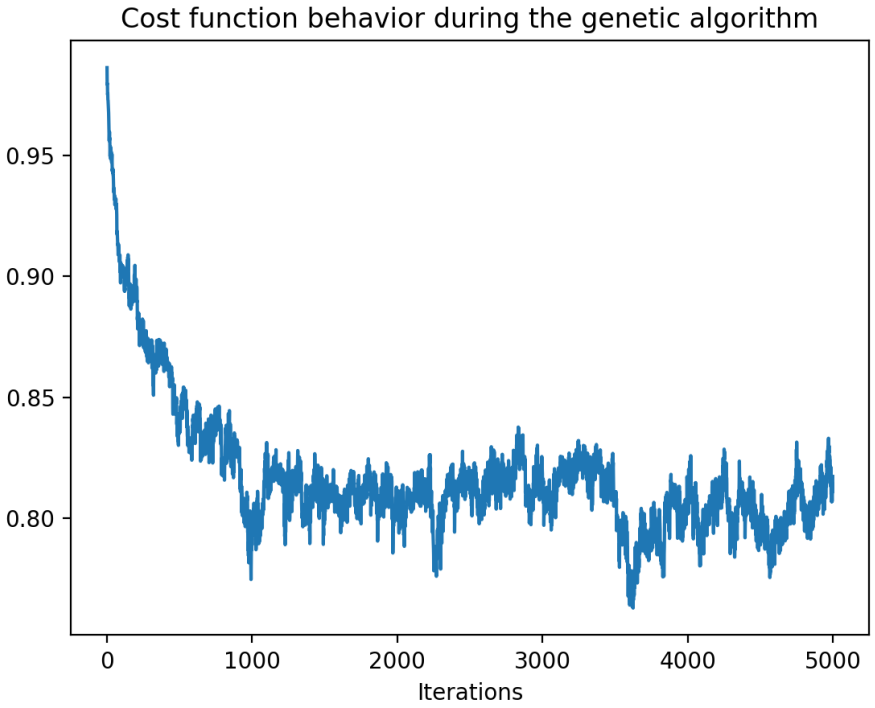

Figure 3-4: Sample cost function decrease during the genetic algorithm in the initialization module

K-means clustering

Oil field clusters exhibit the property of being co-located in basins and other common geographical areas. This leads to certain crude blends being region specific and consequently being formed from the fields co-located in the respective regions. For a few cases, this common property reflects in the name similarity, but for the vast majority, it does not. This motivates a proximity-based module that groups oil fields based on their geolocations and forms blends from the emerging field clusters.

K-means clustering is used given the nodular nature of the supply chain representation. Oil field nodes represent data points to be clustered and crude blends represent the number of clusters to be formed. In order to emphasize the importance of fields who are major contributors to output, volume-weighted distancing is used to prioritize those with high production

volumes [29].

Table 3.8: Sample setups for K-means clustering

Country	Number of field nodes	Number of clusters = Number of blends
United States	1048	37
China	649	15
Saudi Arabia	35	6
Oman	176	1
Brazil	323	10

The cluster assignments are fed into the initialization algorithm as follows:

$$\Theta_{ij} = \Theta_{ij}\ x\ \epsilon \quad \text{if } i \notin cluster\, j$$
$$\Theta_{ij} = \Theta_{ij}\ x\ M \quad \text{if } i \in cluster\, j$$
$$\text{where } \epsilon = 0.01 \text{ and } M = 100$$

3.2.5 Gradient descent using autodifferentiation

Gradient descent is performed starting with the initialized configuration matrix. Given the computational complexity of the cost function and the intrinsic non-linearity, gradient descent is implemented using autodifferentiation [25, 27].

This module uses the deep learning framework PyTorch [30] for its computational graphs that facilitate the functionality of autodifferentiation. The graph allows for gradients to be computed at every step in the computation, which coupled with the chain rule of differentiation enables the computation of complicated gradients in a feasible manner.

Gradient descent is implemented with the concept of momentum which does effective averaging over the steps of descent. Momentum has been shown to make gradient descent more effective in training neural networks with large parameter spaces as shown by Qian et al [26].

Table 3.9 summarizes the optimized hyperparameters that ensure stable convergence close to the local minima.

Table 3.9: Optimized hyperparameters in the gradient descent module

Hyperparameter Name	**Value**
Number of iterations	30000-50000 (depending on the size of Θ)
Step size	Variable (0.035 to 0.07) - indexed to iterations - step size gets smaller with more iterations
Momentum factor (γ)	0.9

The cost function decrease shown in figure 3-5 illustrates the efficiency of using gradient descent with momentum. Furthermore, using an adaptive step size speeds up the convergence as the number of iterations increase.

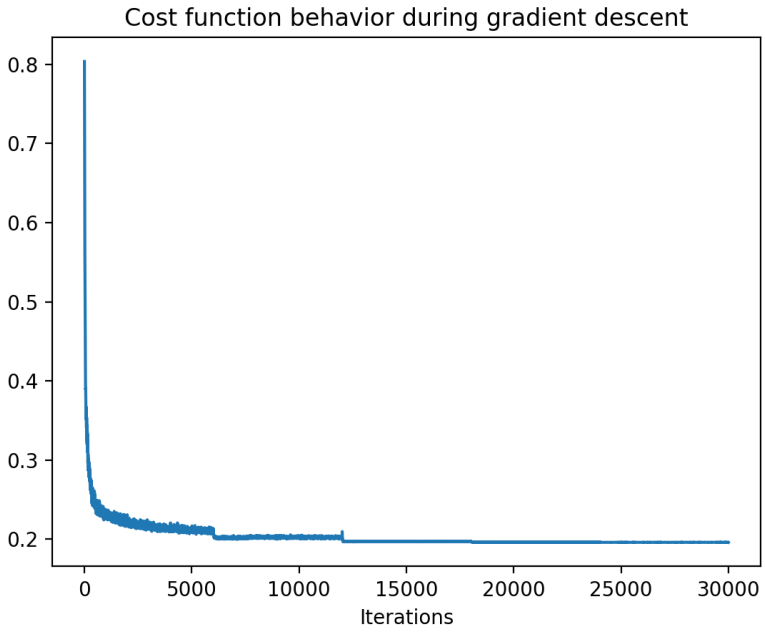

Figure 3-5: Sample cost function decrease during gradient descent

With the hyperparameters shown above, gradient descent is implemented as shown in the

pseudocode below:

---

function $gradient\_descent(\Theta_{\textbf{init}}, hyperparameters, cost\_function)$

$\Theta = \Theta_{\text{init}}$

$delta = \mathbf{0}$
`// momentum factor - same dimensions as theta`

for i in 1:niter
$C = cost\_function(\Theta)$
$delta = [\gamma \times delta] + [step \times \nabla C_{\Theta}]$
$\Theta = \Theta - delta$
`// Gradient gradient descent with the momentum term.`

```
// The backward method in PyTorch is used on the theta tensor inside every
 iteration of the cost function
```

end

3.2.6 Priority mode

For countries with a large number of oil fields and blends (eg: United States, Russian Federation, China), the aforementioned modules of the optimization are run on a priority set of fields. This priority set is generated by sorting fields in descending order of their volume and with a cutoff at 95 percent cumulative volume contribution. As a consequence, this picks the priority fields and leaves out the low volume contributors thus making the gradient descent more stable. The low contributors are allocated to blends based on the corresponding closest field from the priority set. The priority mode can be summarized as:

- Producing country has >100 oil fields
  - If "No"
    - Proceed with baseline Blend Estimation Algorithm

- If "Yes"
  - Sort oil fields in descending order of volume (higher volume first)
  - Filter those fields which cumulatively represent 95 percent of the country's production volume (known as the priority set)
  - Run the Blend Estimation Algorithm on the filtered set of fields and given set of blends
  - For every residual field, complete blend assignment identically as that for the closest field in the priority set

3.2.7 Limitations

The blend estimation algorithm is implemented after filtering out condensates and natural gas liquids (NGLs). These commodities typically have high API (low density) and interact with sister supply chains (eg: natural gas). Thus, this exclusion is a consequence of the available data that is limited to the oil supply chain.

Mass imbalances are observed in the key region of Texas affecting important blends like West Texas Intermediate, Eagle Ford. To mitigate these issues, the algorithm includes a module that selectively calibrates the initialization of the configuration matrix for the United States. This calibration is based on publicly available data about the likely origin oil fields of the major blends.

3.2.8 Sample cases

Blends associated with key regions and/or fields

Figures 3-6 and 3-7 show sample results of the blend estimation algorithm, specific to the blends Wyoming Sweet from the U.S. and Arab Light from Saudi Arabia respectively. Both figures illustrate the oil fields that contribute to the mentioned blends with size indicating the production volume of the field and color indicating the fractional contribution of the field in making the blend.

Figure 3-6 illustrates the regional specificity of the algorithm whereby crude from key fields from Wyoming is blended to form Wyoming Sweet.

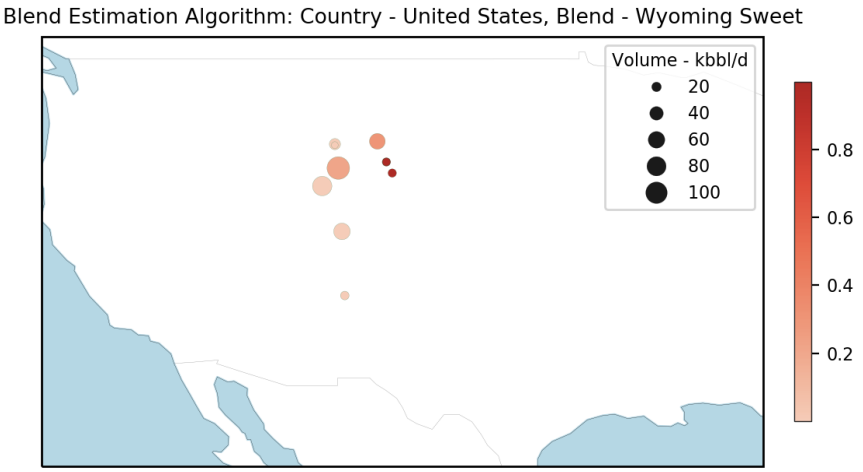

Figure 3-6: Fields in Wyoming predominantly contributing to Wyoming Sweet

Figure 3-7 captures the widely recognized mapping between the field Ghawar (and proximate sister fields) and the blend Arab Light in Saudi Arabia.

