Independent Statistics & Analysis
U.S. Energy Information Administration

World Hydrocarbon Activity Model (WHAM) Component Design Report

April 2022

Independent Statistics & Analysis
www.eia.gov

U.S. Department of Energy
Washington, DC 20585

This report was prepared by the U.S. Energy Information Administration (EIA), the statistical and analytical agency within the U.S. Department of Energy. By law, EIA's data, analyses, and forecasts are independent of approval by any other officer or employee of the U.S. Government. The views in this report should not be construed as representing those of the U.S. Department of Energy or other federal agencies.

Executive Summary

This document presents the Component Design Report for the new World Hydrocarbon Activity Model (WHAM) to be integrated into the World Energy Projection System (WEPS) that we use to produce the *International Energy Outlook* (IEO). WHAM will replace the Global Hydrocarbon Supply Model (GHySMo) that was used for IEO2019.

This Component Design Report reviews the major design elements for WHAM, identifying data input and modeling approach similarities with GHySMo. WHAM will use a linear program to model global refinery operations to satisfy product demand levels imposed by other WEPS modules; the model structure and core mathematical equations are provided in this document. WHAM and GHySMo differ in some key ways:

- WHAM will attempt to improve model robustness and reliability by using a single linear program to perform optimization calculations. This linear program will be responsible for optimizing both global refinery operations and logistics.
- WHAM will not attempt to rigorously model upstream oil and natural gas production, but will incorporate an estimation method based on third-party projected global oil and natural gas production levels and the associated cost of production.

The primary goal of developing a new global hydrocarbon model is to provide WEPS modelers with a model that is more robust and easier to maintain than previous tools. This model meets these needs while maintaining a detailed and accurate representation of global hydrocarbon transport and processing.

Contents

Figures

Tables

Revision History

Name	Date	Reason for Changes	Version
WHAM Project Team	March 9, 2022	New document	1.0

Abbreviations

AIMMS	Advanced Integrated Multidimensional Modeling Software
AEO	*Annual Energy Outlook*
b	barrel
CTL	coal-to-liquids
EIA	U.S. Energy Information Administration
GTL	gas-to-liquids
GLAM	Global Activity Module
GHySMo	Global Hydrocarbon Supply Model
GRTMPS	Generalized Refining Transportation Marketing Planning System
IEMM	International Electricity Market Model
IEA	International Energy Agency
IEO	*International Energy Outlook*
IES	International Energy Statistics (EIA)
LNG	liquefied natural gas
LPG	liquefied petroleum gas
LTEM	Office of Long-Term Energy Modeling
Mb	thousand barrels
MMBtu	million British thermal units
MMb/d	million barrels per day
NGPL	natural gas plant liquids
OECD	Organization of Economic Cooperation and Development
OPEC	Organization of Petroleum Exporting Countries
STEO	*Short-Term Energy Outlook*
Mcf	thousand cubic feet
TBtu	trillion British thermal units
Tcf	trillion cubic feet
WHAM	World Hydrocarbon Activity Model
WEPS	World Energy Projection System
WIM	World Industrial Model
WOP	world oil price
WTI	West Texas Intermediate

Statement of Purpose

The World Hydrocarbon Activity Model (WHAM) is a new model that represents global natural gas and petroleum liquids markets to support the international modeling efforts of the U.S. Energy Information Administration (EIA). Previously, we used the Global Hydrocarbon Supply Module (GHySMo) during the production of the *International Energy Outlook 2019* (IEO2019) and the Oil & Gas Tool for IEO2021. GHySMo consisted of three modules interacting via an Integration module:

- Upstream, run outside of each iteration of the World Energy Projection System (WEPS)
- Conversion, a linear program representing international refining
- Logistics, a linear program representing transportation and trade of hydrocarbons

After extensive evaluation, we determined that the existing GHySMo design did not meet the design objective because the model could not adequately represent the intended markets. As a result, for IEO2021, we built the Oil & Gas Tool, which relied on previous analysis. However, the Oil & Gas Tool cannot model economic behavior and therefore is unable to capture complex international trade interactions.

WHAM will represent core global natural gas and liquid fuel markets by optimizing a single linear program for refining and global movements of hydrocarbon commodities, eliminating the need for convergence between multiple linear programs. This solution will include volumes, prices, and flows to be passed into WEPS.

Purpose of document

The purpose of this document is to facilitate project development by outlining the details of the design, structure, and mathematical formulation of WHAM prior to development.

This document is intended for EIA management, the LTEM team developing WHAM, and future users of WHAM.

Software considerations

We are developing WHAM using the AIMMS platform, our current platform of choice for optimization modeling. This adds some additional constraints to the system, including the formatting and file types for input and output files. Any supporting code or code related to the main linear program will use Python.

WEPS runs inside a Python virtual environment managed by the EIA Integration Team. The virtual environment enables close control of the Python packages used by WEPS.

Other software considerations include:

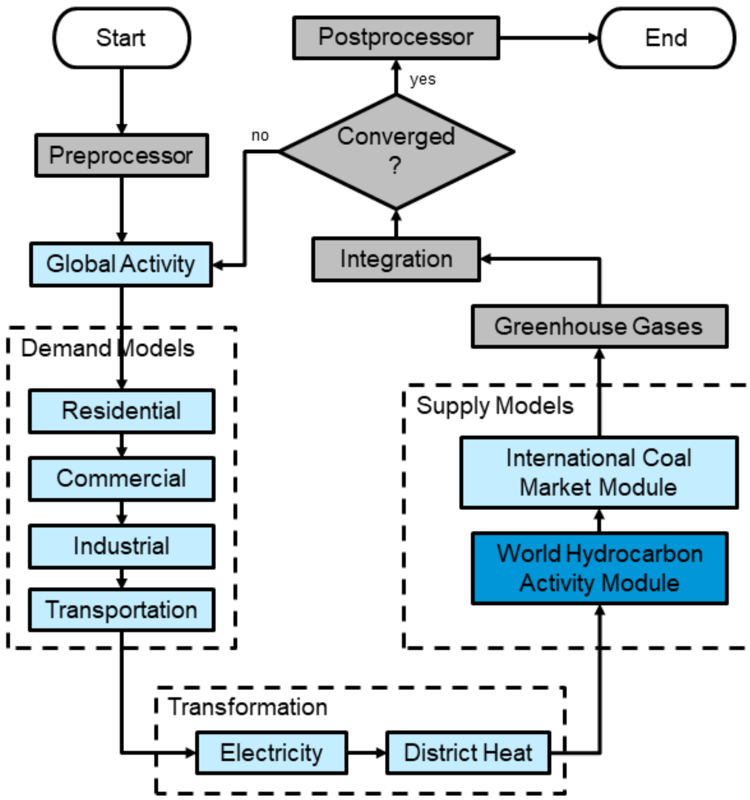
- As of November 20, 2020, WEPS runs in Python 3.7 (64 bit).
- The restart file (*restart.hdf5*) is formatted as a binary *hdf5* file. WHAM must read input variables from and write output variables to the WEPS *hdf5* restart file.
- Associated pre- and post-processing of data can be done either within AIMMS or via Python.
- New model runtime should be no more than 20 minutes.

Input and Output Requirements

General requirements for WEPS

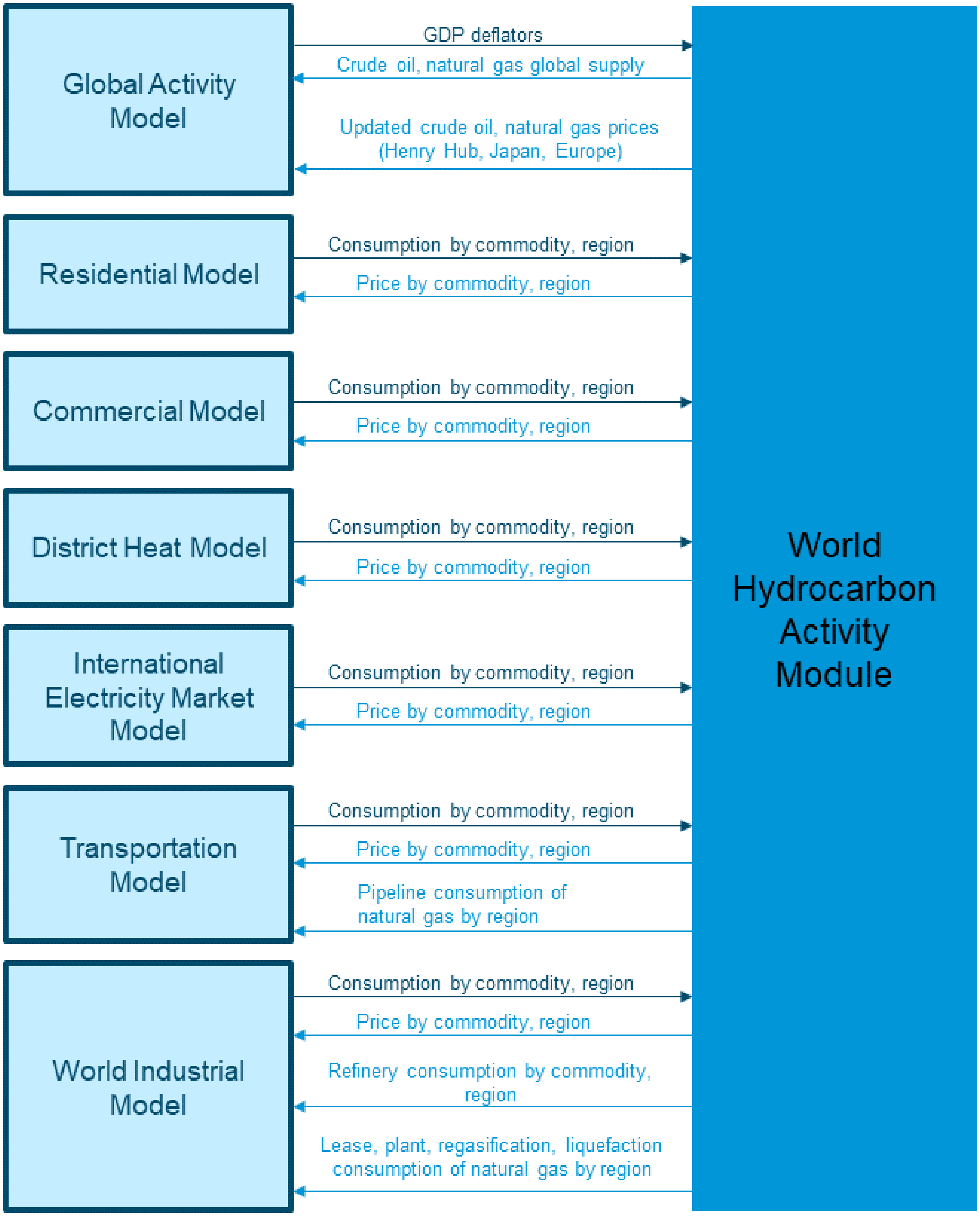
WHAM projects the global price and quantity of liquid and natural gas hydrocarbons within WEPS. It will be one of 11 individual models that make up WEPS (**Error! Reference source not found.**).

