Independent Statistics & Analysis
U.S. Energy Information Administration

# Global Hydrocarbon Supply Model (GHySMo) Requirements

May 2014

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 United States Government. The views in this report therefore should not be construed as representing those of the U.S. Department of Energy or other federal agencies.

# Introduction and Background

Technological advancements in the development of shale oil and gas have had a profound impact on the outlook for oil and gas markets in the United States and have broader implications for world markets as well. This relatively new development has already required the U.S. Energy Information Administration (EIA) to make significant changes to its domestic supply model. However, the broader implications have yet to be adequately addressed within EIA's existing models, and this has led EIA to reconsider its approach to projecting international natural gas and liquids markets in the midstream and upstream. Some of these implications include

- the potential for and economic impacts of US exports of liquid fuels and natural gas
- the application of technologies and practices developed in the U.S. to the production of international shale resources
- the interrelationship of international crude oil and natural gas production
- the potential evolution to a more competitive international gas market
- The degree to which noneconomic factors, such as geopolitical instability and risk, national policies, and strategic and other non-competitive behaviors govern pricing

The Office of Petroleum, Natural Gas, and Biofuels Analysis (PNGBA) of the EIA has been tasked to develop a dynamic representation (referred to here as GHySMo[1]) of the global production, processing, transport, distribution, and storage of natural gas and liquid fuels. The ultimate purpose of this project is to improve EIA's capability to represent international markets (i.e., prices and commodity flows) for liquids and natural gas under a variety of assumptions. The primary function of the model will be to replace the existing upstream and midstream models of petroleum and natural gas within the World Energy Projection System Plus (WEPS+)[2]. WEPS+ has component modules for other areas of the energy system, particularly demand modules and a macroeconomic module that will provide regional, sector-specific annual consumption for natural gas and liquids products and that will be provided corresponding wholesale and/or delivered product prices by GHySMo. GHySMo will need to also produce supply prices and quantities and trade volumes and prices for publication in the IEO. For a more complete description of GHySMo operation within the WEPS+ modeling system, see Appendix I at the end of this document.

While operating within WEPS+ is GHySMo's primary function, a secondary and still very important function of this development project is to identify a reasonably seamless process, based on GHySMo or its results that will allow for a consistent international representation of the gas and liquids markets to be incorporated within EIA's National Energy Modeling System (NEMS)[3]. This will ensure a domestic/international response in NEMS that is consistent with WEPS+ (i.e., that the models produce similar results in the areas where their coverage overlaps) and may require the incorporation of a

[1] The Global HYdrocarbon Supply MOdel, which is generally assumed to be a computational model throughout this document, but which is understood to include any representation of the global liquids and gas markets that meets the requirements set forth in this document.

[2] http://www.eia.gov/forecasts/ieo/models.cfm WEPS+ is a modeling system used to generate the projections published in EIA's International Energy Outlook (IEO).

[3] http://www.eia.gov/oiaf/aeo/overview/

portion of NEMS, or a reduced form of NEMS, in GHySMo, and vice versa. For a short discussion of some of the issues involved with incorporating GHySMo into NEMS, see Appendix II.

Another secondary function of GHySMo is to operate in a standalone fashion. Without the additional overhead of an integrating framework, it is hoped that a standalone version of the model will run dramatically faster. Not only will this improve testing and development turnaround times, but it should enable the targeted use of greater levels of detail to support certain topical analyses that may not critically depend on dynamic feedback from outside the liquids and gas markets. It is envisioned that a standalone GHySMo would be used to perform such analyses as deep-dive analyses of specific countries or regions of the world, changes to specific national policies or pricing strategies, the impact of new fuel standards or formulations, what-if scenarios involving specific pieces of infrastructure, and other topical analyses as the need arises.

EIA recognizes that the development of GHySMo is an ambitious project. As a multiple-purpose analytic platform, GHySMo must retain flexibility in inputs, operational modes, and outputs. This will require a sophisticated understanding and management of data, analyst judgment, and other forms of knowledge (such as what is and is not known, and levels of certainty of the input data). EIA strongly suggests that the model and the data feeding the model be managed separately as much as possible, so that as data sources change, the model logic and code can remain minimally affected. The model should impose requirements on the knowledge management (KM) system, and the KM system should seek to satisfy those requirements as best as possible given the available data. For a more thorough description of the envisioned GHySMo KM system, see Appendix III.

The requirements laid out in the rest of this document focus on those that are needed to satisfy both its primary and secondary functions.

# GHySMo Requirements

## General Requirements

### *Configurable Flexibility*

To satisfy its primary and secondary functions, GHySMo will need to be extremely flexible. On a computational level, it must integrate seamlessly within multiple modeling environments, converging reliably and rapidly in each. From a data perspective, it must be able to incorporate a vast, heterogeneous, and sometimes sparse catalog of data inputs to form an internally-consistent accounting of the liquids and gas markets around the world.

In general, an energy model can be characterized by its scope (i.e., breadth) and granularity along the four dimensions of 1) geography, 2) time, 3) product type, and 4) supply chain (or sector). The scope of GHySMo is 1) global; 2) same projection period as the AEO and IEO (currently 2040); 3) natural gas and liquid fuels; and 4) production (upstream), through processing/refining, down to but not including the determination of end-use demand, including transportation of upstream, midstream, and downstream products between regions. To satisfy its primary function of operating within the WEPS+ environment, the required levels of granularity are 1) 16 WEPS+ regions; 2) annual projections; 3) gas, a single crude-like upstream product, and 13 mid- and downstream products; and 4) unconstrained by the WEPS+ system.

While WEPS+ reports at the level of granularity specified above, GHySMo may need to operate internally at finer levels of granularity in order to generate robust results that reflect actual market behaviors and incorporate the full availability of data. For instance, in order to capture the impact of intra-regional trade, GHySMo may need to represent individual countries, or even sub-country regions in some cases (e.g., U.S., Russia, China), aggregating the results to the regional level for reporting purposes. Similarly, if seasonal factors are important drivers of refinery operation, GHySMo may need to operate at a seasonal level in its representation of refineries, only aggregating to the annual level when communicating with other modules or for reporting. EIA understands that this internal operational level of granularity must be at least as fine at the reporting level of granularity, and that the internal levels of granularity drive the input data requirements of the model. Thus, in this document, EIA will recognize three distinct levels of granularity (along each of the four dimensions mentioned above): reporting level, operational level, and data level.

To ensure the model retains as much flexibility as is practical, EIA would like the GHySMo model to be able to operate at different levels of granularity along each of these four dimensions, depending on data availability and user input (e.g., run-time switches), insofar as this is practical. For instance, to perform an analysis of markets within the Middle East, the user may wish to run GHySMo at the country level within the Middle East, but maintain the WEPS+ regional structure for the rest of the world. This requirement may impose additional considerations for the data that feeds the model. While the minimum levels of granularity for GHySMo are the WEPS+ levels described above, EIA would strongly prefer the possibility of finer levels of granularity as described next.