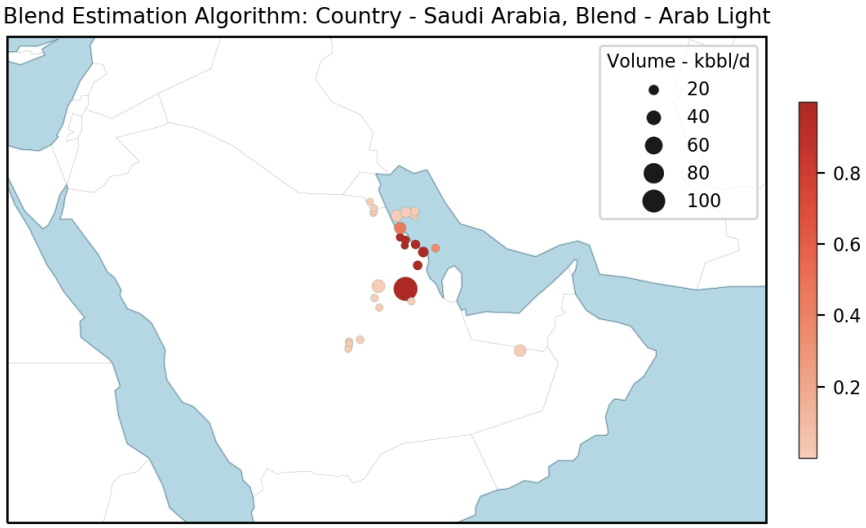

Figure 3-7: Cluster of fields centered around the biggest oil field in the world - Ghawar, contributing to the highest volume blend in the world - Arab Light

Blend formation in select countries

Figures 3-8 and 3-9 represent the aggregate blend estimations in Iran and Saudi Arabia respectively. A common finding in both these cases is the presence of regional clusters both onshore and offshore which points towards zone-specificity in blend formation.

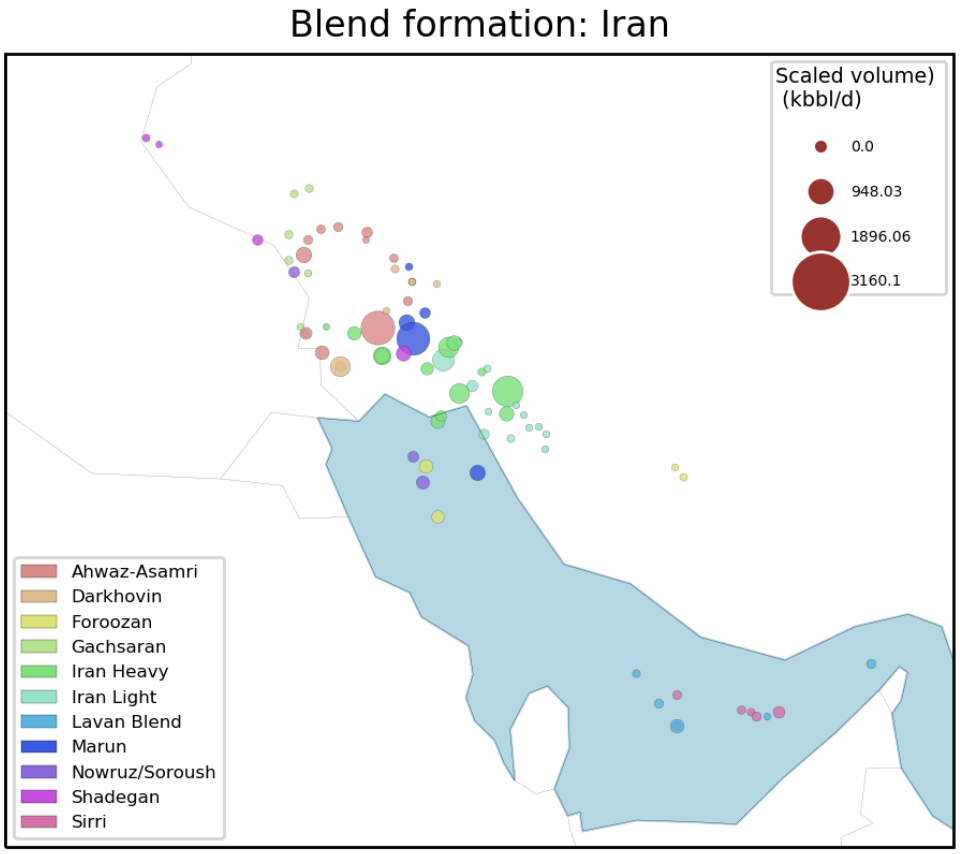

Figure 3-8: Summary of blend formation in Iran

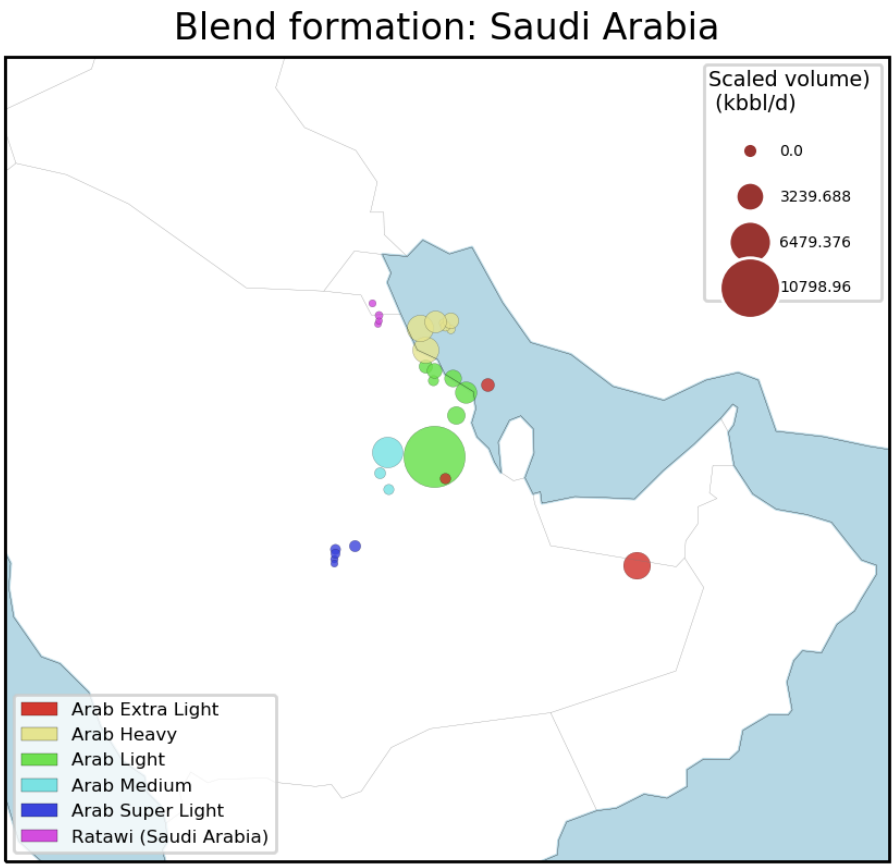

Figure 3-9: Summary of blend formation in Saudi Arabia

3.3 Tracking algorithm

The relationship between oil fields and crude blends along with known data about which blends are consumed by refineries globally leads to inference of the supply-demand mapping from oil fields to refineries.

The mapping of fields to blends and that of blends to refineries leads to the tracking of oil barrels in the supply chain network from source oil fields to destination refineries. This sets

up the foundation for a high-resolution midstream carbon intensity analysis i.e. the emissions associated with crude transportation.

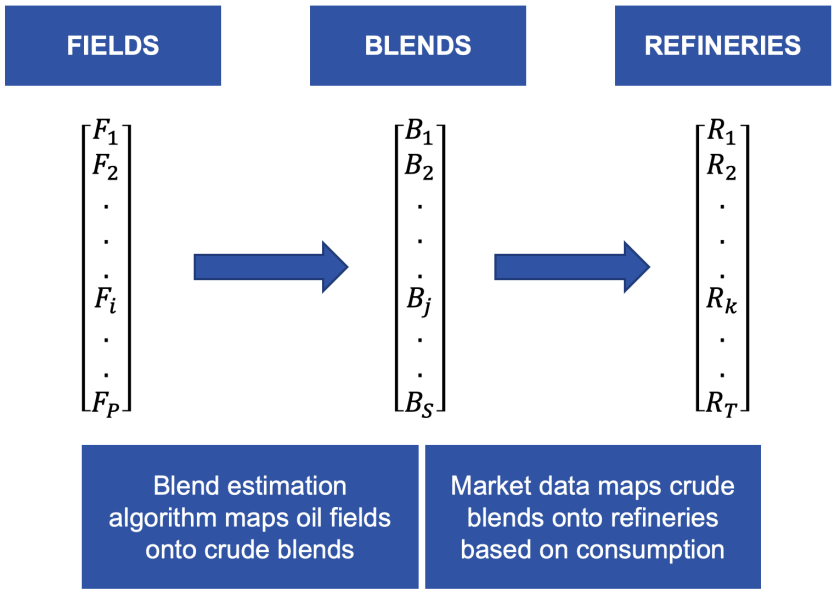

Figure 3-10: Stages in the supply chain - fields, blends and refineries

The tracking of barrels across this mapping is performed using two approaches:

1. **Blend to Refinery Approach** – creating artificial nodes representing blend centers that act as consolidation entities for constituting barrels and subsequent tracking from these blend nodes to refinery nodes
2. **Field to Refinery Approach** – tracking barrels directly from field nodes to refinery nodes without approximating blend centers

To assess which approach is more effective: First, the tracking results are augmented with physical data such as crude viscosity, elevation, temperature and electricity grid carbon intensities. Second, COPTEM, the physics-based emission estimator for pipeline transport is used to compute emissions [14]. Third, tracking results are aggregated along shipping routes to compute shipping emissions using the high-resolution shipping emissions inventory described earlier [22]. And fourth, the two approaches are compared by analyzing with reference data

such as physical elevation change, pipeline capacities, etc.