Figure 1. Schematic of WEPS



WHAM receives data from and provides data to other WEPS modules (Figure 2).

Figure 2. Interactions between WHAM and WEPS by module



WHAM receives annual projections through 2050 of the following:

- Consumption of individual petroleum liquid commodities by end-use sector and WEPS region
- Consumption of natural gas by end-use sector and WEPS region
- Initial Brent crude oil price projection (called the World Oil Price [WOP] path)

WHAM provides delivered hydrocarbon commodity prices back to the demand and transformation models. It also provides additional consumption volumes related to hydrocarbon production, refining, and transportation: lease and plant fuel, pipeline fuel use, liquefaction and regasification volumes, and refinery consumption.

The WHAM project development scope:

- Provides, at minimum, a single aggregated global supply of crude oil and natural gas (in other words, upstream production)
- Balances supply and demand of petroleum and other liquids and natural gas
- Represents global refining to deliver end-use product prices by sector to WEPS
- Represents global transportation between suppliers, refineries, and demand markets consistent with existing infrastructure (Requirement 2.6 in Appendix A discusses expanding infrastructure)
- Uses historical data where appropriate
- Benchmarks to our *Annual Energy Outlook* (AEO) and *Short-Term Energy Outlook* (STEO) where appropriate
- Ensures consistency with our International Energy Statistics (IES) and the International Energy Agency (IEA) historical data processing and data reconciliation process
- Includes documentation to facilitate future use of model
- Results in a generalized model that can be easily accommodate future enhancements

As part of an integrated WEPS cycle, WHAM:

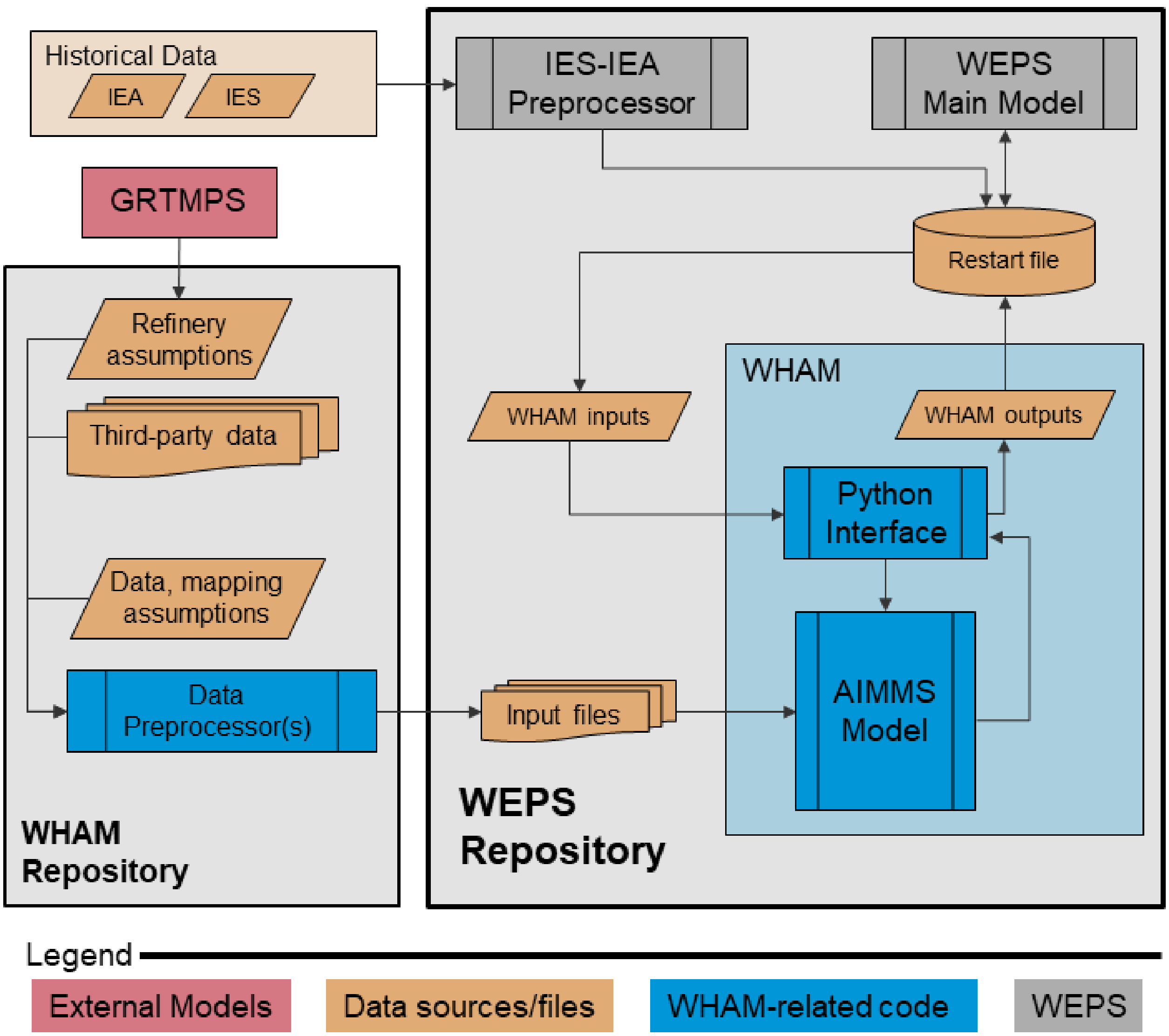
- Provides supply projections for liquid fuels (that is, crude oil, condensate, and other liquids) and natural gas by region for publication in the IEO
- Represents the transformation of crude oil and other liquids in the refining process into petroleum products such that the quantity supplied equals the quantity demanded, as projected by WEPS
- Projects consumption of energy during the production, transformation, and distribution of hydrocarbons
- Projects spot prices as-needed for other WEPS modules and constructs end-use prices
- Projects end-use commodity prices by sector, commodity, and WEPS demand region as expected by other WEPS modules
- Represents trade of hydrocarbons using commodity flows between supply regions, refining regions, and demand regions
- Benchmarks projections to both the latest AEO and STEO, where applicable

Input and output boundaries of WHAM

The requirements classified as integration requirements refer to any other software, interfaces, or systems related to WHAM (Figure 3).**Error! Reference source not found.** WHAM interacts with three ain systems:

- WEPS, via restart file and the WHAM integration wrapper. All files needed for a WEPS model run, including source code and input files, will be part of the WEPS Repository
- WHAM Repository, where preprocessors, input data, and other offline files will be stored
- GRTMPS, offline software to generate refinery yield and operation assumptions

Figure 3. WHAM system context



WHAM reporting and debugging

Currently, the primary WEPS output has five IEO tables designated for WHAM's output:

- Table G01: World petroleum and other liquids production
- Table G02: World crude oil production by region and country
- Table G03: World other liquid fuels production
- Table I01: World natural gas production by region
- Table I02: World net trade in natural gas by region

The information in these output tables is the core requirement for WHAM output results. In addition, WHAM consumption volumes related to hydrocarbon production, refinery operations, and transmission are a component in output tables dedicated to other demand and conversion modules. We will save the WHAM linear program structure and output to for debugging purposes.

WHAM will create summary reports to assist with data and model visualization and debugging. These reports will be created and used through AIMMS and will be the primary method for modelers to view and analyze model results. Reports will typically include price and quantity information and will be generated as part of the development process.

We are considering the following items to facilitate debugging the model, understanding model results, and deconstructing interactions within other modules of WEPS:

- AIMMS report pages to visualize model outputs in the AIMMS environment
- Recording restart file data in .csv format for reference
- Error handling code to prevent WEPS crashes in case of WHAM errors
- Compartmentalization in Integration Wrapper code to ensure that model results are not overwritten

WHAM documentation

Mathematical model documentation will be produced consistent with AIMMS implementation. The WHAM model documentation includes:

- A full set of mathematical equations used in the optimization model
- Plain language description of WHAM's mathematics and its translation to AIMMS
- Pre- and post-processing tools
- Complete list of inputs and outputs
- Visualizations of the modeling system, information flow, and curves constructed as part of the objective function (if applicable)

WHAM data sources

Assumptions and dependencies

The requirements and project schedule include a number of assumptions:

Oil prices

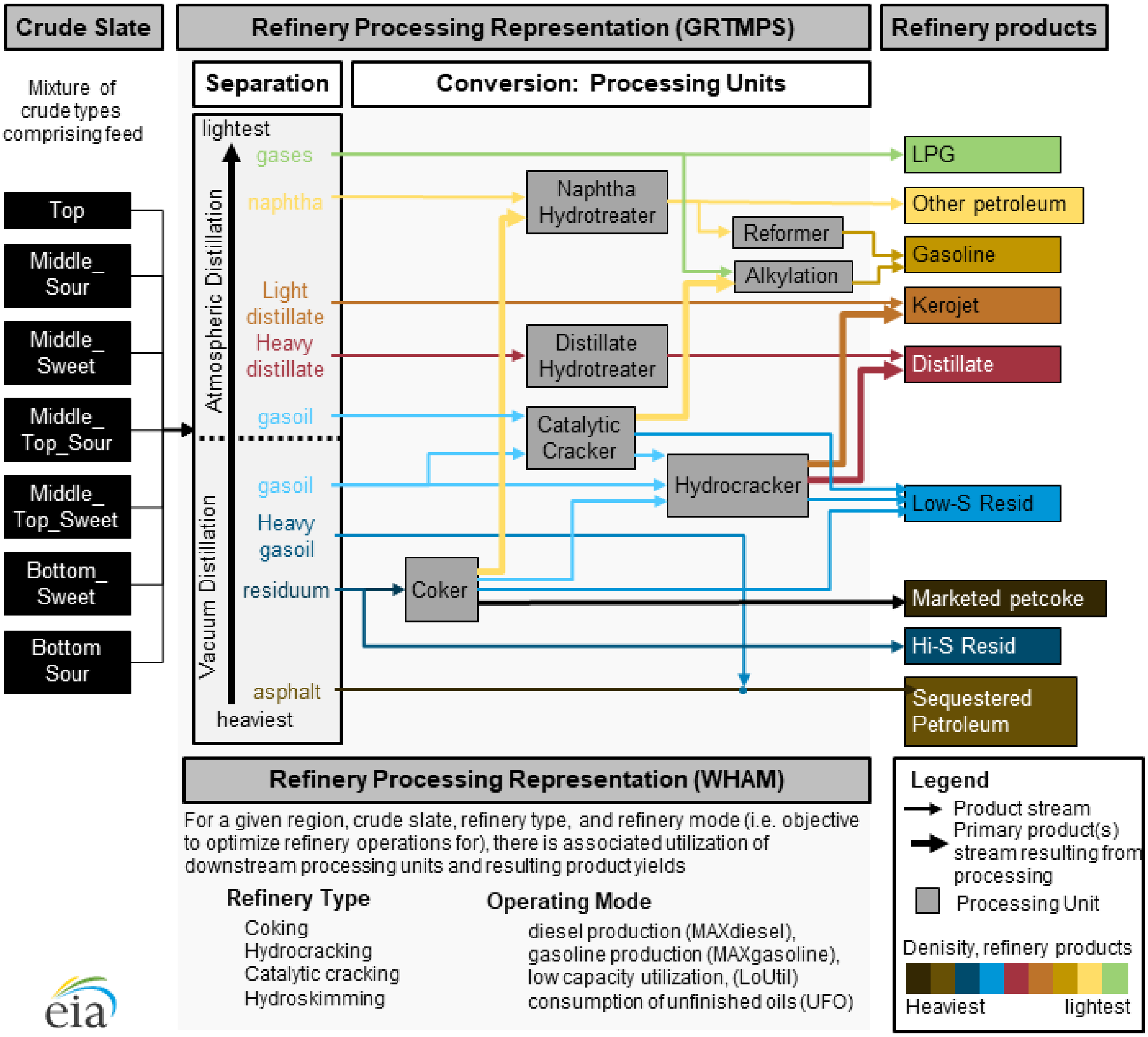
We will continue to provide the World Oil Price (WOP) projections, based on Brent crude oil, as an exogenous input into our domestic and international models; this practice is assumed to hold for both the Reference case and the high and low oil price side cases. This model is not tasked with replacing or evaluating the WOP projections provided; however, the model might adjust the WOP given changes in total global consumption of petroleum products.

WHAM uses Brent oil price projections, in conjunction with price differentials between regions and crude oil quality, to extrapolate the world oil prices for all crude oil types and production regions.

GRTMPS

The Generalized Refining Transportation Marketing Planning System (GRTMPS) refinery model is a critical tool used to generate data for the model. We obtain refinery operations templates from GRTMPS by modeling a sample refinery with varying crude oil slates and refinery configurations. We analyze and condense these templates to provide refinery yield data for a number of different refinery profiles that combine different crude oil slates and refinery types. These data provide the basis for modeling refinery operations in WHAM by serving as the primary operating modes available to the linear program for refinery processing. A simplified representation of how GRTMPS represents a sample refinery is in Figure 4.

Figure 4. Generalized representation of GRTMPS refinery processing



Upstream data handling and generation

The handling and generation of upstream data input for WHAM includes price and production of crude oil, natural gas, and natural gas plant liquids.

For WHAM to allocate the quantity of upstream commodities and the cost of taking quantities of these commodities out of supply regions, a relationship between price and feasible production must be established.

Extensive upstream modeling falls outside of the scope of WHAM, and we will instead aim to simplify the approach to upstream production and pricing. WHAM will leverage third-party data to project breakeven prices for production and reserve quantities aggregated by region and crude oil type through 2050. These data will provide an initial set of production cost curves for input to WHAM. WHAM will receive preprocessed production cost curve and reserves data, which it will use to determine regional production that it can use to meet refinery demand. This determination will provide feasible quantities for a given price point (per WOP), thus constraining supply region production. The WOP path will be passed to WHAM from WEPS. That exogenous price, in conjunction with the regional cost curves, will allow WHAM to allocate regional production through cost minimization.

The most current WOP will remain unmodified. In addition to the WOP, WHAM will take as an input historical differentials to the WOP using the price of Brent crude oil as a proxy. WHAM will use these differentials to calculate the price path by crude oil types. By fixing the price of crude oil to WOP, the model is assuming that upstream suppliers will act as price takers without the ability to affect the price of crude oil.

Using the preprocessed production breakeven price and reserves data, WHAM will apply the relevant WOP to determine a production scenario, which will in turn be used to fulfill refinery region demand requirements. If global production for a given WOP does not meet demand, each region's production will be increased by an amount relative to the region's production share but within the reserve constraint. Reserve constraints will be established based on each region's projected proven reserves.

Many OPEC producers have the potential to produce more low-cost crude oil than they will likely produce in reality. To avoid generating a scenario in which OPEC production is overestimated, WHAM will incorporate an upper and lower constraint to OPEC production equivalent to the highest and lowest share of production that OPEC has realized in the past 20 years.

Prospective data sources

Although access to international data series may constrain the design and development of this model, this document assumes that we continue to have access to existing international data sources. Table 1 identifies data sources that we are considering for use in WHAM, as well as their applications.

Table 1. Prospective data sources for use in WHAM

Data Source	Use
Refinitiv	Refinery capacity and marine transit times
ClipperData	Logistics capacities (pipeline, ship, and liquefaction and regasification) and shipping distances
Bloomberg Terminal	Historical price data
International Gas Union	Natural gas liquefaction and regasification costs as well as capacities
Global Infrastructure Tracker	Pipeline capacities and locations
Oil & Gas Journal	Refinery capacities and locations
GlobalPetrolPrices.com	End-use delivered prices of gasoline and diesel to the transportation sector
Energy Intelligence	Historical crude oil production by API quality, sulfur content, and country; supplemental crude oil price information
Rystad UCube	Global crude oil and natural gas production and reserves

Loss of access to these data providers—or the products provided to EIA staff—may affect both the schedule and requirements of the proposed model.

Classification Plan

Sets

Commodities

Tables 2 through 11 define the set structure for commodities within WHAM as well as within the WEPS system and other EIA products. The root set *Commodity_* is the union of all subsets defined in Table 2. Figure 5 identifies the various commodity sets and their relationships within WHAM and WEPS while Figure 6 provides the key differences for commodity definitions in the following EIA products: IEO publication tables, WHAM, and STEO.

Table 2. WHAM commodity set definitions

Root Set	Index	Subsets	Index
Commodity_	c	CrudeOil_	crude
		UpstreamSupply_	c_upstrm
		RefineryProduct_	c_rf
		PetcokeType_	c_petcoke
		OtherPetroleumType_	c_pet_oth
		RefineryByproduct_	byprod
		RefineryConsumption_	c_cons
OtherLiquids_	liq_oth		
WEPS_Commodity	c_weps		

WHAM uses aggregate crude oil types intended to simulate an average crude oil blend based on a given density (API gravity) and sulfur content. These aggregate crude oils are modeled based on a blend of commonly available crude oil varieties and are designed to provide a variety of feedstocks for the modeled refineries to choose from. The specifications for each crude oil type are defined in Table 3.

Table 3. WHAM crude oil types and definitions

Subset	Index	Description	API gravity	Sulfur content
CrudeOil_	crude_h_hiS	Heavy sour crude oil	≤26	≥1 wt%
	crude_h_loS	Heavy sweet crude oil	≤26	<1 wt%
	crude_m_hiS	Medium sour crude oil	26<API≤36	≥1 wt%
	crude_m_loS	Medium sweet crude oil	26<API≤36	<1 wt%
	crude_l	Light crude oil	36<API≤47	Any
	condensate	Condensate	>47	Any

WHAM will use the set *UpstreamSupply_* to define all supply types that originate from upstream hydrocarbon production activities.

Table 4. WHAM upstream supply type set

Subset	Index	Description
UpstreamSupply_	crude_h_hiS	Heavy sour crude oil
	crude_h_loS	Heavy sweet crude oil
	crude_m_hiS	Medium sour crude oil
	crude_m_loS	Medium sweet crude oil
	crude_l	Light crude oil
	condensate	Condensate
	natgas	Natural gas
	ngpl	Natural gas plant liquids

Table 5. WHAM refinery product set definitions

Subset	Index	Description
RefineryProduct_	pet_seq	Sequestered petroleum
	pet_other_x	Other petroleum (excluding still gas)
	distillate	Distillate
	biofuel	Biofuels
	gasoline	Motor gasoline
	kerojet	Kerosene and jet fuel
	lpg_rf	Liquid petroleum gas from refinery operations
	petcoke_mktd	Marketable petroleum coke
	resid_hiS	High sulfur residuum
	resid_loS	Low sulfur residuum
	ufo*	Unfinished oils

*We are currently debating the precise representation of unfinished oils in WHAM.