WEPS+ operates and reports at a geographic granularity of 16 world regions. While some of these regions are individual countries, most regions are groups of countries. This structure allows for different aggregations of regions to support analyses of OECD and non-OECD countries, differences between continents, and separate accounting of the U.S. with a minimal number of global regions to keep computational and data requirements minimal. However, this level of aggregation does not permit the evaluation of many policy questions involving individual countries, it does not easily allow for changes in the membership of the OECD, and it is largely insensitive to the differing policy environments within different countries. To support more detailed stand-alone analyses of liquid fuels and natural gas markets, and in anticipation of future modeling enhancements of the WEPS+ modeling system, EIA would like GHySMo to support operation at a country level of geographic granularity for most countries of the world. Smaller, geographically proximate countries with similar behaviors may be aggregated for ease of operation, and other country groupings may be proposed where country-level data is unavailable or unreliable, but GHySMo should be structured such that changing these groupings as EIA's analytic needs change and/or as data becomes available should be trivial. In addition, some countries may need a sub-country regional representation to capture the dynamics of the gas and liquid fuels in those countries. EIA would highly encourage a regional representation within the U.S., China, Russia, and Canada; and a GHySMo requirement is that any country may be subdivided into smaller regions to support ad hoc analyses with minimal change to the modeling structure.

While WEPS+ is an annual model, the natural gas market, particularly, has important seasonal drivers that should be considered that will ultimately drive the annual numbers. The International Natural Gas Model INGM breaks the year into winter, summer, and shoulder periods for the purpose of better capturing market dynamics. This should be considered for GHySMo as well. EIA also wishes to build in the flexibility of running the model for user specified time periods, such as monthly or multi-annually (e.g. every five years).

WEPS+ accounts for natural gas, a single representative crude oil stream, and 13 intermediate and final liquid fuels, while the international representation of the liquids markets in NEMS contains 11 different crude specifications, and 20 intermediate and final liquid fuel products. GHySMo must at least operate with the WEPS+ product types, but consideration should be given to how changes to this crude and product slate can easily be incorporated within the model structure. At a minimum, GHySMo should be able to aggregate NEMS-like product information into the WEPS+ categories. Ideally, it should be able to incorporate a larger number of crudes, intermediate and final products, as long as the data are available, with minimal model reconfiguration.

EIA understands that maintaining configurable flexibility imposes requirements on both the data that must be collected and maintained, and also on the knowledge management system that must transform this data to meet the changing input requirements of the GHySMo model. This is especially true for the transportation and logistics functions of GHySMo. Given the scarcity of international data, some data elements will need to be approximated. One approach is to use a known value as a proxy when there are reasons to believe that the two elements are similar. This approach might also require the application of regional scale factors, such as to capture impact of regional labor rates on construction costs. Regardless, if data are not directly available at a reasonable cost to support a proposed modeling approach, the design should also include a proposal for a method for approximating the necessary data.

The model design should include data requirements for maintaining the required configurable flexibility, as well as rules both for aggregating data up to higher levels of granularity and for allocating aggregate data down to constituent components when data at the required level of granularity is unavailable.

### Sectoral Submodules

While the level of detail with which the supply chain is modeled is unconstrained by the WEPS+ environment, EIA would like GHySMo to be broken into submodules to facilitate testing, maintenance, and model administration. As a starting point, EIA suggests submodules for 1) upstream oil and gas production operations including natural gas processing; 2) logistics for transportation (and associated processes such as storage and liquefaction) of primarily natural gas, crude oil, and petroleum products; and 3) transformational processes such as in refineries, and biofuels production facilities. EIA is open to alternate submodules, though, if other groupings would better achieve other model requirements, such as computational speed or convergence. One alternate breakout might be 1) upstream production of gas and liquids, 2) liquids markets (with logistics included), and 3) gas markets (with logistics included). Another alternate submodule grouping might be to separate upstream and downstream logistics modules, potentially including downstream logistics within the downstream (transformation) submodule and the upstream logistics within the upstream (production) submodule.

In theory, the levels of granularity (e.g., regional representations) within each of the GHySMo submodules need not be the same, but they will need to be compatible with one another to facilitate the passing of information among the submodules and the passing of information with WEPS+. When GHySMo acts as a module within WEPS+, a controlling GHySMo routine will be called by WEPS+ to solve for all forecast years and will in turn call the primary submodules. While GHySMo must solve for all years before returning execution to WEPS+, this does not preclude the GHySMo submodules from progressively and iteratively solving one year at a time, reaching internal convergence before proceeding to the next year, if operationally advantageous. The selection of regions should consider impact on quality of projections, data availability, methods for breaking out and aggregating back, and other considerations such as convergence. Since future developments might necessitate a change in the regional breakout, the model should be architected to allow for relative ease of changing the regional structure in the future. Similarly, such things as the breakout of crude types should be considered and selected, but the model should accommodate potential future modifications. EIA’s initial assumption is that the refinery processes will be represented at the WEPS+ regions. On the natural gas side, the more detailed regional representation that has been developed over the years for INGM should provide a good starting point.

### Pricing

The role of GHySMo within the WEPS+ environment is to satisfy regional end-use demands for natural gas and liquid fuels at realistic market-clearing prices across the entire supply chain and in all regions. This requires solving for commodity pricing for upstream and midstream products that is sufficient to drive production to meet the needs of the entire supply chain.

WEPS+ currently represents wholesale prices to the demand modules in each global region, and the initial requirement for GHySMo is to support the same types of prices. Eventually, WEPS+ should project demand based on end-use sector- and region-specific retail prices, which in some countries can be largely influenced by government controls (e.g., subsidized prices). Not only is there an issue about how subsidized prices might change in the future, but there are implications on the regional structure if a subset of countries within a region subsidize and others do not, what a regional price means in such a case, as well as how subsidized pricing effectively results in non-price-responsive demand. GHySMo should be designed to support this transition to retail pricing in WEPS+, which may include providing these prices to the demand modules. Therefore, mechanisms should be built into the architecture of GHySMo for representing end-use prices by sector, even if these mechanisms are initially unused within WEPS+.

### *Costs/Technology*

Assumptions about costs are an important driver in any energy model: capital costs such as those to drill a well, build a liquefaction facility or LNG tanker, or add pipeline or refinery capacity, as well as operations costs for these types of facilities. Costs related to adding capacity are important for making the build decision, but also play a role in the rates that will be charged once in operation. Depending on the overall algorithm selected and the decision being made (i.e., build decision versus operating decision), costs could be specified by investment dollars (requiring additional assumptions, such as exchange rates, rates of return, and recovery period) or as a dollar per unit rate (where such things as rates of return are often assumed inherently). Ideally, GHySMo should allow the user to assume such things as rates of return and endogenously set per unit costs, if practical.

For build decisions, an understanding of the complete cost picture of a project is generally required. However once a facility is built, the operating costs can play a greater role in the decision about how much the facility is actually used as some of the costs are effectively “sunk”. A good example of this is the Sabine Pass liquefaction facility, where users of the facility have purchased the right to use the facility regardless of whether they actually do so. This fee is sunk, and while important in a buyer’s decision to sign a contract, it presumably will not impact their economic decision to use the facility. However, the liquefaction fee is likely included in the charge to the end-use customers. Such distinctions should be included in GHySMo as deemed significant.

Costs and operating efficiencies change across time for a variety of reasons, such as improvements in technology, competition for factor inputs (e.g., high rig utilizations within a region increase drilling costs), externalities such as geopolitical risks, or economic factors (e.g., exchange rates or increased labor rates). To the degree such factors are significant, GHySMo should capture these effects.