As illustrated by figure 3-11, a direct field-to-refinery tracking approach works significantly better than the blend-to-refinery approach.

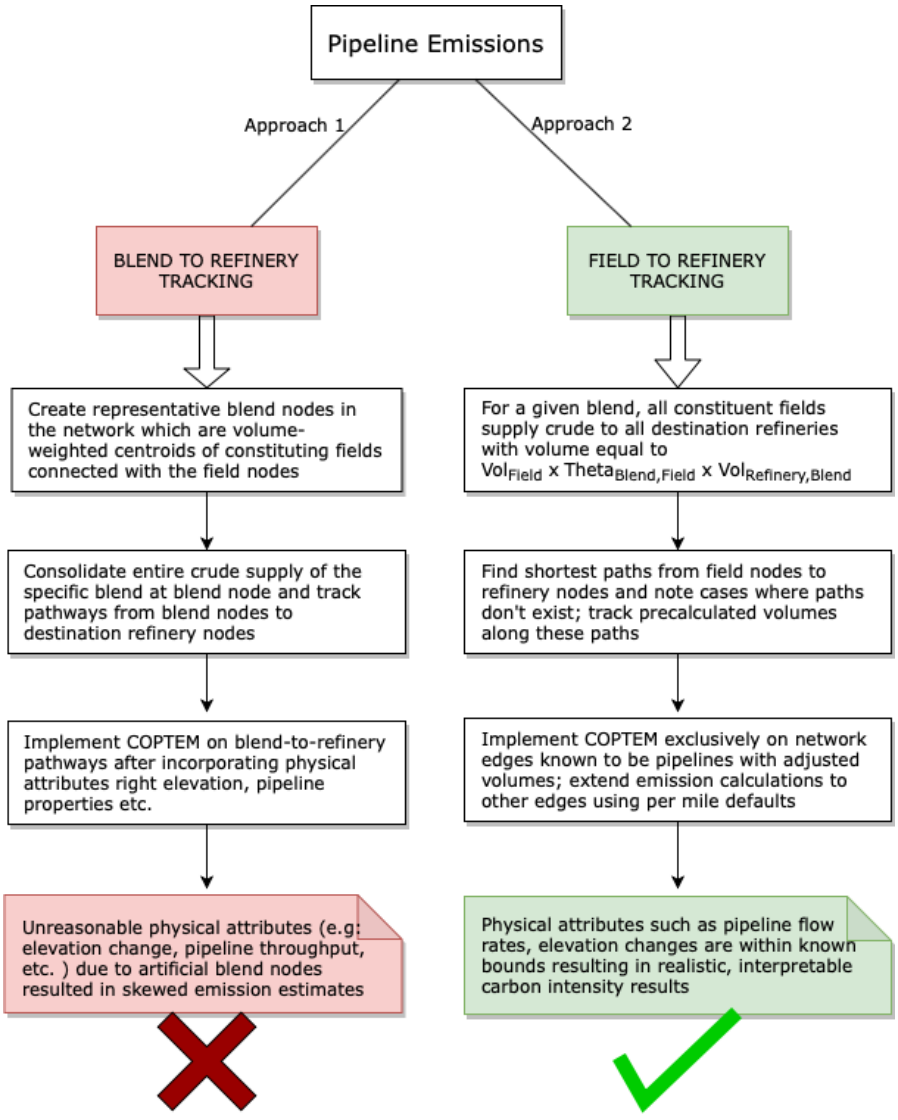

Figure 3-11: Comparing the two approaches tracking approaches

The results from tracking are consolidated and fed into the different mode-specific emission estimators mentioned earlier. Barrels transported along pipeline edges of the network are the inputs for COPTEM and those along shipping edges are inputs into the shipping emissions estimator.

3.4 Emission Estimation

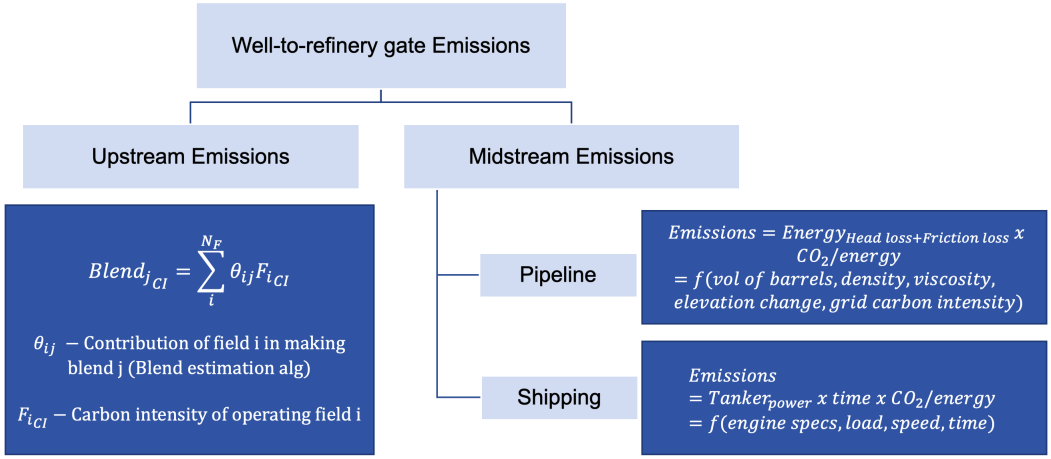

Figure 3-12: Emission Estimation - Upstream and Midstream

3.4.1 Upstream blend carbon intensities

The upstream carbon intensities for crude blends are derived from the blend configuration matrix. For a given blend, the upstream carbon intensity is the volume-weighted sum of the contributing fields' carbon intensities pre-multiplied by the configuration matrix as shown below:

$$Blend_{j_{\mathrm{CI}}} = \sum_{i=0}^{N_F} \Theta_{ij} \times V_i \times F_{i_{\mathrm{CI}}} \text{ where}$$

Θ_{ij} is the contribution of field-i in blend-j,

V_i is the volume of field-i and

$F_{i_{\text{CI}}}$ is the carbon intensity of field-i

The uncertainties in upstream carbon intensity are quantified by varying the weights in the blend estimation algorithm. Specifically, the four weights are varied from [0 to 0.7] such that the weights sum to 1. This leads to a distribution of configuration matrices which in turn leads to a distribution of carbon intensities.

3.4.2 Pipeline emissions

COPTEM, as indicated in the methodology overview, is a first-principles, fluid mechanics-based crude oil pipeline transportation emissions model [14].

The tracking results which route barrels from fields to refineries, after filtering for pipeline pathways, are the input for this model. The pipeline pathways are used to estimate emissions as follows:

- Compute energy associated with overcoming pressure losses encountered in pipeline transport
- Define emission factors, get country grid carbon intensity
- Derive emissions from energy intensity

This approach is especially useful to incorporate the impact of features such as linear velocity of crude transport, pipeline diameter, viscosity of crude on carbon intensity as shown by Choquette et al. [14].

3.4.3 Shipping emissions

A bottom-up estimation of crude shipping CO_2 emissions is conducted using an integrated dataset of terrestrial and satellite Automatic Identification System (AIS) data along with a global-level ship parameter database. As described in Chapter 2, the raw AIS data of the year 2015 acquired from IHS Markit and ship parameters from the World Register of Ship (WRS) database. The estimation of shipping emissions for crude oil tankers consists of the following procedures:

Extracting data specific to crude tankers

The data field Statcode-5 is used to identify crude oil tankers in the WRS database, and the corresponding IMO number, a unique seven-digit identifier for a vessel is used to extract AIS records of the identified crude oil tankers.

Categorizing crude tankers based on size/type

The barrel-capacity of tankers is used to derive per-barrel emissions from the absolute trip emissions. Within the extracted data, crude oil tankers are categorized according to their deadweight tonnage (DWT) provided by the WRS dataset.

Table 3.10: Crude tanker types with DWT values [31, 32]

Type	DWT
Small tanker	<10000
Handy	10000 – 60000
Panamax	60000 – 80000
Aframax	80000 – 120000
Suezmax	120000 – 200000
VLCC	200000 – 320000
ULCC	>320000

Identifying trips between shipping terminals

First, a geographical matrix including terminals' name / label and corresponding latitude and longitude is generated. Second, the AIS records are sorted by time in ascending order and are attached to the terminal labels according to their geolocation. And finally, trip labels are generated for each record according to their adjacent terminal labels.

Estimating emissions

Emission modeling described in [33, 34] based on power calculations is performed on the processed trip data. Specifically, the emissions are estimated as a function of the engine power demand, activity time, and emission factor. The engine power demand for propulsion engines

is calculated using the propeller law which estimates the power associated with propulsion, while the power demand of auxiliary engines and auxiliary boilers are determined according to their corresponding ship class, ship capacity, and activity mode. Table 3.11 shows the emission factors used in the modeling.

Table 3.11: Emission factors used in the estimation of shipping emissions

Engine Type	**CO_2 Emission Factor (g/kW-hr)**	
	Heavy Fuel Oil (HFO)	**Marine Gasoil (MGO) / Marine Diesel Oil (MDO)**
Slow Speed Diesel (<130 rpm)	620	589
Medium Speed Diesel	683	649
Gas Turbine	970	922
Auxiliary Engines	683	649
Auxiliary Boilers	970	922

Chapter 4

Results

The presentation of the results starts with upstream carbon intensity followed by midstream carbon intensity. In the former subsection, CO_2 emissions are aggregated at the level of crude blends. In the latter subsection, different levels of aggregation ranging from granular distributions to country aggregates are shown.

