(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
02 July 2020 (02.07.2020)

WIPO | PCT

(10) International Publication Number
WO 2020/139346 A1

(51) International Patent Classification:
E21B 43/26 (2006.01) *G06N 20/00* (2019.01)
E21B 41/00 (2006.01) *G06N 3/08* (2006.01)

(21) International Application Number:
PCT/US2018/067688

(22) International Filing Date:
27 December 2018 (27.12.2018)

(25) Filing Language: English

(26) Publication Language: English

(71) Applicant: **HALLIBURTON ENERGY SERVICES, INC.** [US/US]; 3000 N. Sam Houston Parkway E., Houston, Texas 77032 (US).

(72) Inventors: **HEIDARI, Peyman**; 1501 Harvey Rd., Apt. 578, College Station, Texas 77840 (US). **BHARDWAJ, Manisha**; 2750 Holly Hall Street, Apt. 1703, Houston, Texas 77054 (US). **WALTERS, Harold Grayson**; 11715 Teal Hollow Ln., Tomball, Texas 77377 (US). **FULTON, Dwight David**; 14407 Anada Bay Ct., Cypress, Texas 77429 (US).

(74) Agent: **JUSTISS, Joel J.** et al.; Parker Justiss, P.C., 14241 Dallas Parkway, Suite 620, Dallas, Texas 75254 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report (Art. 21(3))*

(54) Title: HYDRAULIC FRACTURING JOB PLAN REAL-TIME REVISIONS UTILIZING COLLECTED TIME-SERIES DATA

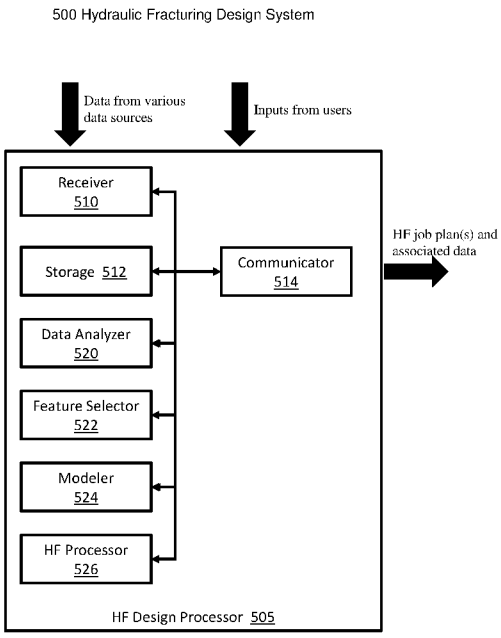

FIG. 5

(57) Abstract: The disclosure is directed to methods to design and revise hydraulic fracturing (HF) job plans. The methods can utilize one or more data sources from public, proprietary, confidential, and historical sources. The methods can build mathematical, statistical, machine learning, neural network, and deep learning models to predict production outcomes based on the data source inputs. In some aspects, the data sources are processed, quality checked, and combined into composite data sources. In some aspects, ensemble modeling techniques can be applied to combine multiple data sources and multiple models. In some aspects, response features can be utilized as data inputs into the modeling process. In some aspects, time-series extracted features can be utilized as data inputs into the modeling process. In some aspects, the methods can be used to build a HF job plan prior to the start of work at a well site. In other aspects, the methods can be used to revise an existing HF job plan in real-time, such as after a treatment cycle, a pumping stage, or a time interval.

WO 2020/139346 A1

HYDRAULIC FRACTURING JOB PLAN REAL-TIME REVISIONS UTILIZING COLLECTED TIME-SERIES DATA

TECHNICAL FIELD

[0001] This application is directed, in general, to designing a hydraulic fracturing stimulation treatment job plan, and, more specifically, to utilizing multiple data sources and machine learning to design and modify a hydraulic fracturing stimulation treatment job plan.

BACKGROUND

[0002] Understanding the impact of stimulation and completion variables, *i.e.*, features, on well production is important to improve the efficiency of hydraulic fracturing (HF) jobs. Some examples of stimulation features include fluid type, fluid volume, additive type and concentration, proppant type, size, concentration and mass, and pumping rate. Some examples of completion features include perforated length, number of stages, and number of perforation clusters. Additional features affecting well production include operator production practices and reservoir or spatial characteristics. The large number of features for a single well and across numerous wells in a region, location, basin, or county presents the need for large scale complex mathematical and statistical modeling. The relationships and interactions between these and other additional features are neither simple nor linear. Simple bilinear or multivariate linear models may not adequately explain the effect of the multiple input features on the well productivity outcomes. Well productivity and cost, together with efficiency and cost of the fracture stimulation treatment, determine the economic viability of the well and return on investment for the well operator.