For a long-term model such as GHySMo, assumptions about technological advancement and penetration can make a significant difference in the results. While scenarios can be run to test the implications of such assumptions, the model should be built to allow for the incorporation (and ideally projections) of technological advancement and penetration. This is likely to be of greater importance in an upstream submodule and perhaps a biofuels module where the technology is relatively new.

### *Transportation*

At a minimum the logistics submodule will need to set rates for the transportation of hydrocarbons, including natural gas, crude oil, petroleum products, and other related liquids by type (as represented elsewhere in the model) as well as by mode of transport, between the producing region and the end-use region. For petroleum, this includes both the transportation of crude oil from the producing region to the refinery region and for liquid products between the refinery region and end-use region, as appropriate. While the mode of transport should ultimately be readily changeable in the model code, initially a reasonable list of transport modes should be included that reflect the primary modes used in the real world and a level of disaggregation (e.g., ship sizes, "dirty" versus "clean" ships, Jones Act vessels) that provides a reasonable representation of key market factors. Any physical limitations of transport links should be identified (e.g., pipelines, shipping lanes) and thereby be used to limit flows and/or set transportation rates.

If there are multiple possible routes for a given transportation mode to get from one location to another (e.g., by tanker through a canal or around a cape), each option should be represented. In some cases minimum flows might be necessary due to contracted supply or other constraints. Any geopolitical limitations on trade (e.g., embargoes, sanctions, tariffs) should be similarly identified and represented by model constraints or added/subtracted costs, as appropriate. While some routes and infrastructure chosen for flows might initially be based largely on historical relationships and/or geopolitical factors, the model should allow for initial assumptions to change over time to become less of a factor and presumably be based more on economic factors, either endogenously or through explicit analyst specification.

### *Capacity Expansion*

Capacity expansion decisions can involve the entire oil and gas supply chain; thus, modeling these decisions likely requires interactions between GHySMo submodules. GHySMo should accommodate the setting of capacity expansions exogenously, primarily for projects under construction or likely to go forward in the near term. However, each of the submodules will need some mechanism for determining additional (i.e., unplanned) capacity expansions endogenously as well. It is anticipated that the model would use adaptive expectations or rational expectations but not perfect foresight for any capacity planning. Whichever method is chosen in the design, capacity expansion choices in the model should be consistent with the way producers, investors, and consumers in the market appear to be making their decisions.

One approach is to apply a similar structure across all GHySMo submodules, such as the one used in the NEMS Liquid Fuel Market Module (LFMM) where the same model structure is used to assess needs in future years for the purpose of setting expansion requirements, in which case the whole GHySMo could be run for the capacity planning years. For marginal increases in existing capacity another approach is to assume something like a hurdle rate, on top of the rate charged for existing capacity, to apply when capacity exceeds current levels, such as is used in pipeline capacity expansion in the NGTDM. This approach allows capacity decisions to be made simultaneously when equilibrating the model, but does not account for expected future prices or demands (which is not necessarily an issue if demand and pricing does not fall over the projection period). The major drawback to such an approach can be an inability to represent expansions in new areas and large discrete expansions (e.g., such as adding a pipeline to an area where gas has previously been stranded). In many cases such projects involve joint decisions between the upstream and downstream (e.g., pipelines and liquefaction facilities), which will necessitate some interaction between submodules. In the case of the NGTDM, these decisions are handled separately. The goal, short of making reasonable capacity expansion decisions, is for the process of deciding on capacity expansions to be incorporated as seamlessly as possible within the overall modeling structure.

### *Geopolitics/Noneconomic factors*

Energy economic models are generally driven by basic economic principles: balancing supply and demand, maximizing revenue, minimizing cost, etc. subject to a number of constraints. Frequently these simple models assume something like perfect competition between market players. Within NEMS, EIA generally assumes that U.S. markets are competitive, subject to current laws and regulations that can place limitations on the market (e.g., emissions standards) or involve adding costs (e.g., taxes) or subtracting costs (e.g., tax credits). Noncompetitive elements in international markets are much more common. The ability to capture the dynamics of geopolitical and other noneconomic influences on these market will be essential within GHySMo. This might require such factors as structural considerations (e.g., modeling OPEC's role in oil markets), setting limits (e.g., disallowing trade between some countries or assuming fixed product prices for a period of time), or adding risk premiums or other cost adders. GHySMo should make explicit allowances for an analyst to incorporate geopolitical factors within the model structure and identifying them as such (i.e., not just using a structure or adding to a value that was intended for a different purpose.) The types of levers that most correspond to the non-competitive elements of these markets should be identified and incorporated explicitly.

While representing existing contracts is essential in a short-term model, it can be challenging for a longer-term model. In some cases contracts lock participants into actions that might be significantly different than those they might take if not constrained by the contract; however, this may be less binding when many contracts contain mechanisms for flexibility (e.g., renegotiating clauses). In such cases, being able to model the impact of existing contracts could be important and some consideration should be made towards incorporating such information into the model. Given the large investments associated with natural gas infrastructure, it is conceivable that contract arrangements will continue well into the future as they provide greater assurances for investors and allow buyers greater security of supplies, albeit at potentially higher prices. The model should include an explicit representation of long-term supply contracts, as well as a mechanism, potentially driven by analyst judgment, for changing the impact of current and future contracts over the projection period in a way consistent with observed market behaviors.

### *Benchmarking*

Opinions vary concerning the need for a longer-term model to align with historical values. In general, a longer-term model captures market drivers that might not reflect what might drive markets in the shorter-term. Nonetheless, it can be difficult to explain why a model produces results that are significantly different from historical levels. The requirement for GHySMo is to align within a specified time period with available historical values, and in some cases to benchmark to forecast values from the Short-Term Energy Outlook using a documented benchmarking standard. The standard will be produced as a part of the model development process.

### *Policy Levers/Options*

The initial focus for the GHySMo design is to support standard WEPS+ modeling efforts (i.e., in generating an IEO Reference case demand or macroeconomic based side cases). Another potential type of scenario analysis which is overarching is carbon emissions restrictions. The standard cases that will most involve GHySMo are the high/low world oil price cases. Running such cases can be challenging as EIA typically adjusts both supply and demand elements to develop a plausible scenario that will drive and keep prices higher or lower. GHySMo would be used in the process of developing these cases and setting up likely market assumptions to reach alternate price paths. The GHySMo design should facilitate this process.

It is anticipated that most scenario analysis requiring changes to GHySMo will involve changing some of the geopolitical levers or modifying input assumptions that are uncertain to better understand their importance. Some potential types of analyses follow: high/low oil and gas resources (e.g., high/low estimated ultimate recovery per well), high/low costs (specific or varied), high/low taxes or royalty rates, higher limits on constrained elements (e.g., limits on access to resources), faster/slower move away from contracts or to competitive based markets (e.g., in natural gas markets).