Lastly, CO_2 emissions are aggregated from the perspective of consumer countries to examine carbon intensity reduction potential through policy levers in the refining industry. This point of view is the aggregated view of the supply chain since it shows the net sum of impacts from sources (oil fields) to destinations (refineries) by combining upstream and midstream emissions.

4.1 Upstream carbon intensity: emissions associated with crude extraction aggregated at the level of crude blends

Upstream carbon intensity results are organized according to key producer regions - Middle East, Russia, Latin America and North America.

Figure 4-1 captures the upstream carbon intensity heterogeneity in the Middle East at both the blend level and the country level. With a range from 9.7 to 124.5 kg-carbon/barrel, the region shows significant variability, largely down to operational practices at the field level.

Furthermore, the uncertainties in Iranian blends are higher than some of the other countries due to greater number of blends in the country and less degree of differentiation between crude properties. On the other hand, the presence of a predominant blend in Saudi Arabia and Iraq (Arab Light and Basrah Light respectively), result in low uncertainties.

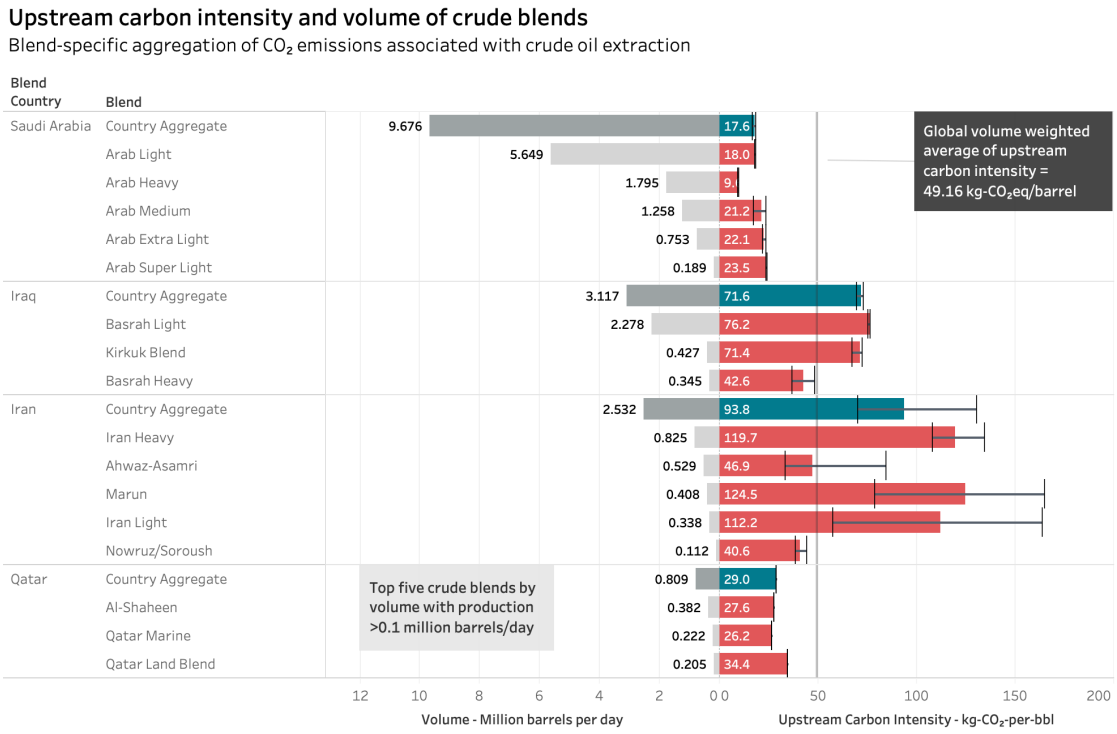

Figure 4-1: Blend upstream carbon intensities - Middle East

Similar degree of variability is seen across North and Latin America as illustrated in figures 4-2 and 4-4. Within Latin America, blends from Mexico, Brazil and Argentina are found to be in the neighbourhood of the global volume weighted average, whereas Venezuelan blends have significantly larger carbon intensities. On the other hand, as shown in figure 4-3, blends in Russia show much less variation in carbon intensities and are close to the global volume weighted average. This is primarily explained by the expansive field clusters and common large-scale infrastructure such as the ESPO pipeline network.

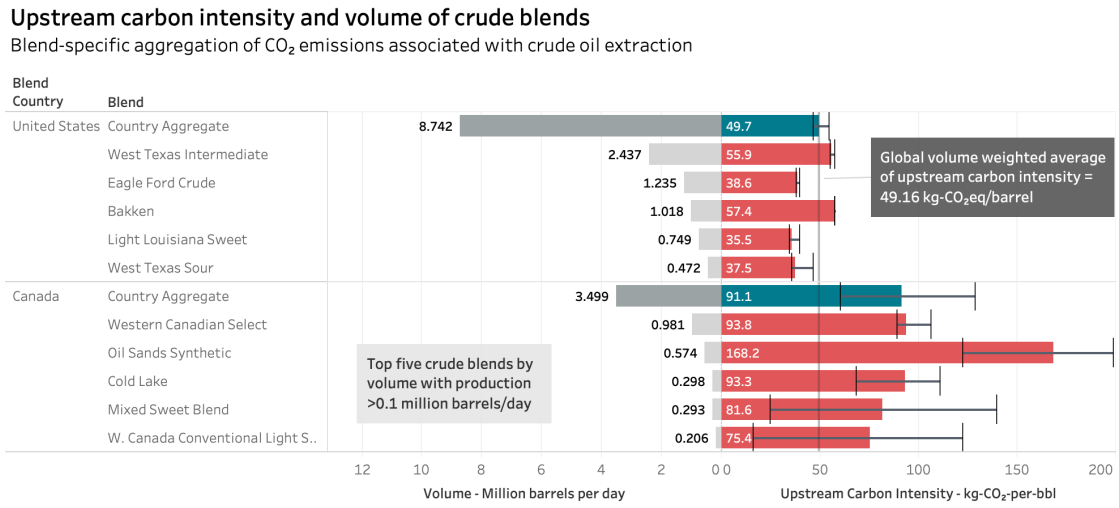

Figure 4-2: Blend upstream carbon intensities - North America

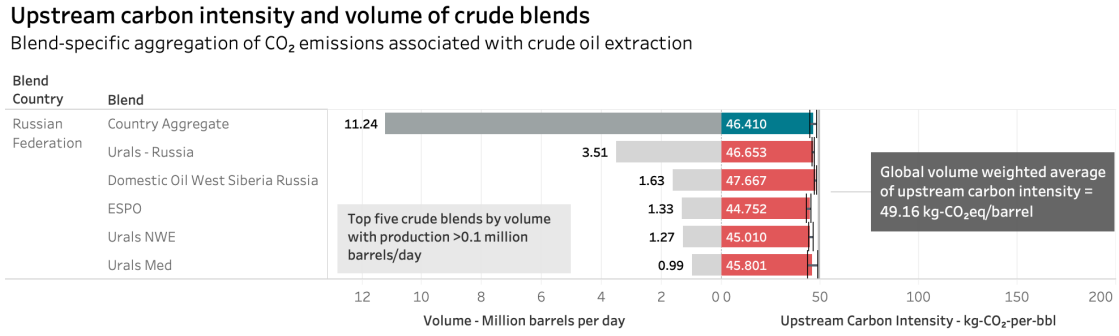

Figure 4-3: Blend upstream carbon intensities - Russian Federation

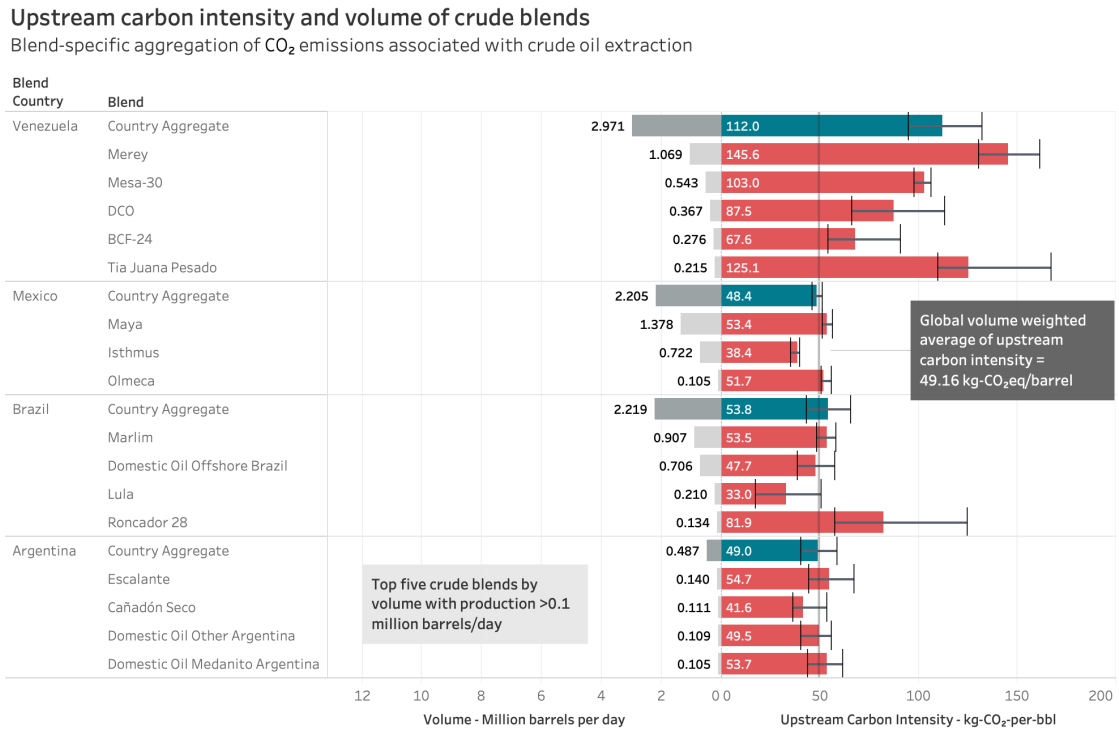

Figure 4-4: Blend upstream carbon intensities - Latin America

In addition to the country aggregate heterogeneity which has been presented by Masnadi et al.[10], figures 4-1, 4-2, 4-2 and 4-4 capture the heterogeneity at the level of blends. These inter-blend differences are the foundation for policy action such as defining incentives for market-based decarbonization. Specifically, what these results show is that policy makers now have the ability to get granular about source crudes and make decisions on the relative carbon intensities.