Table 6. WHAM petroleum coke variable definitions

Subset	Index	Description
PetcokeType_	catcoke	Catalytic coke
	petcoke_mktd	Marketable petroleum coke

Table 7. WHAM other petroleum variable definitions

Subset	Index	Description
OtherPetroleumType_	stillgas	Refinery fuel gas
	pet_other_x	Other petroleum products

Table 8. WHAM refinery byproduct variable definitions

Subset	Index	Description
RefineryByproduct_	stillgas	Refinery fuel gas
	catcoke	Catalytic coke
	sulfur	Sulfur

Table 9. WHAM refinery utility variable definitions

Subset	Index	Description
RefineryConsumption_	electricity	Electricity
	natgas	Natural gas
	stillgas	Refinery fuel gas
	catcoke	Catalytic coke

Table 10. WHAM other liquid fuels supply set definitions

Subset	Index	Description
OtherLiquids_	biofuels	Biofuels, including ethanol and biodiesel
	oth_ctl	Liquid fuels from coal-to-liquids operations
	oth_gtl	Liquid fuels from gas-to-liquids operations
	ref_gain	Refinery gain
	oth_oth	Other miscellaneous (for example, methanol)

Table 11. WEPS commodity set definitions

Subset	Index	Description
WEPS_Commodity_	crude_oil	Crude oil
	natural_gas	Natural gas
	lpg	Liquid petroleum gas
	distillate	Distillate
	gasoline	Motor gasoline
	kerosene	Kerosene
	jet_fuel	Jet fuel
	resid	Residuum
	pet_seq	Sequestered petroleum
	petcoke	Petroleum coke
	pet_other	Other petroleum products
	qSTEO	STEO total liquid fuels

Figure 5. Commodity and commodity set relationships between WEPS and WHAM

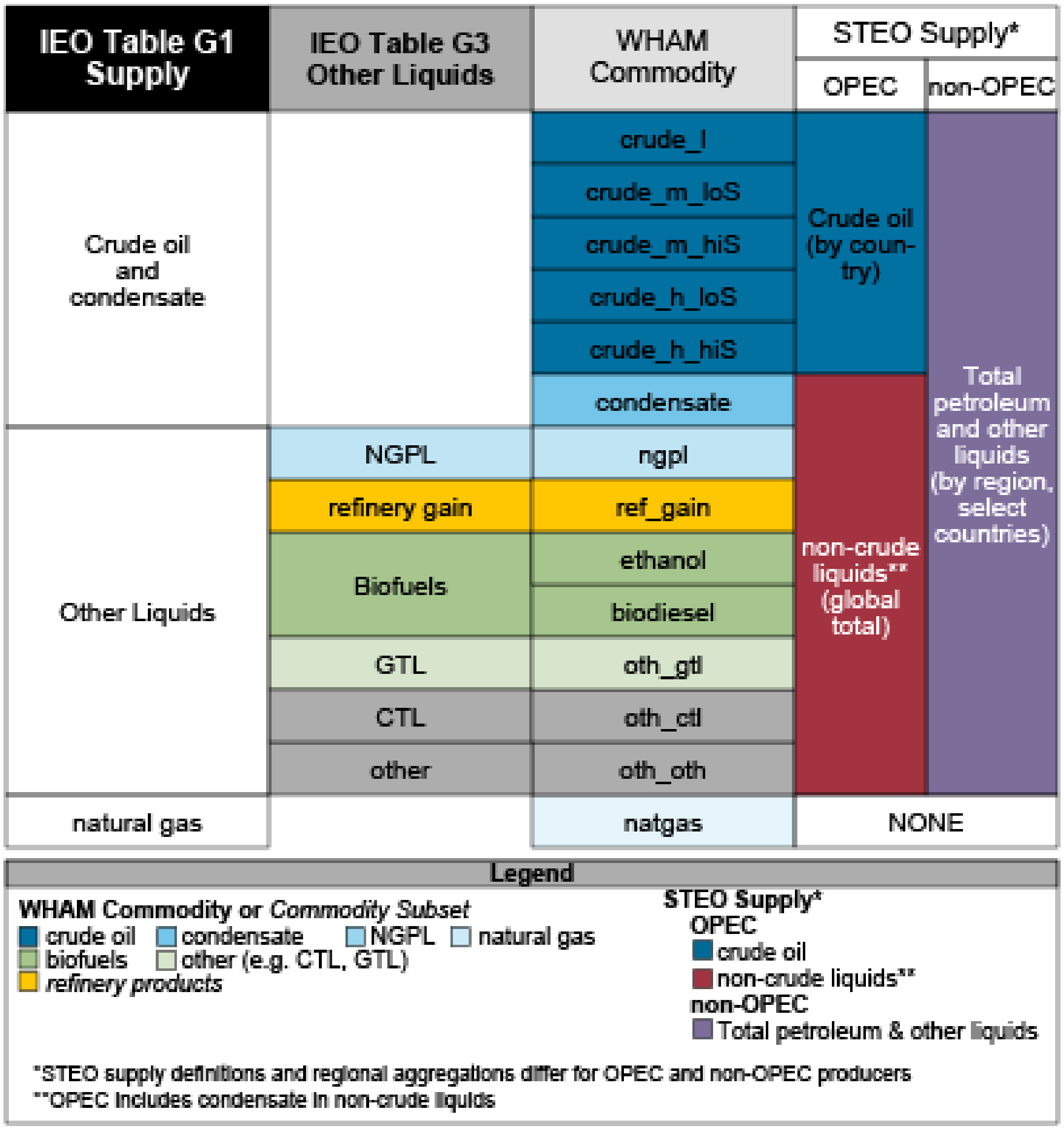
WEPS Product	WHAM Commodity	Description/ Subcomponent	WHAM Commodity Subset
crude oil	crude_l		Supply Types (suptype) (upstream production)
	crude_m_loS		
	crude_m_hiS		
	crude_h_loS		
	crude_h_hiS		
	condensate		
natural gas	natgas		
LPG	ngpl		
	lpg_rf		Refinery Products (c_rf)
motor gasoline	gasoline		
	oth_gtl	GTL supply	other liquids supply, exogenous (liq_oth)
	oth_ctl	CTL supply	
	oth_oth	other (e.g. methanol)	
	biofuels	Ethanol	
distillate		Biodiesel	
	distillate		Refinery Products (c_rf)
jet fuel	kerojet		
kerosene			
other petroleum	pet_other_x		
	stillgas		Refinery Byproducts (byprod)
	sulfur		
petroleum coke	catcoke		
	petcoke_mktd		Refinery Products (c_rf)
sequestered petroleum	pet_seq		
resid	resid_loS		
	resid_hiS		
petroleum (qSTEO)	c_rf	unspecified liquids = STEO consumption	

Legend

Commodity Subset
- Upstream Production
- Other liquids supply
- Refinery products
- Refinery byproducts

Commodity
- crude oil
- condensate
- NGPL
- natural gas
- biofuels (ethanol, biodiesel)
- other (CTL, GTL, methanol)

Figure 6. Commodity and commodity set relationships between WHAM, IEO, and STEO



Refinery process representation

Table 12 lists all the root sets that will be required to define and represent refinery operations within WHAM. Subsequent tables describe the various options that a refinery may have to choose from when optimizing the mathematical program.

Table 12. WHAM refinery set definitions

Root Set	Index	Subsets	Index
RefineryType_	rf_type		
RefineryMode_	rf_mode		
CrudeSlate_	rf_slate		
RefineryUnits_	rf_unit		

Table 13. WHAM refinery type variable definitions

Index	Variable	Description
rf_type	cracking	Refineries with no coking capacity but predominant cracking capability
	coking	Refineries with coking capacity
	hydrocracking	Refineries with no coking capacity but predominant hydrocracking capability
	simple	Refineries with no coking, cracking, or hydrocracking capability

Table 14. WHAM crude slate operating mode definitions

Index	Variable	Description
rf_slate	Bottom_Sour	Composite heavy sour crude oil slate
	Bottom_Sweet	Composite heavy sweet crude oil slate
	Middle_Sour	Composite medium sour crude oil slate
	Middle_Sweet	Composite medium sweet crude oil slate
	Middle_Top_Sour	Composite light-medium sour crude oil slate
	Middle_Top_Sweet	Composite light-medium sweet crude oil slate
	Top	Composite light sweet crude oil slate

Table 15. WHAM refinery operating mode definitions

Index	Variable	Description
rf_mode	MAXdiesel	Refinery operates to maximize ultra-low sulfur diesel production
	MAXgasoline	Refinery operates to maximize motor gasoline production
	LoUtil	Refinery operates at low utilization
	UFO	Refinery operates to consume unfinished oil feedstock

Regions

Within WHAM, different types of regions (supply, refining, and demand) need to be represented. In addition, due to regional differences among EIA publications, different regional representations of the world need to easily relate to one another within the code.