[0003] Current approaches in modeling well production involve incorporating controllable (such as stimulation) features and uncontrollable (such as well location) features in the same statistical model. Often, well location, which can serve as a proxy for reservoir properties and operator practices, dominates the model accuracy and percentage of variance explained in the model error. Therefore, teasing out the effect of controllable features on production outcomes can be difficult. A more accurate modeling technique would be beneficial to predicting production outcomes using the many input feature types.

BRIEF DESCRIPTION

[0004] Reference is now made to the following descriptions taken in conjunction with the accompanying drawings, in which:

[0005] FIG. 1 is an illustration of a diagram of an example well site location;

[0006] FIG. 2 is an illustration of a diagram of an example multiple well system;

[0007] FIG. 3A is an illustration of a diagram of an example data preparation flow;

[0008] FIG. 3B is an illustration of a diagram of an example feature engineering flow;

[0009] FIG. 3C is an illustration of a diagram of an example production data processing flow;

[0010] FIG. 3D is an illustration of a diagram of an example statistical or machine learning predictive modeling flow;

[0011] FIG. 3E is an illustration of a diagram of an example hydraulic fracturing (HF) job design flow;

[0012] FIG. 4 is an illustration of a flow diagram of an example HF job design method;

[0013] FIG. 5 is an illustration of a block diagram of an example HF design system;

[0014] FIG. 6A is an illustration of a diagram of an example time-series data processing flow;

[0015] FIG. 6B is an illustration of a diagram of an example non-temporal data processing flow;

[0016] FIG. 6C is an illustration of a diagram of an example time-series predictive model flow;

[0017] FIG. 6D is an illustration of a diagram of an example time-series HF job design flow;

[0018] FIG. 7 is an illustration of a flow diagram of an example time-series HF job design method;

[0019] FIG. 8A is an illustration of a diagram of an example pressure response feature flow;

[0020] FIG. 8B is an illustration of a diagram of an example flow distribution feature flow;

[0021] FIG. 8C is an illustration of a diagram of an example fracturing dimensions flow;

[0022] FIG. 9 is an illustration of a diagram of an example general response feature flow;

[0023] FIG. 10 is an illustration of a diagram of an example job design adjustment flow;

[0024] FIG. 11 is an illustration of a flow diagram of an example response feature method;

[0025] FIG. 12A is an illustration of a diagram of an example graph demonstrating a selection of treatment and cycle start and end time points;

[0026] FIG. 12B is an illustration of a block flow diagram of an example treatment and cycle process flow;

[0027] FIG. 13 is an illustration of a diagram of an example graph demonstrating a machine learning selection of treatment cycles;

[0028] FIG. 14 is an illustration of a diagram of an example ensemble model flow with independent data sources;

[0029] FIG. 15 is an illustration of a diagram of an example ensemble model flow with independent and combined data sources;

[0030] FIG. 16 is an illustration of a diagram of an example ensemble model flow with independent data sources and data sources as direct input into the ensemble model;

[0031] FIG. 17 is an illustration of a diagram of an example ensemble model flow with a two-step ensemble model; and

[0032] FIG. 18 is an illustration of a diagram of an example ensemble model flow with independent and combined data sources, and a two-step ensemble model.

DETAILED DESCRIPTION

[0033] In the oil and gas industry, methods have been developed to estimate well characteristics using various diverse modeling approaches. A well can be one of various types, such as an oil or natural gas production well. Well characteristics can include, for example, productivity, completion efficiency, drilling direction, and many other characteristics. These modeling approaches encompass and utilize concepts from many different technical domains, such as numerical method, physics-based modeling, earth and reservoir modeling, chemical engineering, and data driven statistical and mathematical modeling.

[0034] Each of these modeling approaches relies on unique assumptions, concepts, hypotheses, and parameters from that technical domain to estimate well characteristics. Attempts have been made to incorporate some of these approaches together to obtain a combined estimate that may increase accuracy in estimating the well characteristic of interest. Most of the combination techniques have been heuristics or rule based and may involve manual or subjective processing of information from different domains.

[0035] The modeling, simulation, or prediction insights gained from the various techniques can be useful to well engineers. The model predictions can be utilized prior to the start of development of a well site. The information can be used to build or modify a job design plan, *i.e.,* a stimulation treatment operation plan. Once a well site has been started, other models can be used to monitor the progress of the well development and provide insight into modifications or changes to the job design plan that can provide a benefit. The benefit can be decreasing the time frame to hit a production goal, decreasing the cost to develop and work the well during the development and production phases, increasing the projected barrel of oil equivalent (BOE), increasing the net present value, return on investment, rate of return, or other economic indicator for the well following completion and production for a period of time, and impacting other key performance indicators or measures (KPI).