4.2 Midstream carbon intensity: emissions associated with crude transportation

This subsection begins with the highest level of aggregation i.e. net transportation carbon intensities between producer and consumer countries. This is followed by a more granular blend-level aggregation i.e. overall upstream and midstream emissions along with inter-pathway distributions.

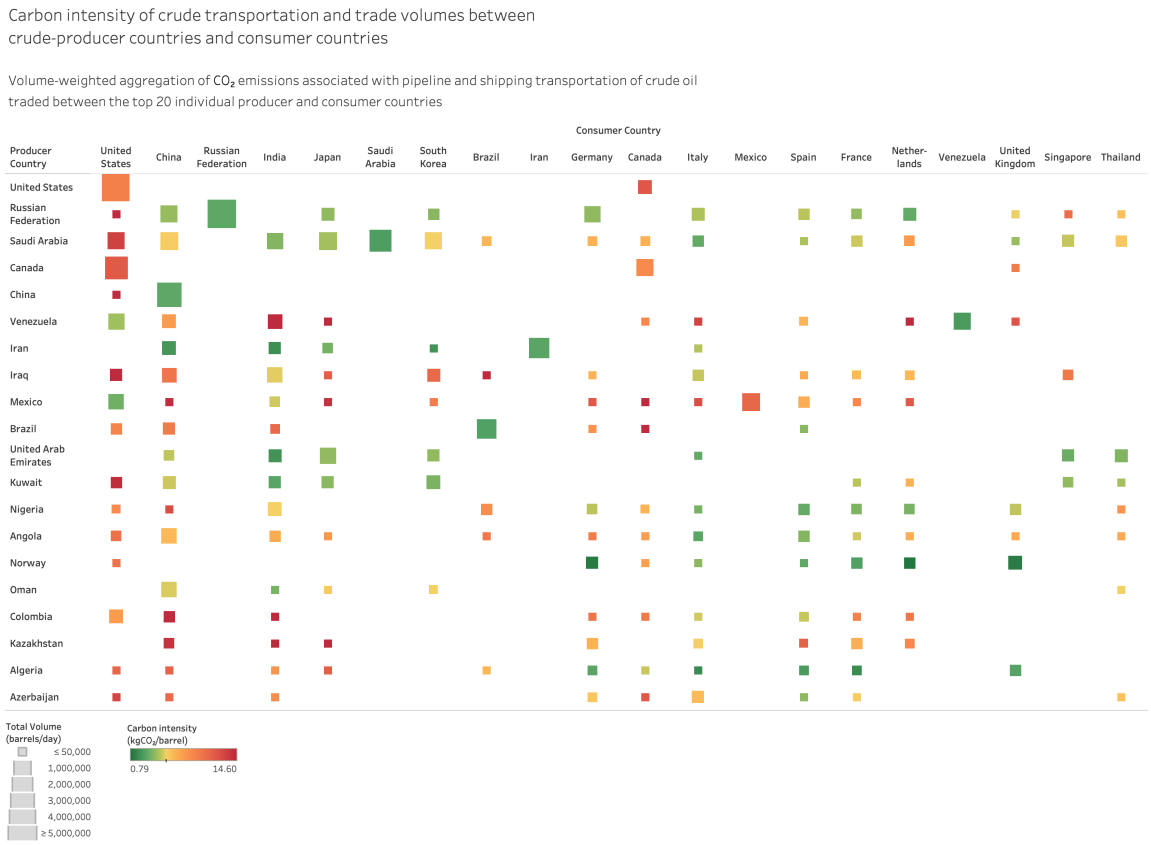

Figure 4-5: Transportation carbon intensities aggregated along supply chain pathways from producer to consumer countries

Insights from the above aggregation matrix include the shipping inefficiencies from Latin America to Asia relative to the Middle East which result in correspondingly high transportation

carbon intensities, the dominance of pipeline transport over shipping for Russian crude and the high carbon intensity of North American pipeline systems as a consequence of the complexity and extent of pipeline infrastructure. The latter is primarily down to the presence of several land-locked fields in the U.S. and Canada which lead to extensive pipeline-miles traversed between sources and destinations.

To further examine emissions from major producer regions to major consumer regions based on source, the above aggregates are segmented by pipeline transport, shipping transport and other transport (the "other" category includes transport along non-pipeline, non-shipping edges in the network and is calculated by using the country-specific default per-mile carbon intensity).

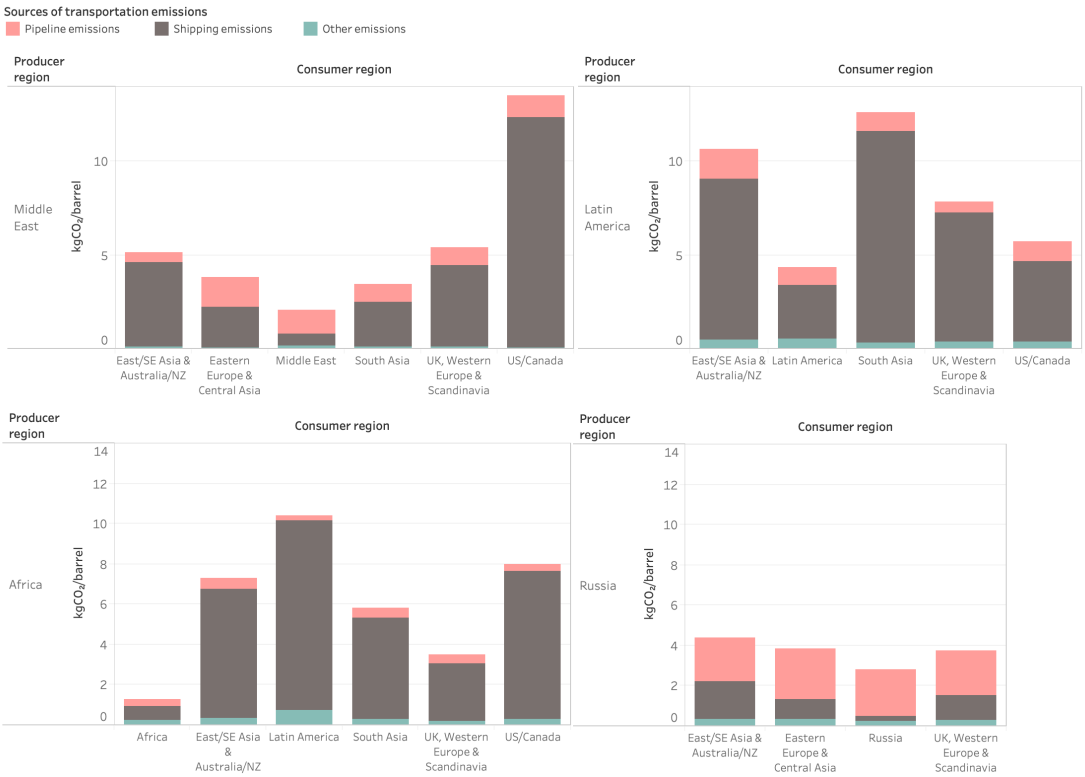

Figure 4-6: Transportation carbon intensities from producer regions to consumer regions broken down by sources of emissions

The mode-specific segmentation of emissions illustrates the carbon heterogeneity in the supply chain. While on balance, the carbon intensity of shipping emissions is greater than that of pipeline emissions, the overall CO_2 impact depends on which mode is more prominent.

For instance, the extensive pipeline coverage in Russia leads to pipeline transport into Europe and China thereby skewing transportation emissions away from shipping.

Within the space of shipping emissions, the observed variability is down to three main factors: distance, capacity utilization of tankers, and intrinsic tanker engine efficiencies. In a similar vein, the main drivers for pipeline emissions are distance, properties of transported crude, and pipeline diameter. The inter-pathway variability i.e. the most granular carbon intensity estimate across all distinct supply chain pathways, as shown in figure 4-7, is a consequence of the underlying variability in these features.