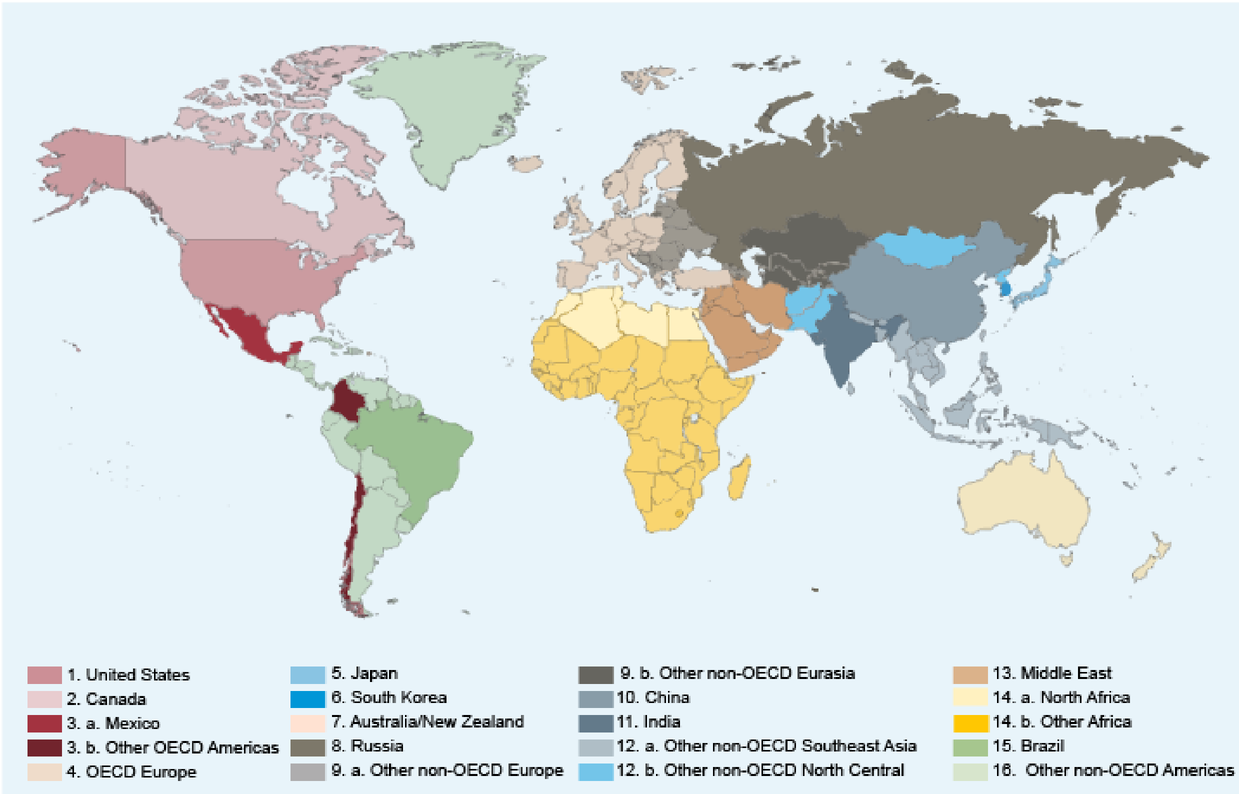
Table 16. WHAM region set definitions

Root set	Index	Subsets	Index
Regions_	reg	SupplyRegion_	s
		RefiningRegion_	rf
		DemandRegion_	d
		OPEC_	opec
		nonOPEC_	nonopec
		OECD_	oecd
		nonOECD_	nonoecd
WEPS_Regions_	reg_weps		
Country_	cntry		
Continents_	cont		

The regional representation of WEPS limits the representation of WHAM because all the data received from WEPS—and all data provided back to other modules—must adhere to these geographic aggregations.

WHAM will use a disaggregated version of WEPS regions (Table 17) for its demand region subset. WHAM needs the disaggregated WEPS regions to effectively represent fundamental aspects of global hydrocarbon activity (Figure 7).

Figure 7. Proposed disaggregation of selected WEPS regions (for WHAM demand region subset)



Note: The legend refers to WEPS regions and does not separate out splits of regions (3), (9), (12), and (14). The four split regions specified share the same color scheme as the legend.

Table 17. WHAM demand region definitions

ID	Name	Example countries
USA	United States of America	United States of America
CAN	Canada	Canada
MXC_MX	Mexico	Mexico
MXC_CC	Colombia and Chile	Colombia, Chile
EUR	OECD Europe	United Kingdom, France, Germany, Italy
JAP	Japan	Japan
ANZ	Australia and New Zealand	Australia, New Zealand, Micronesia
SKO	South Korea	South Korea
RUS	Russia	Russia
URA_EUR	Other non-OECD Europe	Ukraine, Belarus, Romania, Croatia
URA_ASIA	Other non-OECD Europe	Georgia, Azerbaijan, Kazakhstan
CHI	China	China
IND	India	India
OAS_ASIA	Other non-OECD Asia	Pakistan, Mongolia, North Korea
OAS_OCEAN	Other non-OECD Asia	Bangladesh, Thailand, Malaysia, Indonesia
MID	Middle East	Saudi Arabia, Iraq, United Arab Emirates
AFR_N	North Africa	Egypt, Libya, Tunisia, Algeria, Morocco
AFR_S	South Africa	All other Africa not defined above
BRA	Brazil	Brazil
CSA	Other non-OECD Americas	Venezuela, Argentina, Dominican Republic

Solution Methodology

Modeling approach

WHAM's optimization routine should solve, producing feasible results, across a wide range of inputs and assumptions. It should only produce infeasible results if given erroneous or *zero* results from other modules through the restart file.

The solution to WHAM's mathematical program should be a unique solution; alternatively, WHAM should have a procedure in place to handle alternative solutions and modify the mathematical program in order to arrive at a unique, reproducible solution. This is a requirement to ensure that the WEPS system can reach convergence over multiple iterations.

Furthermore, WHAM should strive to maintain an average model runtime shorter than 20 minutes to ensure that a fully integrated WEPS run can be completed overnight.

Mathematical representation of model design

WHAM is a multiple time period, multiple region, linear programming model. We designed it to return solutions that:

- Meet volume requirements as specified by the demand modules
- Satisfy all other specified constraints
- Minimize the total delivered cost for the current year in real dollars

All costs subject to minimization appear in the model's objective function.

Because refineries use one primary input—crude oil—to produce an entire slate of outputs, it becomes ambiguous to use traditional economic marginal costs of production to assign product prices. As such, WHAM will use shadow prices obtained by solving the linear program for these prices. These shadow prices will result in a wholesale price of each refined product for each demand region that can be further adjusted by industry if needed. This methodology lends itself to the application of a cost minimization objective function as opposed to one based on profit maximization.

Index definitions

Element, index	Description
$t \in T$	Time in years
$c \in C$	Commodities
$c \in CD\ (\subset C)$	Commodities to be mapped to a WEPS consumption restart variable
$c \in CREFI\ (\subset C)$	Commodities consumed by refineries for production
$c \in COIL$	Oil commodities
$c \in CREFUTIL$	Utility commodities to be consumed in the refinery process (for example, natural gas, steam, electricity)
$c \in CREFOUTPUT\ (\subset C)$	Commodities produced by a refinery region
$m \in M$	Transportation mode
$m \in PIPE\ (\subset M)$	Pipeline based transportation modes
$m \in SHIP\ (\subset M)$	Ship based transportation modes
$r \in R$ $R = \{N\}$	Regions
$r \in RD\ (\subset R)$	Demand regions to be supplied products
$r \in RS\ (\subset R)$	Supply regions
$r \in RF\ (\subset R)$	Refining regions

$n \in N$	Country
$f \in F$	Refinery type
$o \in O$	Refinery operating mode
$i \in I$	Refinery crude oil slate input
$u \in U$	Refinery unit

Parameters

Parameter	**Definition**	**Data source**
$pcape_{c\,n\,n'}$	Export pipeline capacity of *c* from *n* to *n'*	Research and prior GHySMo
$pcapi_{c\,n\,n'}$	Import pipeline capacity of *c* from *n* to *n'*	Research and prior GHySMo
$pcapri_{c\,r\,r'}$	Total pipeline import capacity from *r* to *r'*	Research and prior GHySMo
$pratei_{c\,n\,n'}$	Import pipeline cost/rate from *n* to *n'*	Research and prior GHySMo
$prateri_{c\,r\,r'}$	Import pipeline cost/rate from *r* to *r*	Derived
$bihaul_{c\,m}$	0-1 parameter; 1 if *c* can be shipped by *m*	Configuration
$bilng_{c\,m}$	0-1 parameter; 1 if *c* is constrained by regasification and liquefaction capacity when transported by *m*	Configuration
$bidom_{r\,r'}$	0-1 parameter; 1 if *r* to *r'* is considered domestic shipping with zero capacity constraints and logistic costs	Configuration
$scap_{m}$	Total shipping capacity of *m* ($m \in SHIP$)	Refinitiv/S&P Platts
$carc_{r\,r'\,c\,m}$	The cost of moving product *c* from *r* to *r'* via method *m*. Time invariant	Derived
$ioil_{f\,c\,c'}$	Total units of *c'* produced by a single unit of *c* in *f*. The yield vector from upstream GHySMo	GRTMPS
$liquecap_{r}$	Natural gas liquefaction capacity for *r*	prior GHySMo, NGMM, International Gas Union
$liquecost_{r}$	Natural gas liquefaction variable cost for *r*	prior GHySMo, NGMM, International Gas Union
$regascap_{r}$	Liquefied natural gas (LNG) regasification capacity for *r*	prior GHySMo, NGMM, International Gas Union
$regascost_{r}$	LNG regasification variable cost for *r*	prior GHySMo, NGMM, International Gas Union

$lease\_loss\_prc_r$	Percentage wet natural gas lost through lease fuel usage in *r*	Research
$liq\_loss\_prc_r$	Percentage of dry natural gas lost in the liquefaction process in *r*	Research
$regas\_loss\_prc_r$	Percentage of dry natural gas lost through the regasification process in *r*	Research
$pipe\_loss\_prc_c$	Percentage of *c* lost through the pipeline arcs.	Research
$plant\_loss\_prc_r$	Percentage of natural gas lost through a natural gas processing plant in *r*.	Research
$ngpl\_prc_r$	The percentage of natural gas plant liquids produced from wet natural gas in *r*.	Research and derived
ship_loss_prc	Multiplier of natural gas lost per day of travel by ship	Research
$temission_{f\,r\,t}$	Tax on emissions by refinery type *f* in *r* and *t* (placeholder parameter for future improvements)	Null, future implementation
$num\_days_{r\,r'}$	The number of days it takes a ship to travel from one region *r* to another *r'*	Refinitiv
$yield_{i\,o\,f\,c}$	The amount of refinery product *c'* created from a single unit of crude oil slate *c*, refinery type *f*, and operating mode *o*	GRTMPS
$slate\_mix_{i\,c}$	The amount of crude oil *c* required in a unit of crude oil slate mix *i*	GRTMPS
$qutility_{i\,o\,f\,c}$	The quantity of utility commodity *c* required to process a unit of crude oil slate *i*, using operating mode *o* and refinery type *f*	GRTMPS
$catchem_{i\,o\,f}$	The dollar cost of required catalysts and other chemicals required to process a unit of crude oil slate *i*, using operating mode *o* and refinery type *f*	GRTMPS
$unit\_capacity_{r\,t\,u}$	The total capacity of *u* in *r* and period *t*	Oil & Gas Journal
$unit\_utilization_{i\,o\,f\,u}$	The amount of *u* capacity utilized when processing a unit of crude oil slate *i* through refinery type *f* in operating mode *o*	GRTMPS