### *Resources to Run and Maintain*

Since GHySMo will be part of a larger system and run multiple times to obtain convergence, the model run time and memory footprint should be minimized. By themselves, the demand modules in WEPS+ run in less than one minute on a Windows 7 PC with a Core i7 processor (greater than 3 GHz) and at least two GB of RAM. With the macroeconomic module added, the system takes close to an hour to solve, usually in 6 iterations. In this mode, the demand modules in WEPS+ are using natural gas wholesale price curves supplied by INGM. While the addition of GHySMo is expected to add significantly to the overall run time, the goal is for the whole GHySMo to solve in less than one hour per iteration within the WEPS+ environment, with each standalone module taking a fraction of this time, so that WEPS+ with the macroeconomic module and GHySMo converge and solve in no more than 8 hours total (accounting for 5 to 6 iterations). Since WEPS+ convergence usually involves multiple iterations, within a single iteration GHySMo should of course run much faster. Issues related both to processing efficiency and to convergence are therefore important elements in the overall design. Otherwise, the computational formulation is unrestricted, with the caveat that it must be solvable with conventional, well-established, commercially available software and hardware.

While EIA anticipates needing considerable resources to develop and implement GHySMo, the modules should be designed and implemented to allow for ease of use and maintenance. Since it is recognized that significant resources might be needed simply to keep up with relevant events throughout the world, particularly to enable an analyst to update geopolitical drivers, a GHySMo requirement is that this information should be tracked in a well-structured knowledge management (KM) system, and not embedded in the model logic itself. Therefore, the input and output structure between the KM system and the model should be designed to support ease of updating, testing, modifying, and debugging, so that the model results are reasonably easy to trace back to either input assumptions or model logic. Once GHySMo is fully operational in a production environment, EIA expects to assign approximately 10 staff, who will apply 30% of their time to run and maintain the GHySMo model and knowledge management system. In addition, EIA expects to have minimal budget available beyond existing resources for the purchase of ongoing data in support of the module. A list of current data subscriptions will be provided. If any additional data is required beyond EIA’s current subscriptions, please describe the data series and why it is needed.

## Specific Upstream Requirements

The upstream submodule is intended to represent the economics of and geological factors involved in development (primarily drilling, but also mining/upgrading for oil sands and kerogen, etc.) and production of oil and gas, including that drilling decisions are impacted by prices for all products extracted from the ground. While it is anticipated that the submodule will project oil and gas production volumes for a given price or set of prices, the ultimate inputs and outputs will depend on the design of the overall system. Supply and demand balancing and the setting of equilibrated prices and volumes will occur elsewhere. So the submodule will effectively, if not in fact, be providing supply curves for oil and gas by type at the wellhead for use in downstream balancing routines, and respond appropriately to equilibrated prices and volumes.

Crude oil prices should respond to specified events inside supply and demand modules, according to clear predefined rules. For side case scenarios, these rules might or might not rely on the use of data results from the corresponding reference case. Assumptions on the interactions between WEPS+ regions at different levels should play an important role. They will become relevant when the equilibrium in some regions is broken. A set of instructions will be given to the GHySMo model and as a consequence, a new set of prices and quantities will reflect the new situation.

### *General module requirements*:

- Able to benchmark production to history and STEO forecast (either 2 or 5 year)
- Operate in two modes of U.S. production:
  - solve for US endogenously
  - benchmark to NEMS results for US
- Operate in two world oil pricing modes:
  - world benchmark crude oil price is specified exogenously
  - endogenously solves for oil prices and production volumes
- Represent uncertainty in resource size, technology improvement rates and costs
- Provide annual supply potential as well as longer-term potential to other modules in GHySMo for evaluating the need for new facilities or transport mechanisms (i.e. LNG liquefaction/regasification facilities, refineries, etc.) or for adding capacity to existing infrastructure.
- Report writing for main table needs to create Table 21 [4]in the AEO.

### *Resource representation:*

- Input assumptions for all resources by some level of geographic "region" (country, supra-country region, basin, field, play, etc.) with a quantitative representation of total future production potential (e.g., remaining volumes of hydrocarbon material) and a flexible structure to represent the assay characteristics of each region

[4] http://www.eia.gov/oiaf/aeo/tablebrowser/#release=AEO2014&subject=0-AEO2014&table=19-AEO2014&region=0-0&cases=ref2014-d102413a

  - The model will need to distinguish between hydrocarbons based on ranges of API and sulfur contents (e.g. bitumen (API <10), medium sour crude, light sweet crude, etc.)
  - Wells can produce a spectrum of hydrocarbons in a variety of mixtures from heavy oil (e.g. Canadian oil sands) all the way up to mainly methane (e.g. Haynesville dry shale gas) and any number of mixtures in between. All production is in either a liquid or gaseous phase
  - Avoid aggregate NGL/HGL measures and instead specify ethane, propane, butane, C5+. While the current WEPS+ only includes a single demand category related to HGL (referred to here as liquid petroleum gases or LPG), this might change in the future
- Delineated by onshore, shallow offshore, deepwater offshore.
- Depth of drilling below surface to help with economic decisions
- Indications of resource type: large fields like Prudhoe Bay with a gas cap, small fields, low permeability formations (shale, tight), coalbed methane, etc.

***Open to ideas regarding:***
- Distinction between known and undiscovered resources
- Represent production from different types of technology: CO2 EOR, horizontal wells, many types of offshore well and gathering systems
- Associated-dissolved vs non-associated gas distinction, only if deemed necessary
- Reservoir mechanics over the life of the production of a deposit, such as first losing gas drive
- Estimates of greenhouse gas (GHS) emissions from production activities

***Production decision requirements:***
- Control of ramp rates, maximum levels, decline rates by region and liquid stream
  - Allow maximum supply increments to increase with profitability of tranche. Instead of setting a maximum supply increment for each resource tranche that is fixed in size for all years of the model, it seems to make sense to allow this amount to increase over time as a function of the current year's profitability for a tranche, reflecting the idea that more capital would be put to use in extracting more of the most profitable tranches of supply sooner, rather than later.
- Exploration and development costs (Is anything unique required for stranded gas?)
- Production (lifting) costs
- Region-specific (e.g., country-level) above-ground costs and restrictions
- Non-economic determinants of production
- Technology improvement rates

***Open to ideas regarding:***
- Cash flow tracking and option to require positive cash flows
- Account for economics and impact of reinjection of natural gas to enhance oil production, as well as account for emissions from venting and flaring of natural gas
- Seasonal constraints especially for resources located in cold areas (thaws and ice flows), but also wind events, hurricanes
- How are new regions or projects initiated? Do they need to see commodity prices for a period of time at particular level? How are the start-up costs involved with setting up a new field handled? Is there competition for resources or capital with other potential projects?

    - This is particular important for areas with zero current production but they have a known resource.
- How are tight/shale formations developed when the production from the whole region is the sum of many small capital investment decisions (individual wells) which have steep decline curves? Are there resource constraints (e.g., capital, labor, steel) that limit the development of these formations?
- Should production be a function of several different time periods, such as y/y change and average over 5 or 10 years?
- Should rig counts be included or a more generic unit of capacity?

### *Geopolitical requirements:*

- Revenue requirements for countries with national oil companies
- Tax and royalty rates
- Access limitations (outright bans or effective bans via prohibitive costs)
- Drilling restrictions or quotas
- Return to pre-conflict production levels after outages
- Mineral vs land rights
- Other policy and regulatory representations
- Types of operators (NOC domestic, NOC foreign, IOC, independents)
- Ability to target an OPEC market share
- Investigate structuring model to represent non-market based decision making that may influence production. For example, this would allow for certain OPEC countries and other to not produce at full capacity. This may affect price formation, production quantities, and resource development decisions in the model.