[0036] For example, conventional pumping decisions during hydraulic fracturing operations are generally not governed by data and statistics. Many decisions are subjective to the subject matter expertise (SME) of the fracturing crew. The decision-making process is geared towards executing a pre-determined pumping schedule (such as pumping 100,000 pounds (lbs.) of proppant in 30,000 gallons (gal) of fluid at 75 barrels per minute (bpm) for each treatment) or reacting to catastrophic deviations from expected outcomes (such as lowering the flow rate during a pressure increase caused

by a screenout). This fixed and reactive approach does not optimize job results. It seeks to complete the job as designed, with little mechanism for accepting feedback during the job and modifying the design to enhance the outcome.

[0037] The techniques and models of the disclosures herein can be applied to various types of well sites. This application will use the hydraulic fracturing (HF) of horizontal wells as the well type to demonstrate the disclosures herein. The same methods could be applied to vertical, deviated, or otherwise oriented wells. Conventional approaches to HF design require a large number of parameters as input that is usually unknown or not reliable. There can be over 10,000 parameters, *i.e.,* features, that can be input into the techniques and models. Some of the features can be well parameters, reservoir parameters, HF stimulation parameters, environment factors (such as geological formations, ground stability, offshore vs onshore, and other factors), equipment factors (such as the type of well equipment and monitoring equipment in use at a well site, and other equipment factors), operational feasibility (such as the ability to bring certain equipment on site, availability of SME, potential impact on neighboring locations, and other operational feasibility factors), and legal factors (such as laws in effect at the well site location, and other legal factors).

[0038] The significance and impact of each of these features on the KPIs can vary such that there can be 10, 100, or another number of features that are important for the analysis needed. The well engineers must often estimate several parameter values, which may add significant bias and uncertainty in the modeled predictions.

[0039] One dataset collected during HF pumping is time-series data containing information regarding fluids, proppants, chemicals, pressure, and other on-site field measurements and observations. This dataset can be leveraged for various applications and analysis. The HF time-series data captures information at regular and irregular time intervals. A grouping of time-series data can be an HF event. An HF event can be a treatment cycle, *i.e.,* a job stage, pumping stage, diversion cycles, minifracs, step-up, step-down, sand slugs, screenouts, water hammer, and other events. Each HF event captured can have different characteristics, such as shape, scale, magnitude, and other characteristics. Currently, there is no defined process to identify and detect these events. Current industry practice includes manual detection or marking of the events by a user that introduces subjective variability, or a user guided method in a software system. These methods are not automated nor provide consistent and robust results.

[0040] This disclosure describes several enhancements to the process of determining a HF job design prior to beginning work at a well site and during active conducting of an HF stimulation on a well site. One aspect provides for data driven statistical modeling and machine learning algorithms that can be utilized to perform the modeling and generate prediction outcomes. Other aspects provide for a systematic, scalable, and robust framework of combining different modeling and estimation techniques to provide an enhanced estimate of the well characteristics. For example, well

productivity can be estimated from earth models based on geological and reservoir information. It can also be estimated based on physics-based modeling techniques that rely on well geometry, fracture topology, and other factors.

[0041] Various data sources can contain unique and common information and parameters about the well site. When more than one data source is available, it would be desirable to join the different data sources to generate a combined data set for use in prediction and estimation of the well productivity. Modeling of predicted well production, or other KPI, can involve multiple different modeling assumptions and techniques, ranging from straightforward to very complex. One aspect of this disclosure can use an ensemble modeling technique. Ensemble modeling can be used to build a systematic way to incorporate diverse modeling techniques, methods, and data sources. This can be used to provide a more accurate estimate of well characteristics since the estimates can encompass a larger variety of information and parameters than analyzing each data source independently.

[0042] In another aspect, an automated detection of HF events based on HF time-series data can be generated. In addition to the presence or absence detection of an HF event, the start and end times of the HF event (HF event timeframe) can be identified. The HF event information can be utilized as modeling features for further estimations and predictions for the well site job operation plan, *i.e.,* job plan.

[0043] From a well productivity perspective, identifying HF events can be a step in building a data driven model to determine the effectiveness of the stimulation job plan. From the time markers, aggregated values, such as the average pump rate, total proppant mass, fluid volume, and more complex derived properties, such as perforation and near wellbore friction, and other factors, can be determined or computed for use in the modeling process. From a surface efficiency perspective, the HF events can serve as operational metrics, such as the number of screenouts per 1000 treatments and planned sand slugs vs the actual number. In addition, an automated post job report generation after a HF treatment can be an optional component.