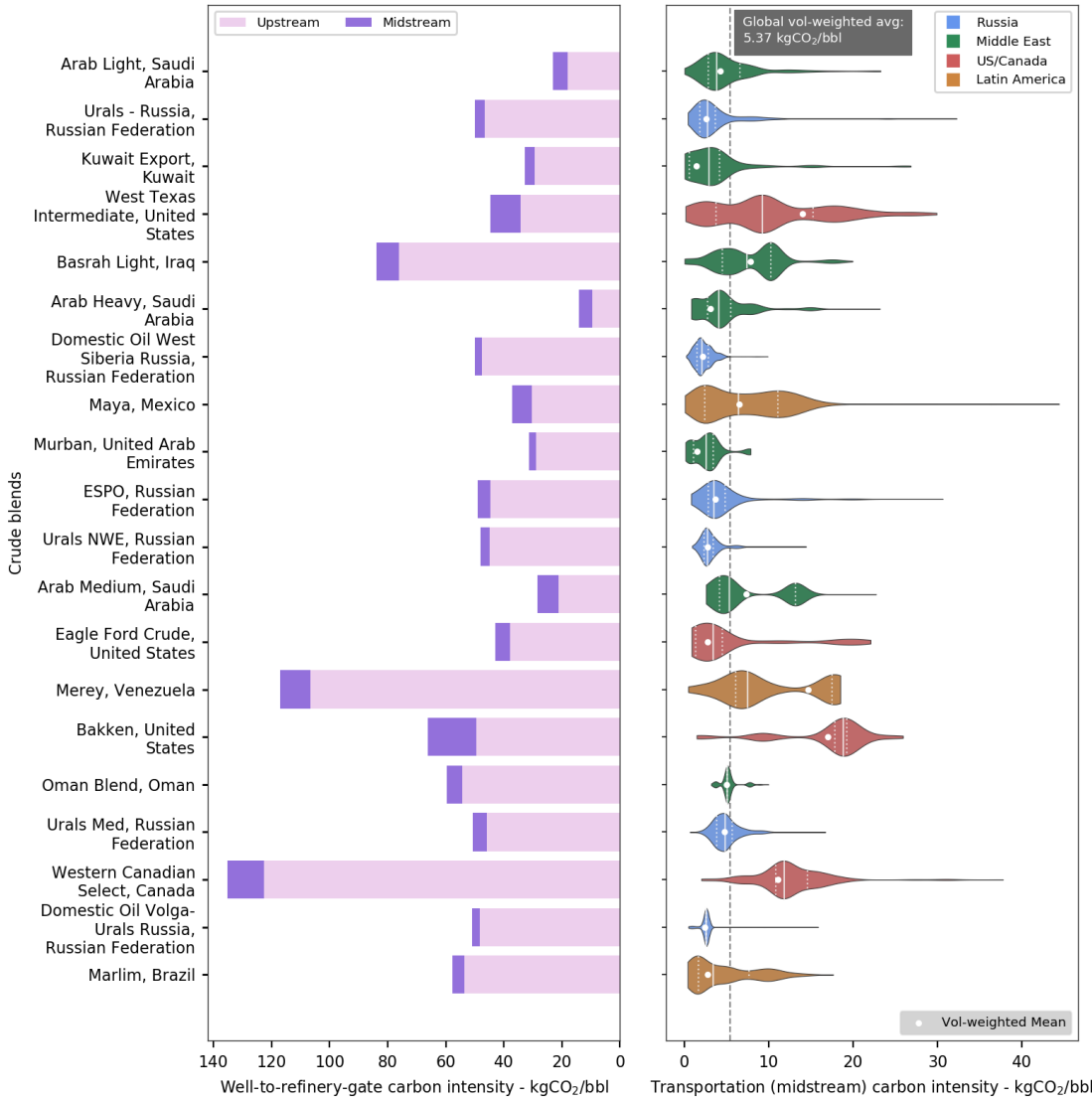

Figure 4-7: Blend-level aggregation for the top 20 blends globally - net upstream, midstream carbon intensity and distributions of midstream carbon intensities

4.3 Net CO_2 emissions attributed to consumer countries

With respect to the consumer perspective, the global volume weighted average is found to be 54.53 kg-CO_2/barrel. Among major consumers (refining volume of >1 million barrels per day), the spread around the average is considerable - 17.54 to 92.21 kg-CO_2/barrel.

Figures 4-8 and 4-9 show the spatial distribution of carbon intensities aggregated at refineries thereby accounting for all the carbon in the well-to-refinery-gate scope.

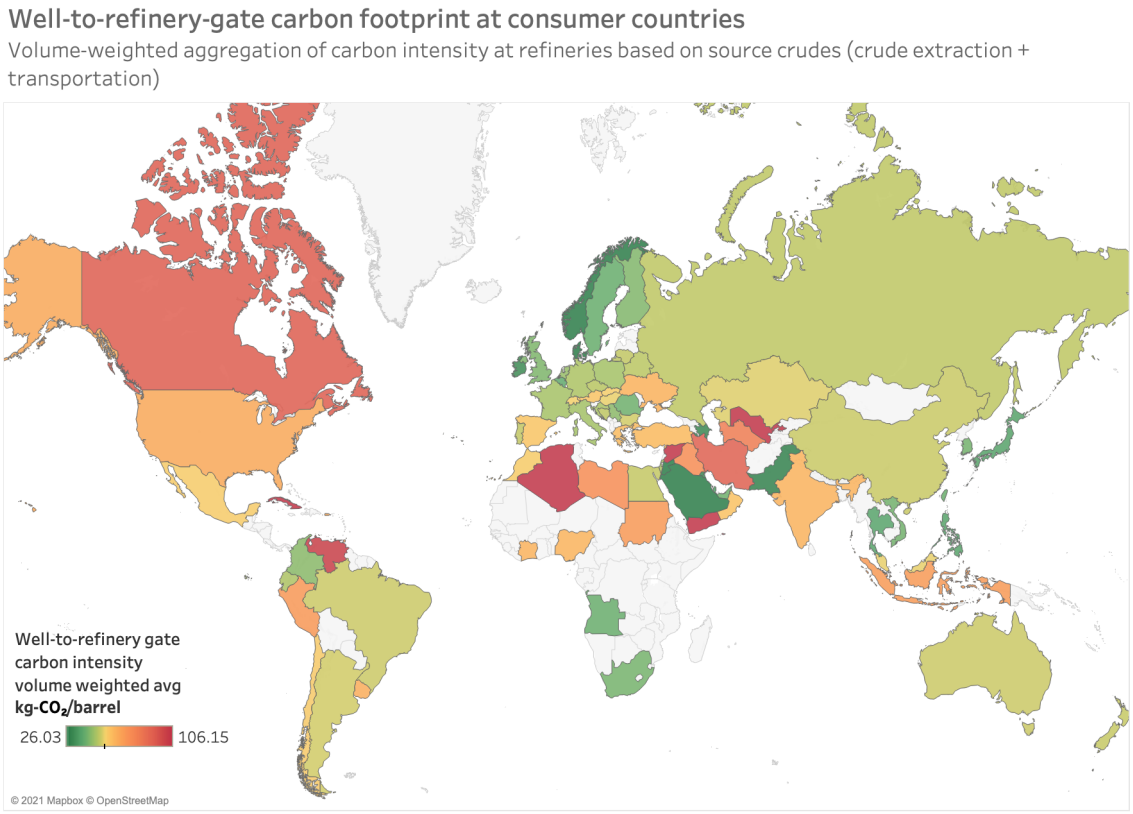

Figure 4-8: Overall carbon intensity of source crudes for consumer countries

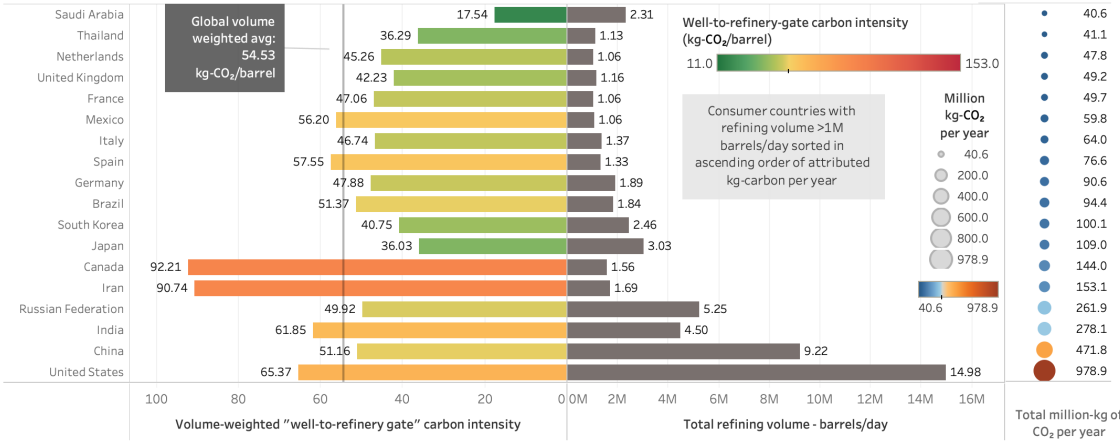

Figure 4-9: Carbon intensity and net annual CO_2 emissions at the level of consumer countries - for countries with >1 million-barrels/day refining volume

Chapter 5

Policy Implications

The carbon intensity estimates at different levels of aggregation confirm the hypothesis of the thesis - there is significant heterogeneity in the life cycle emissions of the oil supply chain. Pathway-level carbon intensities vary from 1.80 to 32.92 gCO_2/MJ with a volume weighted mean of 9.73 gCO_2/MJ.

This insight is significant not just in the present but also in the future. Within the existing sources of primary energy, carbon-based differentiation is an active policy lever as discussed in Chapter 1. Most notably, the Low Carbon Fuel Standard by the California Air Resources Board [2] and the Fuel Quality Directive by European regulators [3] are examples of regulatory efforts to examine the carbon intensity of primary energy sources with the vision of steady decarbonization in the near-term. In the future, with the ongoing and imminently accelerating energy transition, this heterogeneity in carbon intensity has immense potential to act as a powerful decarbonization tool as discussed below.

Table 5.1 below shows oil supply forecasts depending on different models and scenarios. The SSPs or the Shared Socioeconomic Pathways are global scenarios of projected socioeconomic changes [35].