## Specific Downstream Liquid Market Requirements

The downstream representation of the liquids market will include representations of petroleum and non-petroleum refinery processing, biofuels production, accounting for hydrocarbon gas liquids (HGL), the transport of crude oil from production regions to refinery regions, and the transport of petroleum products from refinery regions to demand regions. GHySMo will include a mechanism for balancing supply and demand for each of the liquid products represented in the model, and for establishing the associated trade flows and prices. The selected algorithm should operate in conjunction with the upstream submodule to reflect key market dynamics of the real world market.

The requirements below are defined for two liquids production categories: 1) refineries and 2) biofuels and other non-petroleum technologies. Because of the impact, requirements within each category are presented from two perspectives: a) logistics designed outside the model and b) logistics included as part of the model requirements. This latter distinction is important because the regional requirements of the crude supplied and products demanded will change depending on the logistics option.

### *Refineries*

Description: Represents the conversion of crude oil into petroleum products, and given crude supplies and product demands by region determines petroleum product supplies and crude consumption by region to balance the regional markets. Specifically, the refinery representation must return crude oil consumption and end-use product prices to WEPS+. To aid convergence, it would be convenient if the refinery representation could also provide WEPS+ with an indicator of the elasticities at those points.

#### *Generic requirements independent of how logistics is modeled:*

- Realistically define the crude oil conversion process or algorithm (i.e., petroleum refining)
  a. Must be responsive to product demands
  b. Must realistically match product characteristics to corresponding input crude characteristics
  c. Must include production/processing costs
  d. Optimize refinery crude slates to produce product demands at least cost, subject to over-riding geopolitical constraints and regional inefficiencies
     i. Design may differ by region depending on price formation (i.e. market-driven or policy-driven region)
  e. The model should be able to balance refinery inputs and outputs in both mass and energy
  f. The model should be able to consume crude oil and intermediate products, and produce both intermediate and end-use products (some intermediate products can be either/both refinery inputs and refinery products)
     - Depending on data availability, the model reports should provide details of all environmental discharges (GHG, SOx, NOx, thermal discharge, and other toxic chemicals) from processing facilities
     - The model should report fuel consumed during the refining process, by WEPS+ regions, and fuel type, including but not limited to natural gas, still gas, LPG, distillate, residual fuel oil, catalyst coke, and electricity
- Define methodology for capacity expansion of refinery production processes
  a. Cost-based for market driven regions
  b. Endogenously executed
- Must have a flexible number of world refining regions
- Must have a flexible number of crude types represented (NEMS currently has 11)
- Must have a flexible number of product types represented (at a minimum, include those represented in WEPS+)
     - Must determine the appropriate level of product specification
  a. Is it feasible to require specified levels for octane, RVP, sulfur, cetane, MTBE limits, etc.?
  b. Is it feasible to define multiple classes of a single product instead of using standard product specifications requirements?
     i. How would country-specific demands be classified into these categories?
     ii. How would production of these special classes of products be produced from each crude type?
     - Consider the feasibility of representing unusual use of liquid fuels (e.g., direct burn of crude)

#### *Assuming logistics is modeled separately from refining:*

- Inputs from Logistics model

a. Receive from the product logistics model, net product demands, by type, by refining regions
   i. Net product demand represents local product demands, plus product exports, minus product imports
b. Receive from crude logistics model, a value for crude lay-down price separately for each refining region by crude type (maximum (net) crude availability could also be specified)

- Outputs to Logistics model
  a. Generate a product price for each product type at each refining region (FOB the refinery)
  b. Generate a net crude requirement/demand for each crude type for each refining region
     i. Net crude demand represents crude produced locally, plus crude imports, minus crude exports (i.e., crude processed by refinery in this region)

### *Assuming logistics is modeled in one module with refining:*

- Must have a flexible number of crude supply and product demand regions
  a. can be the same as the refining regions
  b. must be able to translate results into WEPS+ regions
- Must have flexible representation of transport modes, and regional transport routes, by mode
  a. Define which combinations of crudes and products can share transport modes
- Define transport capacity expansion methodology and transport costs for crude and products (endogenously executed)
- Interaction with WEPS+
  a. Input from WEPS+: product demands, by type and demand regions
  b. Output to WEPS+: wholesale and retail product prices, by WEPS+ regions and sectors
- Interaction with upstream crude supply model
  a. Input from upstream model: crude prices, by type and supply region
  b. Output to upstream model: crude demands, by type and supply regions

## *Biofuels and non-petroleum technologies*

Description: Represents the conversion of non-petroleum feedstock into blend components or intermediate liquid streams, and determines blend component prices, by refining regions.

### *Generic requirements independent of how logistics is modeled:*

- Realistically define renewables and other non-petroleum liquid fuels conversion processes or an algorithm that performs similarly
  a. Number of technologies represented must be flexible, and include, but not be limited to, the production of ethanol, biodiesel, CTL, GTL , BTL
  b. Must be responsive to blend component and/or product demands
  c. Yields should be based on mass and energy balance
  d. Must include production/processing costs
  e. Design may differ by region depending on market driven or politically driven region
- Regional representation must be a subset of refining regions, but also must be flexible
- With the constraint of availability of data, the model reports should provide details of all environmental discharges (GHG, SOx, NOx, thermal discharge, and other toxic chemicals) from production as well as processing facilities
- Represent current renewable laws and regulations and have the flexibility to add or change
- Provide chemical and energy characteristics of liquids produced, for product blend purposes

- Define methodology for capacity expansion of biofuels and non-petroleum liquid fuels production processes
    a. Cost based
    b. Endogenously executed
- Receive from the natural gas (NG) logistics model, NG supply curves, by refining region, for GTL feedstock
- Define methodology for representing biofuels and other non-petroleum feedstock supply curves, by processing region, including coal for CTL, biomass for ethanol and other liquids production

***Assuming logistics is modeled separately from biofuels and non-petroleum technologies:***

- Inputs
    a. Receive separate demand from logistics or coordinate demand satisfaction with the refining section comprehending local renewable regulation requirements
- Outputs
    a. Report feedstock and non-refinery liquid fuel production levels, by type and region
    b. Report fuel consumed during process operation, by type and region
    c. Report feedstock and non-refinery liquid fuel production costs, by type and region

***Assuming logistics is modeled in one module with biofuels and non-petroleum technologies:***

- Inputs
    a. Coordinate demand for non-petroleum liquid fuels production with refinery production blends to satisfy overall product demands
    b. Represent any limitations associated with local renewable regulation requirements and blend restrictions
- Outputs
    a. Report feedstock and non-refinery liquid fuel production levels, by type and region
    b. Report fuel consumed during process operation, by type and region
    c. Report feedstock and non-refinery liquid fuel production costs, by type and region

### Logistics Module

Assuming that Logistics is a standalone module, the module will need to model both crude movement from multiple crude production regions to multiple refining regions and product movement from multiple refining regions to multiple demand centers.