[0044] In this aspect, a machine learning or data drive statistical modeling approach can be utilized, with the time-series pumping data, such as treating pressure, slurry rate, proppant concentration, user defined event flags, and other factors, to characterize the HF event (such as treatment, cycle, mini-frac, and other event types) and derived parameters (such as fracture gradient, slopes, averages, weighted averages, and other derived parameters). This information can be utilized to build a well site job plan.

[0045] Utilizing real-time or near real-time information, monitoring and improvements of the well site job plan can be implemented. Real-time, for the purposes of this disclosure, means information and data collected or received from a well site equipment or monitoring equipment at a well site within a time interval, for example, 1 second, 2 hours, or another time interval. Real-time processes and events as used herein is inclusive of near real-time processes and events. Non-real-time

collected data and information is received from stored data or historical data and may not be sourced from the target well site.

[0046] In another aspect prior to implementation of a job plan, a multi-stage data analytics method to maximize well productivity can be used to develop the job plan. The analytics method can automate processing of large-scale data from a variety of public and proprietary sources to prepare and extract standard and novel parameters. It can decouple spatial variability, *i.e.,* uncontrollable features, from stimulation parameters, *i.e.,* controllable features. It can also utilize uncharacteristic design parameters that capture variation in job designs along the lateral section of horizontal wells for building predictive machine learning models. The analytics method can also identify and recommend optimal and customized design features, *i.e.,* variables, for HF job plans.

[0047] These various aspects can be utilized separately or in combination to maximize value for a well site production. Data driven decisions prior to, or in real-time can be determined to satisfy customer defined KPIs by predicting likely outcomes and recommending well site job plan design changes.

[0048] In one aspect of this disclosure, training a model can be conducted to lead to higher quality outputs of the various respective models. First, a determination of manual or previous machine learning based identification of HF events from historical or other completed HF treatments can be collected into a dataset. The dataset can be one or more datasets, or an ensemble dataset. The dataset can be split into two or three portions. A training portion, a validation portion, and an optional testing portion. The training portion can be used to train a machine learning model. The validation portion can be used to validate the training model.

[0049] After training and validation have been completed, various models can be run against the dataset and the model can be selected that yields the outcomes desired by the customer KPIs. The testing portion can be used for this analysis. This model can then be used for the well site job plan. This type of HF event detection methodology is used in the industry for fault and anomaly detection with respect to equipment failure, such as drilling pipe stuck events and electrical submersible pump (ESP) failure. It is not being used for a machine learning based approach to solve multi-variate HF time-series event detection and its timeframe determination situations.

[0050] Turning now to the figures, FIGs. 1 to 2 demonstrate example well systems to which the various aspects disclosed herein can be applied. FIGs. 3 to 5 demonstrate examples of applying the disclosure to the pre-job design phase, *i.e.,* non-real-time application. HF job plans can be developed and modified based on the modeling demonstrated in these figures. FIGs. 6 to 7 demonstrate examples of applying the disclosure to real-time job executions. HF treatments in progress can be monitored and revised models can be generated that can confirm the HF job plan or recommend changes to the HF job plan. FIGs. 8 to 11 demonstrate alternate techniques to model various treatments and job design parameters, specifically around response features, such as pressure,

flow distribution, and fracturing dimensions. FIGs. 12 to 13 demonstrate techniques to detect and monitor response features and events within a well system. FIGs. 14 to 18 demonstrate variations of ensemble modeling.

[0051] FIG. 1 is an illustration of a diagram of an example well site location 100. Well site location 100 can be the target well site for the HF job plans that are designed or modified using the methods disclosed herein. Well site location 100 includes well equipment 110 located at surface 105 and well controller 108 located proximate to well equipment 110. Well equipment 110 can include, for example, drilling rig, surface production wellhead, and other equipment to construct and produce the well. Well controller 108 can include, for example, drilling equipment, completions equipment, pumping equipment, monitoring equipment, and production equipment. Prior to the stimulation treatment, well controller 108 may also include computing equipment to execute the job plan and to perform the modeling as disclosed. Alternatively, the modeling and data computing equipment can be located a distance from well site location 100, such as in a data center, server, or other remote location.

[0052] Extending below well equipment 110 is a wellbore 120. Wellbore 120 bends and becomes more aligned approximately horizontally, as shown by horizontal wellbore 122. A fracture 125 is shown for demonstration of a location of a planned hydraulic fracture to be created by a planned HF job design.