The scenarios are:

- SSP1: Sustainability (Taking the Green Road)
- SSP2: Middle of the Road

- SSP3: Regional Rivalry (A Rocky Road)
- SSP4: Inequality (A Road Divided)
- SSP5: Fossil-fueled Development (Taking the Highway)

Table 5.1: Oil supply projections under different projection models and policy scenarios

Model	Scenario	SSP	Temp target	Unit	2010	2020	2030	2040	2050
IEA	Stated Policy			EJ/yr		214.07	223.38	232.32	241.25
IEA	Sustainable Development Scenario		2	EJ/yr		208.86	192.11	147.43	102.75
BP	Rapid		2	EJ/yr		220.40	205.51	163.07	116.16
BP	Net zero		1.5	EJ/yr		220.40	205.51	129.56	67.01
BP	BAU			EJ/yr		224.10	227.85	212.21	205.51
AIM/CGE 2.0	SSP1-19	1	1.5	EJ/yr	168.76	181.96	146.99	112.19	83.69
AIM/CGE 2.0	SSP2-19	2	1.5	EJ/yr	175.93	208.33	160.22	123.75	116.21
IMAGE 3.0.1	SSP1-19	1	1.5	EJ/yr	171.72	154.89	86.59	43.60	33.69
AIM/CGE 2.0	SSP1-26	1	2	EJ/yr	168.79	181.96	184.75	161.07	146.47
AIM/CGE 2.0	SSP2-26	2	2	EJ/yr	176.08	208.27	208.45	193.14	192.77
AIM/CGE 2.0	SSP4-26	4	2	EJ/yr	171.18	192.16	192.79	172.32	160.01
AIM/CGE 2.0	SSP5-26	5	2	EJ/yr	174.60	205.95	200.74	196.13	223.53
IMAGE 3.0.1	SSP1-26	1	2	EJ/yr	171.72	171.15	149.20	122.02	110.06

Table 5.1 continued from previous page

IMAGE 3.0.1	SSP2-26	2	2	EJ/yr	171.87	167.89	122.66	92.65	82.17
IMAGE 3.0.1	SSP4-26	4	2	EJ/yr	173.71	164.42	125.95	77.66	44.41

Under the subset of decarbonization scenarios with 1.5 - 2 ^{o}C temperature targets, oil supply forecasts show decreasing and/or plateauing behaviors. This exhibits an opportunity for significant additional decarbonization as quantified below. This additional decarbonization can be realized by virtue of prioritizing the phase-out of supply chain pathways with higher carbon intensities.

Figure 5-1 shows the additional annual CO_2 savings that can be realized under different scenarios by prioritizing trade for low carbon intensity pathways. The curve is generated by sorting the granular pathway carbon intensities and eliminating barrels of crude sequentially in order of higher-to-lower carbon intensities. The scatter points correspond to the 2050 supply values for the different model, scenario combinations and are placed at the corresponding locations on the curve.

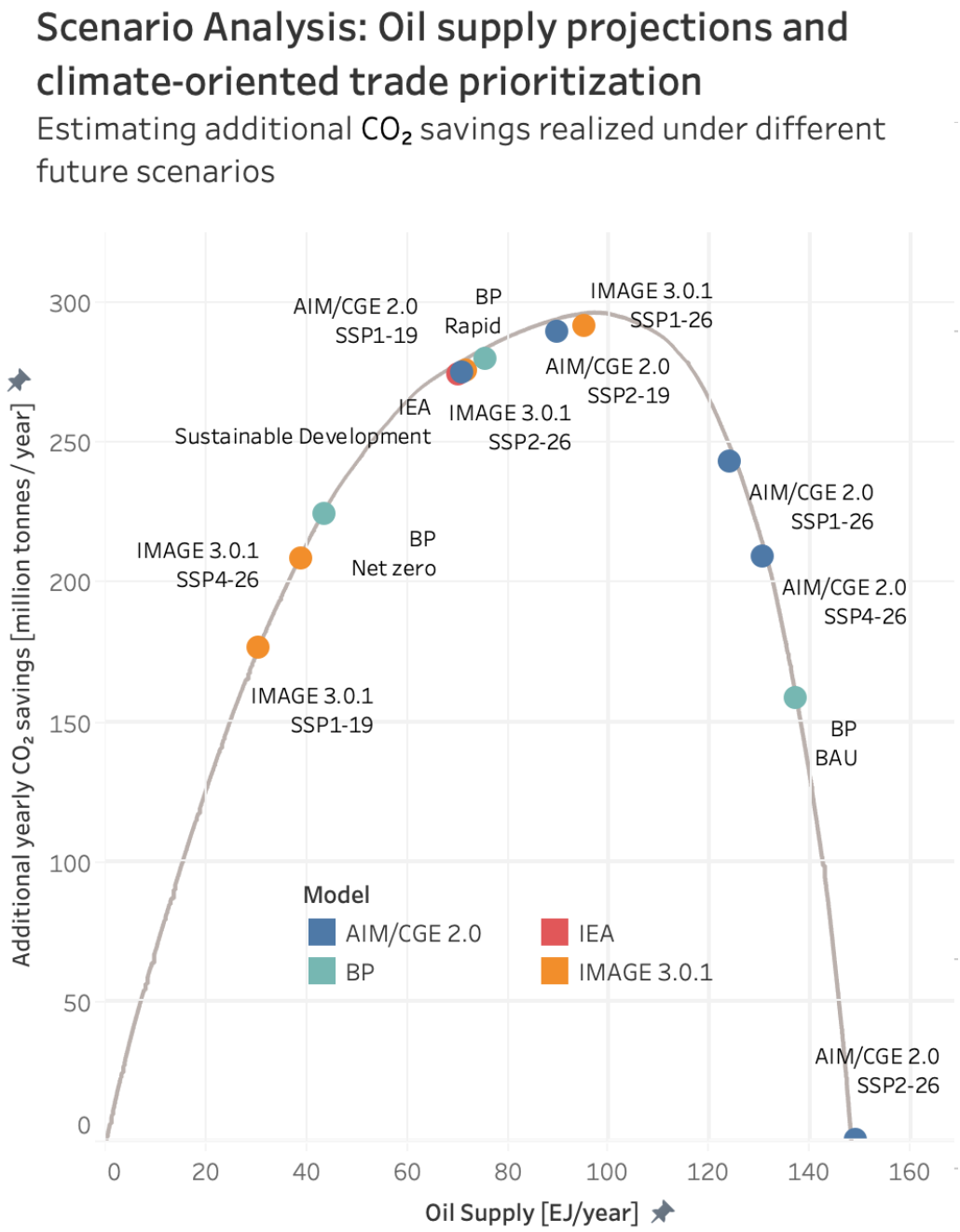

Figure 5-1: Scenario analysis - trade prioritization optimized for the climate

These model, scenario combinations correspond to trends in future carbon intensities based on different supply projections. The carbon intensity time series in turn lead to different cumulative carbon savings as shown in figure 5-2. The study thus provides the foundation for future supply fulfilment optimized for CO_2 emissions.

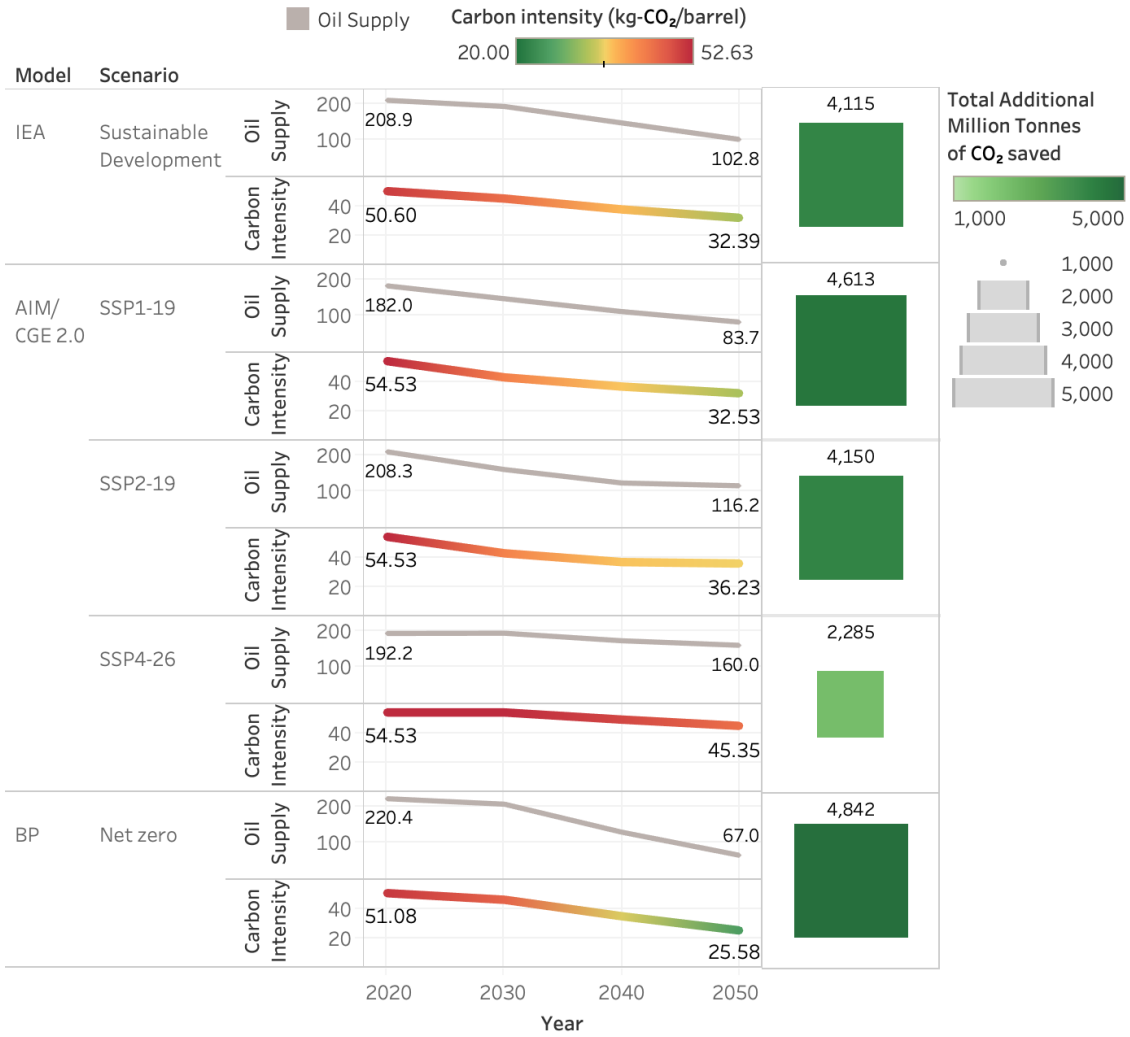

Figure 5-2: Scenario analysis - time series of crude carbon intensity and cumulative CO_2 savings

Thus, on a spectrum from SSP-4 to net-zero scenarios, the additional CO_2 savings amount to 2-5 GT. To put this into context, this is comparable in magnitude to removing ~100 million passenger cars, assuming a typical car is driven for 10 years and emits 4.6 tonnes of carbon per year [36].