Column definitions

Exogenous

Variable	Description
$PSUP_{c\,r\,t}$	The base price of *c* in *r* and *t* provided the upstream supply curve
$PUTIL_{r\,c\,t}$	The price of utilities *c* used in refining, excluding natural gas in *r* and *t*
$QSUP_{c\,r\,t}$	The quantity of *c* in *r* and *t* provided by the upstream supply curve
$QDEM_{c\,r\,t}$	The quantity of *c* in *r* and *t* demanded by a demand region

Endogenous

Variable	Description	Application
$UTILIZATION_{r\,t\,i\,o\,f}$	The quantity of crude oil slate *i* run through each refinery type and operating mode at a given *r* and *t*.	Used in conjunction with yield vectors from GRTMPS to determine refinery output. Also determined the quantity and type of oil imported, the amount of natural gas imported, and the amount of electricity and steam used in refinery production.
$QR_{c\,r\,t}$	The amount of commodity *c* produced by refinery region *r* in *t*	Sets the amount of refinery products that can leave the refinery region.
$QSUP_{c\,r\,t}$	Quantity of *c* supplied by *r* in *t* (quantity supplied upstream)	Constrains the quantity of crude oil and natural gas available to leave a supply region.
$RC_{r\,t}$	Total regional refinery cost of production for *r* and *t*	Used in the cost minimization objective function.
$QARC_{r\,r'\,t\,c\,m}$	Quantity of *c* exported from *r* to *r'* via method *m*	One of two decision variables. Controls the movement of commodities between regions.
$LEASE\_LOSS_{r\,t}$	Quantity of wet natural gas lost for lease fuel usage in *r* and *t*	Determined by a lease fuel parameter, used to balance natural gas volumes.
$QNETGAS_{r\,t}$	Quantity of natural gas available in *r* and *t*. This variable provides the supply basis for natural gas arc quantities.	Limits the amount of natural gas that can be moved on an arc, is used in refinery processes, and is supplied to demand regions.
$PLANT\_LOSS_{r\,t}$	Quantity of dry natural gas lost in *r* and *t* during the process transforming wet natural gas into dry natural gas.	Determined by a plant loss parameter, used to balance natural gas volumes.

$QNGPL_{r\ t}$	Quantity of natural gas plant liquids produced in *r* and *t*.	Limits the amount of natural gas plant liquids (NGPL) that can be moved from a supply region. NGPL is proportionally produced from the initial natural gas volumes supplied by a region.
$LIQUEFACTION\_LOSS_{r\ t}$	Natural gas loss from the liquefaction processes in *r*.	Determined by a liquefaction loss parameter, used to balance natural gas volumes.
$SHIPPING\_LOSS_{r\ t}$	Natural gas loss from shipping to *r* via ship.	Determined by a shipping loss parameter, used to balance natural gas volumes.
$REGASIFICATION\_LOSS_{r\ t}$	Natural gas loss from the regasification process in *r.*	Determined by a regasification parameter, used to balance natural gas volumes.
$PIPELINE\_LOSS_{r\ t}$	Natural gas loss from piping to *r*.	Determined by a pipeline loss parameter, used to balance natural gas volumes.
$COIL_{c\ r\ t}$	The total cost of oil type *c* used in refinery region *r* in *t.*	This variable tracks the total cost of all types of oil imported into a refinery region. This is used in the refinery cost function.
$CGAS_{r\ t}$	The total cost of dry natural gas used in refinery region *r* in *t*.	This variable tracks the total cost of natural gas imported into a refinery region for production. This is used in the refinery cost function.
$TOTAL\_LOGISTIC\_COST_{t}$	The total cost of all logistic activities in period *t*.	The cost of all arc activity ending at a demand region. Activity from supply to refinery regions is not included because those costs are already reflected in the refinery cost function and would create double counting.
$TOTAL\_REFINERY\_COST_{t}$	The total cost of all refinery activities in period *t*.	Includes the cost of importing natural gas and crude oil, as well as chemical and utility costs, according to GRTMPS.
$TOTAL\_COST_{t}$	The total cost of all refinery and logistic activities in period *t*. The objective function.	The sum of total logistic costs and total refinery costs. This value is minimized in the objective function.

Objective function

Logistic cost

$$TOTAL\_LOGISTIC\_COST_t = \sum_r \sum_{r'} \sum_c \sum_m QARC_{r\,r'\,t\,c\,m} * PARC_{r\,r'\,t\,c\,m} \qquad (r' \in RD)$$

The total cost of logistics in t is the sum of all arc activity ending at a demand region multiplied by the price of using that arc. This total cost includes the initial upstream price of that commodity with various logistics adders. Refinery activity costs are not included in this summation of arc activity because the total cost of refinery activity cannot be distributed clearly among its multiple outputted commodities.

Refinery Cost

$$TOTAL\_REFINERY\_COST_t = \sum_r RC_{rt} \qquad (r \in RF)$$

where

$$\begin{aligned} RC_{r\,t} = & \sum_{c\,\in\,CREFOIL} COIL_{c\,r\,t} + CGAS_{r\,t} \\ & + \sum_i \sum_o \sum_f \sum_c UTILIZATION_{r\,t\,i\,o\,f} * qutility_{i\,o\,f\,c!=natgas} * \boldsymbol{PUTIL_{r\,t\,c}} \\ & + \sum_i \sum_o \sum_f UTILIZATION_{r\,t\,i\,o\,f} * catchem_{i\,o\,f} + \sum_f temission_{f\,r\,t} \end{aligned}$$

$[$*total regional refining cost*$]$ = $[$*oil input costs*$]$ + $[$*natural gas input costs*$]$ + $[$*exogenous utility costs*$]$ + $[$*catalyst and chemical costs*$]$ + $[$*cost of emissions*$]$

The total cost of refining activity is the sum of previously calculated regional refinery costs, which are a function of their total output produced.

and

$$COIL_{c\,r'\,t} = \sum_m \sum_r QARC_{r\,r'\,t\,c\,m} * PARC_{r\,r'\,t\,c\,m} \qquad (r' \in RF, r \in RS, c \in COIL)$$

$$CGAS_{r'\,t} = \sum_m \sum_r QARC_{r\,r'\,t\,c=natgas\,m} * PARC_{r\,r'\,t\,c=natgas\,m} \qquad (r' \in RF, r \in RS)$$

Objective Function

$$\boldsymbol{Minimize}\ TOTAL\_COST_t = TOTAL\_LOGISTIC\_COST_t + TOTAL\_REFINERY\_COST_t$$

Or

$$\begin{aligned}\boldsymbol{Minimize} \sum_r \sum_{r'} \sum_c \sum_m \boldsymbol{QARC_{r\,r'\,t\,c\,m}} * PARC_{r\,r'\,t\,c\,m} \\ + \sum_{c\,\in\,CREFOIL} \sum_m \sum_r \boldsymbol{QARC_{r\,r'\,t\,c\,m}} * PARC_{r\,r'\,t\,c\,m} \\ + \sum_m \sum_r \boldsymbol{QARC_{r\,r'\,t\,c=natgas\,m}} * PARC_{r\,r'\,t\,c=natgas\,m} \\ + \sum_i \sum_o \sum_f \sum_c \boldsymbol{UTILIZATION_{r\,t\,i\,o\,f}} * qutility_{i\,o\,f\,c!=natgas} * PUTIL_{r\,t\,c} \\ + \sum_i \sum_o \sum_f \boldsymbol{UTILIZATION_{r\,t\,i\,o\,f}} * catchem_{i\,o\,f} + \sum_f temission_{f\,r\,t}\end{aligned}$$

Constraints

1. Demand constraint for each commodity (c) in each region (r') in time period (t)

$$\sum_m \sum_r \boldsymbol{QARC_{r\,r'\,t\,c\,m} = QDEM_{c\,r'\,t}} \qquad \boldsymbol{r' \in RD}$$

where

$QARC_{r\,r'\,t\,c\,m}$ = quantity of commodity (c) exported from region (r) to region (r') via method *m*;
$QDEM_{c\,r'\,t}$ = the quantity of commodity (c) in region (r) and time (t) demanded by a demand region. Directly determined by the WEPS restart file.

2. Total supply capacity for all oil products in (c) in an upstream region (r) in time period (t) cannot exceed the total quantity of (c) exported from that region.

$$\sum_m \sum_{r'} \boldsymbol{QARC_{r\,r'\,t\,c\,m} \leq QSUP_{c\,r\,t}} \qquad \boldsymbol{c \in COIL}$$

where

$QARC_{r\,r'\,t\,c\,m}$ = quantity of commodity (c) exported from region (r) to region (r') via method (*m*);
$QSUP_{c\,r\,t}$ = quantity of (c) supplied by (r) in (t).

3. The quantity of commodity (c) transported between region (r) and (r') by pipeline cannot exceed the pipeline capacity between those regions in a given time period (t)

$$\boldsymbol{QARC_{r\,r'\,t\,c\,m} \leq pcapri_{c\,r\,r'}} \qquad \boldsymbol{m \in PIPE}$$

where

$QARC_{r\,r'\,t\,c\,m}$ = quantity of commodity (c) exported from region (r) to region (r') via method *m*;
$pcapri_{c\,r\,r'}$ = total yearly pipeline import capacity from (r) to (r').

4. The quantity of all commodities (c) transported between regions (r) and (r') via ship must not exceed the global shipping capacity of ship type (m)

$$\boldsymbol{\sum_{r}\sum_{r'}\sum_{c}(QARC_{r\,r'\,t\,c\,m} * bihaul_{c\,m}) \leq scap_{m}} \qquad \boldsymbol{m \in SHIP}$$

where

$QARC_{r\,r'\,t\,c\,m}$ = quantity of commodity (c) exported from region (r) to region (r') via method (*m)*;
$bihaul_{c\,m}$ = binary parameter allowing commodity (c) to by transported via method (m);
$scap_{m}$ = total global shipping capacity for ship type (m).