***This module will need to:***

- Realistically represent marine movements among regions accounting for:
    a. Ship size
    b. Availability
    c. Port restrictions (draft, air draft, LOA, DWT)
    d. Optimize freight movements representing various routes and tonnage restrictions for multiple trade routes (canal passage versus maximum vessel size along with vessel cost)
    e. Account for current fleet and usage of that fleet
    f. Allow for freight rate (WS percentage) as fleet utilization changes

- Provide crude price (availability) by refining region by crude type to the Refining module
- Provide product demand (price) by refining region by product type to the Refining module
- Optimize the distribution of crude supply from the Crude module to the Refining module and the distribution of products from the refining module to demand centers considering
    a. Refinery module feedback on product production and production cost and preferred crude type and volume
- Input to the Logistics model is expected to be:
    a. Crude availability (quantity) by-grade by-crude-production-region (from the Crude module)
    b. Crude price by-grade from the Crude module
    c. Product demand by-product by-region (demand region)
- Given non-refining product production (from renewables and xTL centers):
    a. Account for transport of those non-petroleum products to demand centers
    b. Account for renewable mandates met by renewable fuels on a by-demand center basis
    c. Represent current renewable laws and regulations and have the flexibility to add or change
    d. Coordinate non-petroleum liquid fuels production with refinery production blends to satisfy overall product demands. Represent any limitations associated with local renewable regulation requirements and blend restrictions

### *Hydrocarbon gas liquids*

Given prices for crude oil, natural gas, and natural gas plant liquids, the upstream model will project the production of natural gas processing liquids and their component products (e.g., propane), as required. Currently, WEPS+ demands an aggregate product, referred to as liquid petroleum gases (LPG), although this could change in the future. The underlying assumption is that natural gas will be processed in the region that it is produced. These products will be combined with refinery gas liquids produced in the region and transported like the rest of the petroleum products to markets to satisfy demands. This component of the model will need to set a minimum processing fee for the liquids at the natural gas processing plant, to be used in establishing a floor price for the associated products delivered to the consumer. EIA anticipates that supply and demand for these products will largely be balanced on the demand side by adjusting the price to consumers (and a consistent price to be sent to producers), based on whether there is a surplus or shortage of these products in a given iteration, until convergence is achieved. Demand curves will likely be advantageous in this process.

## Specific Downstream Gas Market and Gas Logistics Requirements

The purpose of the downstream gas market representation is to provide the interface between the upstream producers of natural gas and the downstream consumers of natural gas, while capturing key current and likely future international natural gas market dynamics. Primary model inputs will be annual consumption of natural gas by sector by demand regions (e.g., the 16 WEPS+ regions), as well as supply potential from the upstream model, probably in the form of annual dry natural gas supply curves. The approach taken must complement and support EIA's two existing forecasting models, WEPS+ and NEMS, either through the development of an independent model, a modification of the existing model, or a combination thereof. The primary outputs of the model(s) will be production and trade volumes of natural gas by region that balance supply and demand, infrastructure capacity, and natural gas prices at key elements along the supply chain. Specific model requirements follow in no particular order:

- Capacities and rates should be included for natural gas pipelines, liquefaction facilities, LNG ships of at least two sizes, regasification facilities, and storage facilities.
- For each step in the process, the model must determine associated volumes, capacities, prices, and/or costs.
- As described earlier, some of these cost components might need to be distinguished by fixed and variable rates to properly capture market incentives (i.e., that fixed costs do not always influence flow decisions).
- While the model code should allow for flexible regions, the initial selection of regions should be based on the need to capture key transportation routes and disaggregate regions with distinct disparities, yet keep the regions in line with available data and resources.
- If the transportation regions are more granular than the demand regions, the methodology for disaggregating demands should be relatively simple and consistent with the algorithms in the WEPS+ demand models.
- Some consideration should be made in the model to account for seasonal demand and storage.
- LNG might need to be categorized by a few heat content levels, in which case LNG facilities might need to be a consumer of hydrocarbon gas liquids.
- As relevant, the model should allow for pipeline, LNG, and regasification facilities to be built on a contract basis initially. After that, the algorithm should be able to gradually move the contract relationship to a spot market relationship, as appropriate. In such cases, cost representations should likely increase to capture the transfer of investment risk to the facility owner.
- The model should provide an endogenous mechanism for allowing currently oil-based pricing to move toward gas-on-gas pricing based on market forces, as appropriate.
- The model methodology must be compatible and provide inputs directly usable in NEMS for modeling LNG trade with the United States. The representations of Canada and Mexico should be structured to allow for consistency between the two models.

# Appendix I: Operation of GHySMo within WEPS+

WEPS+ consists of a system of individual sectoral energy models or modules, using an integrated iterative solution process that allows for convergence of consumption and prices to an equilibrium solution. The WEPS+ platform allows the various individual modules to communicate through a shared database and provides a comprehensive, central series of output reports for analysis. Each module is called in turn and produces annual projected values through the forecast horizon given the values projected by the other models in the current or previous iteration (depending on the order in which the modules are called). Currently not all modules are fully integrated within the WEPS+ system, but that is an ultimate goal. When GHySMo is called for execution within WEPS+, it will need to similarly produce annual projected values for all years in the forecast period and be called in each WEPS+ iteration until overall convergence is attained.

The Reference case reflects the underlying relationships incorporated in the complete set of models interacting with each other in supply/demand relationships communicated through macroeconomic variables, prices, and consumption. The system of models is run iteratively to a point at which prices and consumption have converged to a reasonable equilibrium. Accumulated knowledge from the results of other complex models that focus on specific supply or demand issues and analysts' expert judgments also are taken into account and incorporated into the final projections. After the Reference case has been established, WEPS+ is used to run alternative cases that, for example, reflect different assumptions about future economic growth and energy prices. WEPS+ also can be used for other analyses, such as the effects of carbon prices.

WEPS+ currently produces projections for 16 regions or countries of the world, including OECD Americas (United States, Canada, and Mexico/Chile), OECD Europe, OECD Asia (Japan, South Korea, and Australia/New Zealand), Russia, other non-OECD Europe and Eurasia, China, India, other non-OECD Asia, the Middle East, Brazil, and other Central and South America. At the moment, the projections extend to 2040.

### *Demand*

In the individual models, the detail also extends to the subsector level. In WEPS+, the end-use demand models (residential, commercial, industrial, and transportation) and the power generation model project consumption of the key primary energy sources. The end-use models also provide intermediate consumption projections for electricity in the end-use demand sectors. The demand sector referred to as “district heating” consumes relatively small volumes and is not price responsive at this time. Unlike in the domestic market, crude oil is consumed directly in some other countries, likely requiring special handling within GHySMo. Specifically, consumption is currently specified for the following fuels for each sector, with the fuels supplied via GHySMo italicized:

| | |
|---|---|
| Residential | *Natural Gas*, *Distillate*, *Kerosene*, *LPG*, Heat, Coal, Electricity, Biomass, Solar |
| Commercial | *Natural Gas*, *Motor Gasoline*, *Distillate*, *Residual*, *Kerosene*, *LPG*, Heat, Coal, Electricity, Biomass, Solar |