[0053] FIG. 2 is an illustration of a diagram of an example multiple well system 200. Multiple well system 200 includes a well site 201, a completed well site 202, and a computing system 203. Well site 201 includes well equipment 210 located at surface location 205 and well controller 208. Extending below well equipment 210 is a wellbore 220. Wellbore 220 is a horizontal wellbore designed for fracturing operations.

[0054] Completed well site 202 includes a completed well 230 and well controller and monitor 234. Extending below completed well 230 is a wellbore 238. Wellbore 238 is a horizontal wellbore designed for fracturing operations. Interior to wellbore 238 can be a sensing device 240, for example, an acoustic sensor or fiber optic cable. Sensing device 240 can be used to gather time-series feature data and response feature data that can be used as inputs to the various modeling algorithms described herein. Other devices, on surface and in the wellbore, can be used as well to gather the time-series and response feature data.

[0055] Computing system 203 can include one or more computing devices, cloud storage and processing systems, distant data centers, on-site data centers, servers, and other types of computing systems. A computing device can be a server, laptop, smartphone, dedicated well equipment, distributed processing system, and other types of computing devices. Computing system 203 can include a HF design system 250 (*see* FIG. 5) and one or more data sources 252. HF design system 250 can execute the methods, algorithms, and techniques described herein. Well controller

and monitor 234 can relay surface and downhole conditions and measured values before, during, and after the fracture stimulation treatment to the HF design system 250.

[0056] Data sources 252 can be located with HF design system 250, be located proximate to HF design system 250, or be located a distance from HF design system 250. Data sources 252 can be stored in a database, a hard drive, CD, DVD, USB, memory, server, data center, cloud storage, and other storage mediums and locations. Data sources 252 can be public data sources, private or confidential data sources, proprietary data sources, historical data sources, and other data source types.

[0057] HF design system 250 can be communicatively coupled to the data sources 252, and communicatively coupled to one or more well sites, such as well site 201 and completed well site 202, where such communicative coupling is shown by the dotted lines. The information collected from the data sources 252 and from sensing device 240 can be used by the HF design system 250 to build a new or revised HF job plan. Depending on the model being built, the data sources 252 or the sensing device 240 data can be optional. HF design system 250 can execute the methods and algorithms to build models prior to a job design being executed, or in real-time such as after a treatment cycle.

[0058] In multiple well system 200, data gathered from one well, such as the completed well 202 can also be used as input data to the HF design system 250 when building or revising a HF job plan for well site 201. Other combinations are possible as well. For example, there can be more than two well sites at a location, where one or more of those well sites are contributing data to HF design system 250.

[0059] FIG. 3A is an illustration of a diagram of an example data preparation flow 301. Data preparation flow 301 handles large scale ingestion, handling, cleaning, and processing of data from a variety of public and proprietary data sources. The process can include joining and validating data across different sources, performing statistical and descriptive analysis on the data, and identifying outlier data elements.

[0060] There can be one or more data sources, where three data sources are modeled for demonstration purposes as 310a, 310b, and 310m. The data sources 310 can be public data sources, for example, Drillinginfo, IHS, RigData, Rystad, RS Energy, FracFocus, national or state government oil and gas agencies, and other public sources. Data sources 310 can also be proprietary data sources, for example, HF job data from previous HF well sites, customer data, sales data, well cuttings, core, log, survey, and other data from the HF or other wells, seismic and other geology and reservoir data, and other proprietary or confidential data sources. The relevant data from each of the selected data sources 310 can be ingested, cleaned, and processed, as shown in flow 312. A series of quality checks and outlier removal processes can be applied to improve data quality at various processing steps. Imputation can be used to fill in missing values. Outlier data can be excluded. After the data has been processed and passed quality checks, the data from disparate sources can be combined together based

on the common information available between the data sources 310, for example, well API (American Petroleum Institute unique well identifier), well name, well number, and other data elements.

[0061] FIG. 3B is an illustration of a diagram of an example feature engineering flow 302. Feature engineering flow 302 includes extracting one or both standard and engineered features for use within the model building stages. The engineered features can be mathematical or statistical transformation of data, statistical or mathematical computation of derived features, SME knowledge driven, logic or rule-based features, or a combination of features across different data sources.

[0062] The feature engineering flow 302 takes as input the processed data from data preparation flow 301. The data can be analyzed and categorized. Standard features 320 include traditional and commonly available information about the well, for example, the well name, well number, API, spatial location, lateral length, measured depth, true vertical depth, and other information.