Moreover, along with the quantification of the net CO_2 impact, the high resolution of the underlying data points towards appropriate carbon pricing through mechanisms such as credit systems that price in the pathway-level heterogeneity. The different levels of aggregation (pathway, blend trade, country), facilitate policy flexibility which matters given the political challenges of getting pricing schemes to work. This flexibility can manifest through supply contracts with dedicated carbon clauses thus giving greater agency to regulators and policy institutions.

Furthermore, the significance of these emissions over the 30 year horizon as shown in 5-2, can motivate real-time granular carbon reporting from industry. The right reporting systems can lead to better and more data which can enable accurate carbon inventories which in turn points to more effective decarbonization through policy and business strategy.

Chapter 6

Conclusion

6.1 Heterogeneity in life-cycle CO_2 emissions

The thesis concludes that globally, the carbon footprint variability at the pathway-level (pathways are defined as the routes from oil fields to refineries) ranges from 1.80 to 32.92 gCO_2/MJ with a volume weighted mean of 9.73 gCO_2/MJ.

Within the subset of the top 20 crude blends by volume, the carbon intensity ranges from 4.16 to 23.11 gCO_2/MJ.

The two main underlying sources guiding this variability are CO_2 intensive field-level operations along with distances and inefficiencies in transportation networks.

6.2 Policy insights

The heterogeneity in well-to-refinery-gate emissions associated with different marketed crudes is a key decarbonization opportunity in the present and in the near future.

With the increasing policy intent [2, 3], the study lays the foundation to account for these differences through either market-based policy schemes or command-and-control regulation. The former can manifest by pricing in these differences and/or through carbon credits which incentivize best practices, while the latter through demand prioritization.

Coupled with supply forecasts up to 2050 from low-carbon scenarios, the variability in carbon intensity translates to additional CO_2 savings of 2-5 GT. These savings can be realized through a CO_2-oriented supply optimization as described in Chapter 5.

6.3 Future Work

The research fills the information gap of well-to-refinery-gate carbon intensities at a high resolution. This work can be augmented through further research across the modeling pipeline.

First, better high-fidelity data that is granular in time as opposed to annually averaged would improve the pathway tracking estimates. Further, the benefits of having complementary features (geolocations, operations data, etc.) as shown by current data streams, can guide future data collection efforts (e.g: terminal capacities).

Second, improved infrastructure data has the potential to make the physics-based emission estimations more effective. Third, accounting for the spatial extent of supply chain assets to route barrels would represent an improvement over the current implementation that assumes the assets to be point entities. Fourth, the weights that guide the blend estimation algorithm could be customized according to the specifics of supply chains in different countries such that the weights reflect the physical realities of crude blending.

In closing, the study seeks to motivate further improvements in supply chain modeling to improve the quality of life cycle assessments. The data consolidation and modeling pipeline has the potential to be the foundation for allied research into techno-economic analysis, thereby making the analyses more holistic towards guiding decarbonization policy.

Bibliography

1. *World Energy Outlook 2019*. English. OCLC: 1243142150 (OECD Publishing., 2019).
2. *Low Carbon Fuel Standard | California Air Resources Board* Retrieved from https://ww2.arb.ca.gov/our-work/programs/low-carbon-fuel-standard (May 2, 2021).
3. *Fuel Quality* en. Text. Nov. 2016. Retrieved from https://ec.europa.eu/clima/policies/transport/fuel_en (May 2, 2021).
4. Ruehl, C. & Giljum, J. BP Energy outlook 2030. *Energy* **2030** (2011).
5. *ExxonMobil debuts indirect emissions data from consumption of its products* Jan. 2021. Retrieved from https://ihsmarkit.com/research-analysis/exxonmobil-debuts-indirect-emissions-data-from-consumption-of-.html (May 3, 2021).
6. Brandt, A. R. Variability and uncertainty in life cycle assessment models for greenhouse gas emissions from Canadian oil sands production. *Environmental Science & Technology* **46,** 1253–1261 (2012).
7. Cai, H. *et al.* Well-to-wheels greenhouse gas emissions of Canadian oil sands products: Implications for US petroleum fuels. *Environmental Science & Technology* **49,** 8219–8227 (2015).
8. El-Houjeiri, H. M., Brandt, A. R. & Duffy, J. E. Open-source LCA tool for estimating greenhouse gas emissions from crude oil production using field characteristics. *Environmental Science & Technology* **47,** 5998–6006 (2013).
9. Meehan, D. N., El-Houjeiri, H. M., Rutherford, J. S., *et al. Carbon intensity: Comparing carbon impacts of middle east and US shale oils* in *SPE Kingdom of Saudi Arabia Annual Technical Symposium and Exhibition* (2018).
10. Masnadi, M. S. *et al.* Global carbon intensity of crude oil production. *Science* **361,** 851–853 (2018).

11. Vafi, K. & Brandt, A. R. Reproducibility of LCA models of crude oil production. *Environmental Science & Technology* **48,** 12978–12985 (2014).
12. Masnadi, M. S. *et al.* Well-to-refinery emissions and net-energy analysis of China's crude-oil supply. *Nature Energy* **3,** 220–226 (2018).
13. Burnham, A., Wang, M. & Wu, Y. *Development and applications of GREET 2.7–The Transportation Vehicle-CycleModel.* tech. rep. (Argonne National Lab.(ANL), Argonne, IL (United States), 2006).
14. Choquette-Levy, N., Zhong, M., MacLean, H. & Bergerson, J. COPTEM: a model to investigate the factors driving crude oil pipeline transportation emissions. *Environmental Science & Technology* **52,** 337–345 (2018).
15. Nian, V. & Yuan, J. A method for analysis of maritime transportation systems in the life cycle approach–The oil tanker example. *Applied Energy* **206,** 1579–1589 (2017).
16. Olmer, N., Comer, B., Roy, B., Mao, X. & Rutherford, D. Greenhouse Gas Emissions from Global Shipping, 2013–2015 Detailed Methodology. *International Council on Clean Transportation: Washington, DC, USA,* 1–38 (2017).
17. Greene, S., Jia, H. & Rubio-Domingo, G. Well-to-tank carbon emissions from crude oil maritime transportation. *Transportation Research Part D: Transport and Environment* **88,** 102587 (2020).
18. Abella, J. P. *et al.* Petroleum Refinery Life Cycle Inventory Model (PRELIM) PRELIM v1. 3 (2019).
19. *Wood Mackenzie provides the most trusted research in the industry* en. June 2017. Retrieved from https://www.woodmac.com/research/ (May 3, 2021).
20. *Kpler - Leading Commodity Data and Analytics Solution* Retrieved from https://www.kpler.com/ (May 3, 2021).
21. *GlobalData Login* Retrieved from https://login.globaldata.com/login/index/oilgas?ReturnUrl=%2fHomePage (May 3, 2021).
22. *To make better decisions, you need to see the big picture.* Retrieved from https://ihsmarkit.com/products/ship-and-port-data.html (May 3, 2021).
23. Wan, Zhengming, Hook, Simon & Hulley, Glynn. *MOD11C3 MODIS/Terra Land Surface Temperature/Emissivity Monthly L3 Global 0.05Deg CMG V006* type: dataset. 2015. Retrieved from https://lpdaac.usgs.gov/products/mod11c3v006/ (May 2, 2021).
24. *Digital Elevation Data - with SRTM voids filled using accurate topographic mapping* Retrieved from http://www.viewfinderpanoramas.org/dem3.html (May 2, 2021).

25. Paszke, A. *et al.* Automatic differentiation in pytorch (2017).
26. Qian, N. On the momentum term in gradient descent learning algorithms. *Neural Networks* **12,** 145–151 (1999).
27. Ruder, S. An overview of gradient descent optimization algorithms. *arXiv preprint arXiv:1609.04747* (2016).
28. Vose, M. D. *The simple genetic algorithm: foundations and theory* (MIT press, 1999).
29. Likas, A., Vlassis, N. & Verbeek, J. J. The global k-means clustering algorithm. *Pattern recognition* **36,** 451–461 (2003).
30. Paszke, A. *et al.* Pytorch: An imperative style, high-performance deep learning library. *arXiv preprint arXiv:1912.01703* (2019).
31. Stopford, M. *Maritime Economics 3e* (Routledge, 2008).
32. *Oil tanker sizes range from general purpose to ultra-large crude carriers on AFRA scale - Today in Energy - U.S. Energy Information Administration (EIA)* Retrieved from `https://www.eia.gov/todayinenergy/detail.php?id=17991` (May 2, 2021).
33. Zhang, Y., Fung, J. C., Chan, J. W. & Lau, A. K. The significance of incorporating unidentified vessels into AIS-based ship emission inventory. *Atmospheric Environment* **203,** 102–113 (2019).
34. Selin, H., Zhang, Y., Dunn, R., Selin, N. E. & Lau, A. K. Mitigation of CO_2 emissions from international shipping through national allocation. *Environmental Research Letters* (2021).
35. O'Neill, B. C. *et al.* Achievements and needs for the climate change scenario framework. *Nature Climate Change,* 1–11 (2020).
36. US EPA, O. *Greenhouse Gas Emissions from a Typical Passenger Vehicle* en. Overviews and Factsheets. Jan. 2016. Retrieved from `https://www.epa.gov/greenvehicles/greenhouse-gas-emissions-typical-passenger-vehicle` (May 9, 2021).