| | |
|---|---|
| Industrial | *Natural Gas* (includes feedstocks, but excludes lease and plant fuel), *Motor Gasoline*, *Distillate*, *Residual*, *Kerosene*, *LPG* (includes feedstocks), *Petroleum Coke*, *Sequestered Petroleum*, *Other Petroleum* (includes feedstocks), *Crude oil* (consumed directly), Heat, Coal (includes coking coal), Electricity, Waste, Biomass, Geothermal, Solar, Other Renewables |
| Transportation | *Natural Gas* (excludes pipeline fuel), *Motor Gasoline*, *Distillate*, *Residual*, *LPG*, *Jet Fuel*, *Sequestered Petroleum*, *Other Petroleum*, *Crude oil* (consumed directly), Coal, Electricity, *Ethanol (E85)*, *Other Biofuels (Biodiesel)*, Hydrogen<br><br>[Note: the Transportation module is currently being enhanced and is likely to be expanding its slate of fuel types.] |
| Electric Power | *Natural Gas*, *Distillate*, *Residual*, *Crude oil* (consumed directly), Coal, Waste, Biomass, Hydroelectric, Geothermal, Solar, Wind, Other Renewables, Nuclear |
| District Heating | *Natural Gas*, *Distillate*, *Residual*, *Crude oil* (consumed directly), Coal, Waste, Biomass, Geothermal |

While not all of these sector/fuel combinations respond to price, the goal is still to project a wholesale price, and ultimately a retail price by sector, for each of the WEPS+ regions. Currently the WEPS+ integrating module estimates retail prices from wholesale prices projected by the supply models. In addition, it has an algorithm to project natural gas consumption for lease, plant, and pipeline fuel, which are not projected by the WEPS+ demand modules. However, GHySMo should set and account for these consumption levels and effectively override or replace this WEPS+ calculation.

The end-use model projections generally depend on retail supply prices, economic activity as represented by GDP (or gross output in the industrial sector), and population. The transformation models (electric power generation and district heat) satisfy electricity and heat requirements and also project consumption of primary energy sources at resulting price levels.

GHySMo is likely to benefit from using demand curves, however simple, to speed convergence of the WEPS+ system. A simple price/quantity pair from the previous iteration and an assumed elasticity would likely be sufficient.

### *Macroeconomy*

The Oxford Economics Global Economic Model (GEM) and Global Industry Model (GIM) are used to generate projections of gross domestic product (GDP) and gross output (GO) for the various IEO countries and regions and their respective industrial sectors. The theoretical structure of the GEM differentiates between the short and long-run for each country, with extensive coverage of the links between different economies. The GEM outputs GDP for use with WEPS+ and also provides drivers for the GIM. The structure of the GIM is based on input-output relationships, and this model outputs GO in the IEO sectors for each country or region in WEPS+.

The GEM produces a number of macroeconomic variables by country which could be used as input to GHySMo as needed, beyond GDP and population there are disposable income and industrial production as examples.  The GEM macroeconomic projections are based on projected values of such things as world dry gas production.  Currently, since the GEM is a relatively new addition to WEPS+, many of these types of variables that WEPS+ projects are also projected within the GEM by default.  In the future, however, the goal is to supplant these projections with projected values from WEPS+.  Specifically the following values, related to gas and liquids, should be projected and passed to the GEM from GHySMo during an integrated WEPS+ model run:

Natural gas

- Imports and exports by pipeline and LNG to each WEPS+ region
- Production by region
- Hub prices corresponding to NBP, Henry Hub, and Japan
- Wholesale (or ideally retail) natural gas prices by region

Liquid fuels

- Crude oil and petroleum product imports and exports to each region
- Crude oil and petroleum product production by region
- WTI and Brent prices
- Wholesale (or ideally retail) crude oil and petroleum product prices by region

### *Supply*

The current WEPS+ supply models (petroleum, natural gas, and coal) generate supply and wholesale price projections for the key supply sources corresponding to the primary consumption sources. The refinery model makes retail price projections for a variety of petroleum products based on the world oil price.   In addition, the supply models must provide the following projections to the WEPS+ system:

Global world oil price (currently Brent)

Natural gas production by region (i.e., 16 WEPS+ regions)

Petroleum and other liquids production by region:  crude oil, refined products, other liquids

The Generate World Oil Balance (GWOB) application is currently used to create a "bottom up" projection of world liquids supply—based on current production capacity, planned future additions to capacity, resource data, geopolitical constraints, and prices—and to generate conventional crude oil production cases. The scenarios (Oil Price cases) are developed through an iterative process of examining demand levels at given prices and considering price and income sensitivity on both the demand and supply sides of the equation. Projections of conventional liquids production through 2015 are based on analysis of investment and development trends around the globe. Data from EIA's Short-Term Energy Outlook are integrated to ensure consistency between short- and long-term modeling efforts. Projections of emerging liquids production are based on exogenous analysis.

Eleven major streams of liquids production are tracked on a volume basis: (1) crude oil and lease condensate, (2) tight oil, (3) natural gas plant liquids, (4) refinery gains, (5) Canadian oil sands, (6) extra-heavy oils, (7) coal-to-liquids, (8) gas-to-liquids, (9) shale oils, (10) ethanol, and (11) biodiesel.

Currently, projections of global natural gas production and trade are generated from EIA's International Natural Gas Model (INGM),[5] a tool (based on a linear program) that estimates natural gas production, demand (based on consumption projections from WEPS+), and international trade. It combines estimates of natural gas reserves, natural gas resources and resource extraction costs, energy demand, and transportation costs and capacity in order to estimate future production, consumption, and prices of natural gas. The INGM operates at a more granular level than the 16 regions in WEPS+, as well as accounting for three seasons in the year, to more accurately capture the dynamics of world natural gas markets.

INGM incorporates regional energy consumption projections by fuel from the WEPS+ model, as well as more detailed U.S. projections from NEMS, which are used to generate U.S. energy projections for the Annual Energy Outlook (AEO). An iterative process between INGM and WEPS+ is used to balance world natural gas markets, with INGM providing supply curves to WEPS+ and receiving demand estimates developed by WEPS+. INGM uses regional natural gas consumption projections from NEMS for the United States rather than those computed as part of the WEPS+ output, so that the final output for the United States is consistent with AEO projections.

[5] http://www.eia.gov/forecasts/aeo/nems/documentation/ingm/pdf/ingm(2011).pdf

# Appendix II: Operation of GHySMo within NEMS

NEMS is a large scale model, similar to WEPS+ in that it iterates between major supply, demand, and processing modules to arrive at an annual market equilibrium solution, but just for the domestic market. One difference is that NEMS equilibrates in one year before going on to the next, whereas WEPS+ equilibrates across all years simultaneously. Extensive documentation exists for NEMS on EIA's web site.

Some of the supply modules in NEMS include some limited representation of activities in Canada and Mexico to project cross border trade. Necessarily the coal, natural gas, and liquids modules have varying representations of ship borne international trade, which frequently involves the development and use of representative import and export curves for international fuel trade with the United States. These curves generally attempt to isolate out the relationship with the United States and the world, but are limited in their ability to capture the global impact of changes in trade volumes. While some information from WEPS+ is used in the development of these curves or to set exogenous projections in NEMS, the degree to which the model is able to replicate the international response to changes in trade with the United States is limited, which can also be said about the ability of WEPS+ to replicate the domestic response within NEMS.