5. The quantity of all LNG commodities (c) transported out of region (r) must not exceed the yearly liquefaction capacity of (r).

$$\boldsymbol{\sum_{r'}\sum_{c}(QARC_{r\,r'\,t\,c\,m} * bilng_{c\,m}) \leq liquecap_{r}} \qquad \boldsymbol{m \in SHIP}$$

where

$QARC_{r\,r'\,t\,c\,m}$ = quantity of commodity (c) exported from region (r) to region (r') via method (*m)*;
$bilng_{c\,m}$ = binary parameter determining if a commodity (c) is constrained by regasification and liquefaction capacities when transported by method (m);
$liquecap_{r}$ = yearly natural gas liquefaction capacity for region (r).

6. The quantity of all LNG commodities (c) transported into region (r') must not exceed the yearly regasification capacity of region (r').

$$\boldsymbol{\sum_{r}\sum_{c}(QARC_{r\,r'\,t\,c\,m} * bilng_{c\,m}) \leq regascap_{r'}} \qquad \boldsymbol{m \in SHIP}$$

where

$QARC_{r\,r'\,t\,c\,m}$ = quantity of commodity (c) exported from region (r) to region (r') via method (*m)*;
$bilng_{c\,m}$ = binary parameter determining if a commodity (c) is constrained by regasification and liquefaction capacities when transported by method (m);
$regascap_{r'}$ = yearly natural gas regasification capacity for region (r').

7. Total quantity of natural gas exported from a supply region (r) to any region (r') cannot exceed the base supply of natural gas in region (r) and time period (t)

$$\sum_{m}\sum_{r'} \boldsymbol{QARC}_{r\,r'\,t\,c=natgas\;m} \leq \boldsymbol{QNETGAS}_{r\,t} \qquad\qquad \boldsymbol{r} \in \boldsymbol{RS}$$

where

$QARC_{r\,r'\,t\,c\,m}$ = quantity of commodity (c) exported from region (r) to region (r') via method (*m)*;
$QNETGAS_{r\,t}$ = the quantity of natural gas available in region (r) in time period (t). This quantity is the available quantity from initial supply data, net losses from lease fuel and natural gas plant processing.

8. Total quantity of NGPL exported from a supply region (r) to any region (r') cannot exceed the base supply of NGPL in region (r) and time period (t).

$$\sum_{m}\sum_{r'} \boldsymbol{QARC}_{r\,r'\,t\,c=ngpl\;m} \leq \boldsymbol{QNGPL}_{r\,t} \qquad\qquad \boldsymbol{r} \in \boldsymbol{RS}$$

where

$QARC_{r\,r'\,t\,c\,m}$ = quantity of commodity (c) exported from region (r) to region (r') via method (*m)*;
$QNGPL_{r\,t}$ = the quantity of NGPL available in region (r) in time period (t).

9. Total unit capacity in time period (t) for each refinery unit (u) and region (r) must be greater than or equal the total unit utilization used from running all volumes of crude oil slate (i) through each refinery type (f) operating mode (o).

$$\sum_{i}\sum_{o}\sum_{f} \boldsymbol{UTILIZATION}_{r\,t\,i\,o\,f} * \boldsymbol{unit\_utilization}_{i\,o\,f\,u} \leq \boldsymbol{unit\_capacity}_{r\,t\,u}$$

where

$UTILIZATION_{r\,t\,i\,o\,f}$ = the quantity of crude slate (i) run through each refinery type (f) and operating mode (o) at a given refinery region (r) in time (t);
$unit\_utilization_{i\,o\,f\,u}$ = the amount of unit (u) capacity utilized when processing a unit of crude slate (i) through refinery type (f) in operating mode (o);
$unit\_capacity_{r\,t\,u}$ = the total capacity of unit (u) in refinery region (r) and period (t).

Integration wrapper

WHAM requires an integration wrapper to communicate with the greater WEPS system. We will write an integration wrapper in Python, which will handle the following tasks:

- Read in the WEPS working directory and user configuration files
- Read data from the Restart file and pass to WHAM
- Initialize the WHAM code in AIMMS while handling any errors
- Format results from AIMMS and write to the Restart file
- End AIMMS session after a successful run
- Store debugging data

The Python integration wrapper is based on the integration code found in GHySMo; however, it is significantly simplified due to the smaller scope of WHAM. The Python code resides with the AIMMS code in the WEPS repository.

Benchmarking and calibration

Our short- and long-term forecasts need to be consistent, as do our international and domestic energy market projections. We have applied the following rules to projections in the IEO:

- STEO forecasts define EIA's short-term (1–2 year) projections, where available
- AEO projections define EIA's projection of U.S. energy markets, including U.S. energy trade

Furthermore, given that WHAM is required to balance supply and demand, this benchmarking must be endogenous, not be a requirement satisfied by overwriting volumes, and must be able to be easily disabled to allow the model to solve independently from STEO and AEO projections. The benchmarking requirements:

- Use additive and multiplicative factors to benchmark (in other words, not overwrite) in order to track discrepancy and maintain supply and demand balance
- Constrain U.S. model results to match AEO projections
- Provide production that matches STEO total liquid fuels by region
- Benchmark to STEO crude oil and other liquids production forecasts by international country or region as allowed, given inconsistent regional representation and commodity definitions

Calibration is an iterative procedure and involves:

- Establishing model inputs that correspond to reported inputs in each region of interest
- Solving the model with those inputs
- Comparing the model outputs to reported refinery outputs in each region of interest
- Adjusting model coefficients as needed
- Repeating as needed until model outputs match reported outputs within a given degree of precision

Uncertainties and Limitations

Modeling a global network of petroleum logistics and processing carries with it a number of challenges that translate to uncertainties and limitations in the model. WHAM seeks to address as many of these as possible, but some uncertainty is unavoidable in a model of this scale. As such, we made a number of design choices to minimize the impacts of these uncertainties:

- WHAM refineries are aggregated based on regional and geographical location, and they are allowed to operate based on the predominant conversion technology present in these aggregate refineries.
- Interregional transport of liquid fuels and gases is based on existing and planned aggregate transport capacities between regions.
- Upstream production of crude oil and natural gas is not rigorously modeled, but instead, the model will allocate regional upstream production based on logistic cost minimization and will be bound by third-party breakeven price data.

Because global natural gas and liquid fuels markets continue to evolve, WHAM includes a framework that facilitates future development. At this time, WHAM does not address issues such as biofuels, infrastructure investment, or climate policy; however, the model is designed in such a way to facilitate implementation of these modeling concepts in the future.

Testing Plan

WHAM will undergo rigorous testing before we include it in IEO2023. This testing will take place in both standalone (WHAM only) and integrated (WEPS) environments, and will involve a number of separate phases, including, but not limited to:

- Testing WEPS interaction—ensuring that WHAM can communicate and interface appropriately with WEPS, sending and receiving variables and data as needed
- Testing WHAM upstream implementation—ensuring that the implementation of crude oil and natural gas production in WHAM operates correctly and as expected
- Testing price and quantity validity—ensuring that WHAM outputs of prices and quantities for crude oil, natural gas, and finished petroleum products are logical and consistent with historical values
- Testing regionality—ensuring that WHAM's distribution of refining and logistics activity is logical and consistent with historical values

In addition to the dedicated testing phase of the project which occurs in the final three months of development, we will perform ad hoc testing throughout the development process.

Conclusion

The primary purpose for developing the World Hydrocarbon Activity Model (WHAM) is to provide an alternative modeling solution for representing global petroleum processing and logistics in the WEPS environment. We will design this modeling solution to use a single linear program model to provide robust and reliable representations of petroleum production, refinery operations, and interregional transport of petroleum and petroleum products. We will design WHAM to implement this solution by incorporating:

- A single linear program designed to robustly model both global refinery operations and global petroleum logistics
- An estimation method for upstream production of crude oil and natural gas based on regional production volumes and cost of production

WHAM will focus on building a straightforward but flexible model to facilitate future improvements without having to create a new model from scratch. This flexibility will allow the WEPS team to more efficiently tackle future challenges such as rigorous upstream modeling and decarbonization scenarios. Because documentation will be critical to the success of future improvements to WHAM, we will document all critical pieces of the model so that current and future modelers can use WHAM with ease.

Appendix A: Background

Current approach

After IEO2019, we undertook a study to analyze GHySMo's performance and to identify potential shortcoming of the model. As a result of this study, we launched Phase 1 of the GHySMo Overhaul Project. We intended this project to rectify a number of issues and prepare GHySMo for IEO2021. Phase 1 successfully addressed the primary objectives that we identified; during the testing and validation phase, however, the GHySMo team encountered a number of new issues that could not be resolved. As a result, we developed and used the Oil and Gas Tool, a deterministic simulator, in IEO2021.

The primary issue we observed during the testing and validation stage of Phase 1 was price and quantity behavior that didn't adequately represent petroleum markets as GHySMo performed its internal iterations. In its current implementation, GHySMo iterates between the Logistics and Conversion modules in series to meet a demand slate provided by the GHySMo Integration module (and WEPS). The Logistics and Conversion modules have different objective functions, but pass model results back and forth (in other words, the outputs of Logistics become the inputs of Conversion). We believe this strategy of iterating between two competing linear programs is the root cause of the issues. This convergence issue led to our current emphasis on redesigning WHAM to create a better logistics and conversion representation instead of a focus on upstream modeling.

For example, we observed flawed behavior in the pricing of E0, or clean gasoline (without ethanol). From one iteration to the next, the price of E0 would decrease as the quantity produced by the Conversion model increased. This result was caused by the Logistics model "seeing" a lower E0 price and subsequently demanding a larger volume of E0 to fill the overall transportation demand. As a result, E0 prices drop to near zero over the internal iteration cycle, and the model results as a whole are compromised.

Furthermore, the GHySMo codebase does not follow best practices, making it difficult to read and track updates made over time. The presence of deprecated code and a lack of applicable documentation resulted in additional difficulty as we attempted to troubleshoot the issues discovered during Phase 1.

Previous approaches

Prior to the development of WHAM, we used a number of other approaches to accurately project the behaviors of the global upstream production and refining industries (Table A).