[0063] Logic based features 321 can be computed. Logic based features 321 can be computed using statistical or mathematical transformations 325 or computations 326 of standard features 320 or other types of features. The computation and usage of these designed features are distinctive to this disclosure. For example, a computed feature can be a toe-to-heel or stage-to-stage variation of proppant amount, fluid volume, maximum proppant concentration, proppant mesh size, pad fraction, fluid type, and other computed features. SME knowledge-based features 322 can be identified for the targeted well site. For example, for a new location, SME geological knowledge can be utilized since neighboring wells' data may not be available for the modeling process.

[0064] Features can be combined across different data sources as shown in flow 327. This can assist the analysis when certain features are at the well site level and other features are at the stage level, *i.e.,* treatment cycle. The treatment and pumping stage level features can be statistically aggregated at the well level while maintaining enough information to build a well level model. The collective output of feature engineering flow 302 can be a dataset or repository of well level features 329.

[0065] FIG. 3C is an illustration of a diagram of an example production data processing flow 303. Data processing flow 303 includes processing and cleaning of the production data to result in a cleaned production metric which can be used as a target variable for modeling. A type curve can be fitted, such as using the Arps equation or similar, to reduce noise in production data, overcome production reporting issues, and estimate production values per well, for example cumulative 180-day production outcomes and estimated ultimate recovery targets.

[0066] Depending on the region, the production data source 330 may have limited availability in the public domain or it may not be trustworthy. The production data can be ingested, cleaned, and processed, as shown in flow 332. One or more quality checks can be applied. Imputation can be used to fill in missing values. Outlier data can be excluded. In the decision flow

334, a determination can be made whether type fitting would be beneficial for the ingested data. If 'Yes', then a type curve, shown in flow 336, or another type of smoothing method can be applied to reduce noise and correct for erroneous production data, or to identify wells for which the production data is unusable and should be removed from the dataset. An appropriate target production value, *i.e.,* KPI, is computed from the smoothed or fitted data. The KPI can be, for example, a 30-day, 90-day, 180-day, 365-day, a maximum initial production, an estimated ultimate recovery (EUR), or other rolling cumulative production outcome. In flow 338, the KPI can then be combined with other well level features to generate a dataset to be used in the modeling steps.

[0067] FIG. 3D is an illustration of a diagram of an example statistical or machine learning predictive modeling flows 304. This can include a single predictive model or multi-stage model consisting of geospatial and predictive models. Modeling flows 304 include a first step to determine a feature selection. After joining the production data, as described in FIG. 3C, with well-level features, as described in FIG. 3B, a feature selection can be performed to reduce dimensionality of feature space, find relevant input features, or build simpler models using small number of features. Features are selected to be applied within one or both geospatial and predictive models. In some aspects, a single model can be built for the selected features. The geospatial modeling includes building a predictive geospatial model for well productivity based on spatial location parameters using mathematical techniques, such as universal kriging, decisions trees, or other machine learning methods. This can decouple the effect of the well location, which is an uncontrollable feature, from the stimulation parameters, which are controllable features. The predictive modeling includes using stimulation parameters along with additional relevant features to build a predictive model for well productivity. Information can be shared between the geospatial and predictive models.

[0068] The output of the feature engineering flow 302 is represented by the flow 329. The output of the production data processing flow 303 is represented by the flow 338. The combined dataset of production and well level features is represented by flow 342. The output of the model data is flow 344. From the output of the model data flow 344, feature selection can be performed as shown in flow 346.

[0069] The selection process in flow 346 can be achieved through many standard techniques and algorithms, for example, forward step-wise selection, backward step-wise selection, correlation analysis, lasso regression, ridge regression, elastic net regression, other techniques, and a combination of these techniques. Features can also be selected based on domain or SME knowledge. Flow 346 can select the same, different, or some same and some different features for each of the subsequent modeling steps, such as the output for the geospatial model flow 348 and the output for the predictive model flow 349. Flow 346 can also enable dimensionality reduction of the feature space along with decoupling the impact of relevant features from non-relevant features in predicting well production output.

[0070] In the geospatial modeling flow 350, the previously prepared data can be divided into three separate classes: (1) training data, (2) validation data, and (3) test data. Training data can be leveraged to build one or more models. Validation data can be used to evaluate the model performance and select a modeling algorithm. A geospatial model can be built using features that characterize well spatial location, either two or three dimensionally, and well completion parameters. Techniques such as universal kriging, random forest, and other machine learning methods can be used to build the models. The test data can be used to ensure model robustness and predictive power. The model resulting in the highest accuracy can be selected as the geospatial model to be used for the output of well production predictions flow 364.