As an example, as mentioned above INGM sets domestic consumption levels equal to AEO results exogenously. It also aligns somewhat with pipeline trade volumes projected in NEMS. However, on the supply side there is currently no relation between the representation of domestic supply in NEMS and in INGM. Even if one aligns the underlying data assumptions, this does not guarantee that the results would necessarily align (nor do they), particularly since different algorithms are used. Conversely, the Natural Gas Transmission and Distribution Module (NGTDM) uses projected values from INGM as a basis for projecting international LNG prices when making LNG export decisions for the United States, as well as using LNG import curves. It also uses some projected results, largely consumption, from WEPS+ for Canada and Mexico.

The goal for this project relative to NEMS is to improve on the process of aligning the two models to the degree possible. It is recognized that there is no ideal solution to this problem and that there will necessarily be differences regardless. However, the current approaches were largely developed in a piece-meal fashion and not during the process of the separate design processes of the two models. In the case of the NGTDM, a redesign process is occurring simultaneously with GHySMo that should support a better integration design. The expectation is that a more seamless and effective approach at aligning the two models can be developed in conjunction with the GHySMo design.

# Appendix III: Knowledge Management (KM) System Design

EIA understands that maintaining this type of model flexibility requires a sophisticated knowledge management (KM) system. The three basic functions of a knowledge management system are: 1) structuring, updating, and retrieving external data; 2) synthesizing model inputs at the appropriate level(s) of granularity from available data, and 3) ensuring that all constraints and levels of certainty are available during model calibration.

The first function encompasses those of any standard database. The database must be structured for convenient and efficient ingesting, updating, and retrieving of external data, whether that data is structured (e.g., a spreadsheet, external database, report table) or unstructured (e.g., analyst judgment). Ideally, the database will be designed with automated data ingest tools, so that regularly-scheduled new data releases can be quickly and easily incorporated into the database system. In all cases, the source of every piece of data will be identified, the ingest procedures will be documented, and the data itself will be vintaged and stored so that the model may always be run in a "roll-back" mode using only data that would have been available at a previous date. In some cases the data will be geographic in nature, so EIA is open to the idea of using geographic databases that include explicit understanding of graph structures, such as routes and the concept of nearness, if such a design will facilitate the management and manipulation of the data.

The second function of a knowledge management system is to translate the data collected in the database into inputs required by the model. Often, desired model inputs only correspond with available pieces of data if either/both the data and/or the model are constructed that way. In general, the model requirements will not exactly align with available data. In international energy modeling, data availability is not universal, and EIA expects many gaps in the knowledge requested by the GHySMo model. The KM system will need to be designed with a set of logical rules for aggregating, disaggregating where necessary, and filling in missing pieces of information so that the model can be run with a complete set of inputs that incorporate all the best available data. When the model requests a level of granularity higher than that at which the data is stored, aggregating is usually (but not always) straightforward. When the model requests more granularity for a particular input than is available in the data, the KM system must have guidelines for how to disaggregate higher-level data to a lower level. For example, demand at a multi-national regional level might be split out to the country level proportionately with country-level GDP, but VMT might be assumed to be the same across the entire region. These rules need to be established in advance for each of the model inputs and for all levels of aggregation the model could request. To aid interpretation of the model results, an indication should be passed from the KM system to the model of the level of uncertainty associated with each of the inputs. For instance, if an input comes from a reliable source, its level of certainty may be high; but if the input is disaggregated according to a sharing heuristic, then the level of certainty might be very low. It is therefore important for the KM system to have an understanding not just about what it knows, but also in what is less well-known.

The third function of a knowledge management system is done in conjunction with the model. Using information about what is known, what is approximated, and what is completely unknown, a particular model configuration can be calibrated against a known set of results. For instance, not knowing actual storage levels or pipeline flows, but knowing something about capacities, end-use demands, and prices, a model could be run in a mode in which it varies the unknown storage levels and flows until it reflects the measured prices. This mode of operation essentially sets default values of the less-well-known model inputs, and these inputs can then be stored in the KM system for later use in model runs in which the prices are not given as input. The KM system should be designed to facilitate this process of model calibration by indicating the level of uncertainty of each model input (so the model can tell which inputs may be changed in a calibration run), and it must be capable of ingesting and storing for later use the resulting calibrated values.

# References


Energy Information Administration, Annual Energy Outlook 2014, DOE/EIA-0383(2014), (Washington, DC, May 7, 2014), http://www.eia.gov/forecasts/aeo/

Energy Information Administration, International Energy Outlook 2013, DOE/EIA-0484(2013) (Washington, DC, July 25, 2013), http://www.eia.gov/forecasts/ieo/

Energy Information Administration, The World Energy Projection System Plus (WEPS+): An Overview 2011, DOE/EIA-M074(2011), (Washington, DC, August 12, 2011), web site http://www.eia.gov/forecasts/archive/m074(2011).pdf

Energy Information Administration, The World Energy Projection System Plus (WEPS+): Global Activity Module, (Washington, DC, October 2013) http://www.eia.gov/forecasts/aeo/nems/documentation/weps/pdf/m076(2013).pdf

Energy Information Administration, Model Documentation Report: International Natural Gas Model 2011 (Washington, DC, August 2013) http://www.eia.gov/forecasts/aeo/nems/documentation/ingm/pdf/ingm(2011).pdf

Energy Information Administration, Refinery Model of the World Energy Projection System Plus: Model Documentation, DOE/EIA-M084(2011), (Washington, DC, August 12, 2011) http://www.eia.gov/forecasts/archive/m084(2011).pdf

Energy Information Administration, The World Energy Projection System Plus (WEPS+): Greenhouse Gases Model, DOE/EIA-M081(2011), (Washington, DC, August 12, 2011) http://www.eia.gov/forecasts/archive/m081(2011).pdf

Energy Information Administration, National Energy Modeling System (NEMS): An Overview 2009, DOE/EIA-0581(2000) (Washington, DC, October 2009), http://www.eia.gov/forecasts/archive/0581(2009).pdf

Energy Information Administration, International Energy Module of the National Energy Modeling System: Model Documentation 2013 (Washington, DC, October 2013) http://www.eia.gov/forecasts/aeo/nems/documentation/international/pdf/m071(2013).pdf

Energy Information Administration, Model Documentation: Natural Gas Transmission and Distribution Module of the National Energy Modeling System, DOE/EIA-M062(2013) (Washington, DC, July 2013), http://www.eia.gov/forecasts/aeo/nems/documentation/ngtdm/pdf/m062(2013).pdf

Energy Information Administration, Model Documentation: Oil and Gas Supply Module (OGSM) of the National Energy Modeling System, DOE/EIA-M063(2013) (Washington, DC, May 2013), http://www.eia.gov/forecasts/aeo/nems/documentation/ogsm/pdf/m063(2013).pdf

Energy Information Administration, Liquid Fuels Market Model of the National Energy Modeling System: Model Documentation 2013, (Washington, DC, December 2013), http://www.eia.gov/forecasts/aeo/nems/documentation/lfmm/pdf/m059(2013).pdf

Energy Information Administration, Onshore Lower 48 Oil and Gas Supply Submodule (OLOGSS) development record, (2006-2011), http://www.eia.gov/oiaf/emdworkshop/model_development.html.