Table A. Previous global petroleum supply and refining modeling approaches

Modeling approach	IEOs used	Description
Global World Oil Database (GWOB)	2010–2017	Accounting tool that determines global production of crude oil and liquid fuels by region to meet demand
Refinery Model	2010–2017	Projects petroleum product prices based on three key refinery regions based on pricing algorithms and the calculation of market balance (in Excel)
International Natural Gas Model (INGM)/(Reduced Form) Natural Gas Model	2010–2017	Linear program written in GAMS to solve for natural gas production and natural gas prices in response to changes in consumption; reduced form of INGM in WEPS
World Oil Resource Model (WORM)	NA	Stochastic resource-based accounting spreadsheet model designed to be responsive to any given price scenario, while maintaining integrity of known global resource volumes
PAIN	2015	AIMMS linear program model designed to optimize for plausibility of upstream production using an implausibility tax and IHS risk factor coefficient to determine production profile by country
GHySMo	2019	Four-module AIMMS model designed for rigorous accounting of upstream production and refinery operations. Used two linear programs in series to perform optimization calculations
Oil and Gas Tool	2021	Accounting tool designed as a stopgap replacement for GHySMo. Uses analyst judgment to define production and growth. Does not model trade.

Note: NA=not available.

Appendix B. WHAM Hierarchy of Requirements

The matrix of requirements below will be the basis for requirements tracking and traceability during model design and development.

Table B. WHAM hierarchy of requirements

ID	Requirement
1	**Integration**
1.1	WEPS
1.1.1	WHAM-WEPS interface
1.1.2	WEPS runtime parameters
1.1.3	Restart file interface
1.1.3.1	Input variables
1.1.3.2	Output variables
1.1.3.3	Completion of output data received
1.1.4	Debug .csv files
1.1.5	Input file formatting
1.1.6	IES data
1.1.7	IEA data
1.1.8	STEO data
1.1.9	AEO data
1.1.10	No database use
1.2	WHAM Repository
1.2.1	Data processing files
1.2.2	Transfer to WEPS Repository
1.3	GRTMPS
2	**Functional**
2.1	Supply
2.1.1	Crude oil and condensate production
2.1.2	Dry natural gas production
2.1.3	Other liquid fuels production
2.1.3.1	NGPL supply
2.1.3.2	Biofuel supply
2.1.3.3	Refinery gain
2.1.3.4	Coal-to-liquids supply
2.1.3.5	Gas-to-liquids supply
2.1.3.6	Other liquid fuels
2.2	Refining
2.2.1	Petroleum liquids supply equals WEPS demand
2.2.2	Mass and energy balance
2.2.3	Refinery gate prices
2.2.4	Account for differences in refinery operations, equipment, and crude oil feeds
2.2.5	Refinery consumption

Appendix C. Selected WEPS Restart File Variables in WHAM

Table C. Selected WEPS restart file variables in WHAM

Restart variable	Units	Dimension	Input or output
wptpprc	$/b	Year	Both
wptprc	$/MMBtu	Year	Both
sec_qhngin	TBtu	Industry x Region x Year	Input
cnvfacep	MMb/d, Tcf	Region x Year	Input
liqconvforsteo	TBtu per MMb/d	Region x Year	Input
qcddh	TBtu	Region x Year	Input
qcdin	TBtu	Region x Year	Input
qcdpg	TBtu	Region x Year	Input
qcdtr	TBtu	Region x Year	Input
qcdun	TBtu	Region x Year	Input
qdscm	TBtu	Region x Year	Input
qdsdh	TBtu	Region x Year	Input
qdsin	TBtu	Region x Year	Input
qdspg	TBtu	Region x Year	Input
qdsrs	TBtu	Region x Year	Input
qdstr	TBtu	Region x Year	Input
qdsun	TBtu	Region x Year	Input
qettr	TBtu	Region x Year	Input
qetun	TBtu	Region x Year	Input
qjftr	TBtu	Region x Year	Input
qjfun	TBtu	Region x Year	Input
qkscm	TBtu	Region x Year	Input
qksin	TBtu	Region x Year	Input
qksrs	TBtu	Region x Year	Input
qksun	TBtu	Region x Year	Input
qlgcm	TBtu	Region x Year	Input
qlgin	TBtu	Region x Year	Input
qlgrs	TBtu	Region x Year	Input
qlgtr	TBtu	Region x Year	Input
qlgun	TBtu	Region x Year	Input
qmgcm	TBtu	Region x Year	Input
qmgin	TBtu	Region x Year	Input
qmgtr	TBtu	Region x Year	Input
qmgun	TBtu	Region x Year	Input
qngcm	TBtu	Region x Year	Input
qngdh	TBtu	Region x Year	Input
qngin	TBtu	Region x Year	Input
qngpg	TBtu	Region x Year	Input

Restart variable	Units	Dimension	Input or output
qngrs	TBtu	Region x Year	Input
qngtr	TBtu	Region x Year	Input
qngun	TBtu	Region x Year	Input
qobtr	TBtu	Region x Year	Input
qobun	TBtu	Region x Year	Input
qopin	TBtu	Region x Year	Input
qoptr	TBtu	Region x Year	Input
qopun	TBtu	Region x Year	Input
qpcin	TBtu	Region x Year	Input
qpcun	TBtu	Region x Year	Input
qrscm	TBtu	Region x Year	Input
qrsdh	TBtu	Region x Year	Input
qrsin	TBtu	Region x Year	Input
qrspg	TBtu	Region x Year	Input
qrstr	TBtu	Region x Year	Input
qrsun	TBtu	Region x Year	Input
qspin	TBtu	Region x Year	Input
qsptr	TBtu	Region x Year	Input
qspun	TBtu	Region x Year	Input
qsteo	TBtu	Region x Year	Input
suppet	TBtu	Region x Year	Input
conngtcf	Tcf	Year	Input
qclrf_subtype	TBtu	CoalType x Region x Year	Output
netimpngas_2021	Tcf	Region x Year	Output
pcngtr	$/MMBtu	Region x Year	Output
pdscm	$/MMBtu	Region x Year	Output
pdsdh	$/MMBtu	Region x Year	Output
pdsin	$/MMBtu	Region x Year	Output
pdspg	$/MMBtu	Region x Year	Output
pdsrs	$/MMBtu	Region x Year	Output
pdstr	$/MMBtu	Region x Year	Output
pettr	$/MMBtu	Region x Year	Output
pjftr	$/MMBtu	Region x Year	Output
pkscm	$/MMBtu	Region x Year	Output
pksin	$/MMBtu	Region x Year	Output
pksrs	$/MMBtu	Region x Year	Output
plgcm	$/MMBtu	Region x Year	Output
plgin	$/MMBtu	Region x Year	Output
plgrs	$/MMBtu	Region x Year	Output
plgtr	$/MMBtu	Region x Year	Output
plngtr	$/MMBtu	Region x Year	Output

Restart variable	Units	Dimension	Input or output
pmgcm	$/MMBtu	Region x Year	Output
pmgin	$/MMBtu	Region x Year	Output
pmgtr	$/MMBtu	Region x Year	Output
pngcm	$/MMBtu	Region x Year	Output
pngdh	$/MMBtu	Region x Year	Output
pngin	$/MMBtu	Region x Year	Output
pngpg	$/MMBtu	Region x Year	Output
pngrs	$/MMBtu	Region x Year	Output
pobtr	$/MMBtu	Region x Year	Output
popin	$/MMBtu	Region x Year	Output
poptr	$/MMBtu	Region x Year	Output
prscm	$/MMBtu	Region x Year	Output
prsdh	$/MMBtu	Region x Year	Output
prsin	$/MMBtu	Region x Year	Output
prspg	$/MMBtu	Region x Year	Output
prstr	$/MMBtu	Region x Year	Output
qbmrf	TBtu	Region x Year	Output
qctlprod	TBtu	Region x Year	Output
qdsrf	TBtu	Region x Year	Output
qelrf	TBtu	Region x Year	Output
qgtlprod	TBtu	Region x Year	Output
qhtrf	TBtu	Region x Year	Output
qksrf	TBtu	Region x Year	Output
qlgrf	TBtu	Region x Year	Output
qmgrf	TBtu	Region x Year	Output
qnglgin	TBtu	Region x Year	Output
qnglpin	TBtu	Region x Year	Output
qngpptr	TBtu	Region x Year	Output
qngrf	TBtu	Region x Year	Output
qoprf	TBtu	Region x Year	Output
qpcrf	TBtu	Region x Year	Output
qrsrf	TBtu	Region x Year	Output
qsprf	TBtu	Region x Year	Output
qwsrf	TBtu	Region x Year	Output
supngas	TBtu	Region x Year	Output
supngas_tcf	Tcf	Region x Year	Output
totliqkbpd	Mb/d	Region x Year	Output
totnatgastcf	Tcf	Region x Year	Output
svregionalproduction	MMb/d, Tcf	SupplyType x Region x Year	Output
eurngprc	$/MMBtu	Year	Output
hhngprc	$/MMBtu	Year	Output

Restart variable	Units	Dimension	Input or output
jpnngprc	$/MMBtu	Year	Output
svbiofuel	MMb/d	Year	Output
svlpg	MMb/d	Year	Output
svother	MMb/d	Year	Output
svrefinerygain	MMb/d	Year	Output
wptreq	TBtu	Year	Output
wtipprc	$/b	Year	Output
qclrf_subtype	TBtu	CoalType x Region x Year	Unused
svregionalexports	MMb/d, Tcf	Commodity x Region x Year	Unused
svnetgasexports_pipe	Tcf	ContractMode x Region x Year	Unused
svnetgasexports_ship	Tcf	ContractMode x Region x Year	Unused
pngwd	$/MMBtu	Region x Year	Unused
qclrf	TBtu	Region x Year	Unused
sngls	TBtu	Region x Year	Unused
svrefineryproduction	MMb/d	Region x Year	Unused

Notes: Restart variable names for consumption (q*) and price (p*) correspond to the fuel type and end-use sector.

Unused variables may or may not be used by WHAM (To be determined).

$/b=dollars per barrel, $/MMBtu=dollars per million British thermal units, TBtu=trillion British thermal units, MMb/d=million barrels per day, Tcf=trillion cubic feet