[0071] In the predictive modeling flow 360, the data can again be divided into three classes (train, validate, test). The division of the data can be the same or different from the geospatial model division of data. The KPI can be the residual production (predicted well production output from the geospatial model) or it can be the actual well production output. A predictive model can be built using stimulation parameters, therefore decoupling the effect of location from the HF design parameters. One or more statistical and mathematical modeling techniques can be tested to build the model, for example, linear regression, non-linear regression, support vector machine, random forest, gradient boost, neural networks, deep learning, and other techniques. The algorithm resulting in the highest accuracy or with the best fitted data can be selected as the final predictive model. The evaluation of the algorithm and model performance can be achieved through cross validation, such as 5-fold or 10-fold cross validation. The selected model can output the predictive model well production predictions in the output of well predictions flow 364.

[0072] FIG. 3E is an illustration of a diagram of an example HF job design flow 307. HF job design 307 includes building an optimized model based on a list of input parameters, such as the parameters used for the geospatial and predictive models, to maximize or optimize well production in regard to a designated KPI metric. The HF job design 307 process can generate one or more potential job designs, where each optimizes a pre-defined objective function, or KPI, under given constraints for a given set of parameters. The potential job designs can then be validated under operational constraints and feasibility. The final recommended job design can be selected after the analysis is completed.

[0073] A pre-determined well can be selected or a well site can be identified for which the HF job design will be built, and is represented by flow 370. A list of features is selected for optimization in flow 372. These features can be selected from the features used to build the geospatial model and the predictive model. A set of constraints can be selected, as represented in flow 374. The constraints can be pre-defined or customized by a user of the process. The constraints can include operational feasibility, cost factors, and other constraint types.

[0074] An optimization algorithm, for example, a genetic algorithm, pattern search, differential evolution, and other types, can be used for the list of features within the provided constraints to generate multiple iterations and scenarios to optimize the objective function, as shown in flow 376. The most optimized scenario or one or more potential scenarios can be considered as potential designs for the HF job, shown as flow 378. Users of the process can provide their own estimates and ranges for one or more features in the list to optimize. This can create custom scenarios, shown as flow 380. Flow 382 encapsulates the previously presented statistical or machine learning predictive modeling flows 304. In flow 382, the resulting HF designs can be input to flow 304 to generate predicted well production or other KPI based on the scenario parameters. The output of 382 is the resulting output from the output of well predictions flow 364.

[0075] The output from the analyzation process, flow 384, can then be evaluated against the constraints to test that the constraints are satisfied, as shown in decision flow 386. If the constraints are not satisfied, the process can loop back to flows 378 and 380 to adjust the scenarios or to select a different scenario. If the constraints are satisfied, the design models can be converted into an HF job plan, including pumping schedule level job parameters, as shown in flow 388. The HF job plan can be output and utilized by field engineers to implement the HF job design, as shown in flow 390.

[0076] FIG. 4 is an illustration of a flow diagram of an example HF job design method 400. Method 400 starts at a step 401 and proceeds to a step 405. In the step 405, the process can prepare the first data set as shown in FIG. 3A. One or more datasets can be ingested, cleaned, and processed, rendering the datasets available for further analysis within the method 400. Proceeding to a step 410, engineering of features can be conducted, as shown in FIG. 3B. Features can be identified from the datasets, and additional features can be computed or transformed from the datasets. In addition, SME knowledge can be applied to identify additional features.

[0077] In a step 415, a second set of data, *i.e.,* a production data set, can be processed, as shown in FIG. 3C. Production data from various sources can be ingested, cleaned, and processed. The data can then be analyzed to determine if a fitting or smoothing algorithm should be applied to the data. Outliers can be removed if they do not fit the curve. In a step 420, data sets from steps 410 and 415 can be combined to create a larger model data set, as shown in FIG. 3D. The design parameters can be selected from the combined datasets for each of the models that will be built. The selected features can be the same, different, or overlap between the various models that are built.

[0078] Proceeding to a step 425, zero or more geospatial models can be created using the selected features from step 420 that were selected for geospatial modeling (*see* FIG. 3D, elements 348 and 350). Separately, in a step 430, zero or more predictive models can be created using the selected features from step 420 that were selected for predictive modeling (*see* FIG. 3D, elements 349 and 360). Various techniques and algorithms can be utilized to build the models. At least one of the steps 425 and 430 should be selected, with the other step being an optional selection.

[0079] After steps 425 and 430 have completed, the method 400 proceeds to a step 435. In the step 435, the various geospatial and predictive models can be evaluated against various criteria and constraints (*see* FIG. 3E). Models can be optimized for cost, short term production outputs, long term production outputs, and other KPIs. The model that best fits the constraints and KPI goals provided can be selected and converted into a HF job plan. The HF job plan can then be utilized by well engineers and operators, and by well site equipment. The method ends at a step 450.

[0080] FIG. 5 is an illustration of a block diagram of an example HF design system 500. HF design system 500 includes a HF design processor 505. HF design processor 505 includes a receiver 510, a storage 512, a data analyzer 520, a feature selector 522, a modeler 524, a HF processor 526, and a communicator 514. Data from various sources, public sources and confidential / proprietary sources, can be received by receiver 510. Receiver 510 can also receive various user inputs, such as SME knowledge, domain knowledge, user defined constraints, and other information. Receiver 510 can be a conventional communication system, such as a network connection, wireless connection, wired connection, or other communication types. The data can be stored in storage 512. Storage 512 can be a conventional storage medium, for example, a database, a hard drive, CD, DVD, USB memory, a cloud storage, a server, a data center storage area, and other storage medium types.

[0081] HF processor 526 can execute one or more applications and processes to perform the methods as described herein. HF processor 526 can manage the data stored in storage 512, and issue instructions to data analyzer 520, feature selector 522, and modeler 524 to complete the method steps. Data analyzer 520 can clean and process the data received by receiver 510 and store the processed data back into storage 512. Feature selector 522 can select features to be used in the modeling processes from the data stored in storage 512. Modeler 524 can generate geospatial and predictive models using the features selected in feature selector 522 and data stored in storage 512.

[0082] Communicator 514 can communicate the HF job plans selected by the HF processor. This can typically be the most optimized HF job plan based on the constraints provided. In other aspects, more than one HF job plan can be provided. Communicator 514 can communicate with a user, such as using a monitor, display, mobile device, or printer, and the communicator 514 can communicate with other systems, such as a storage medium, data center, server, cloud-based system, well site equipment, and other computing systems. The communication can be via a conventional communication means, such as a network protocol, wireless connection, wired connection, directly to a storage medium, and other communication types.

[0083] HF design system 500 is demonstrating an example system to implement the methods described herein. HF design system 500 is a logical description of the functionality. The functionality can be combined or separate. For example, storage 512 can be located at a distance in a data center. Data analyzer 520, feature selector 522, and modeler 524 can be combined with HF processor 526 where the different processes can be executed by one or more applications on a central

processing unit (CPU), graphics processing unit (GPU), single instruction multiple data (SIMD) processor, and other processor types.

[0084] FIG. 6A is an illustration of a diagram of an example time-series data processing flow 601. The process can leverage time-series data along with well characteristics to build a statistical model to predict KPI in real-time on location while the previous version of the HF job plan is being implemented. This aspect of the disclosure utilizes raw time-series HF pumping data rather than pre-extracted or pre-engineered features to build the revised model. In addition, the model uses other information, such as location, reservoir, and completion parameters. Historical data from a large set of wells is fed into the model for training. Predictions can be made on-location in real-time for a proposed well location or a specific well of interest. The statistical methodology can use various conventional models, such as deep learning, random forest, gradient boost, and other machine learning or statistical methods.

[0085] The first step of the flow is to collect production data and then clean and process the data. This is similar to the process described in FIG. 3C, production data processing flow 303. The next step is shown in the time-series data processing flow 601. The time-series HF data source flow 610 can be raw time-series HF pumping data. This data can span minutes, hours, days, or other time intervals of pumping data available for jobs in the historical data set. The data can be cleaned and processed for quality control as shown in flow 612. The time-series data is collected from the beginning of the data set up to a time t. Time t can be a point of time in the historical data or the present time.

[0086] There can be a variable amount of data available across different sized jobs. A decision point, such as decision flow 614, can be made to determine whether a change in the sampling of the data would be beneficial to the process. If 'Yes', then the data can be down sampled or up sampled to a pre-determined sample density for the wells to be used, as shown in flow 618. If no sampling changes are needed or if the sampling changes have been completed, the time-series data can be shifted to account for differences between surface and bottom hole values, as shown in flow 620.

[0087] A decision point, such as decision flow 622, can be made to determine whether a scaling change of the data would be beneficial to the process. If 'Yes', then the process proceeds to scaling flow 624 where the data can be scaled, *i.e.,* standardized or normalized, to a certain range based on the modeling method used. If no scaling changes are needed or if the scaling changes have been completed, the data can be output as shown in flow 626. The time-series data is presented as data up to a time t. The data can be used to predict KPI at a later point in time, such as $t+1$ or $t+n$, wherein is a selected time interval in the future. The time in the future can be labeled T, which leads to $T > t$.

[0088] FIG. 6B is an illustration of a diagram of an example non-temporal data processing flow 602. The modeling process can use specific information about a well that is not temporal, such as being fixed to the well, as shown in flow 630. For example, the location of the well. This flow may include data sourced and processed as in Fig. 3A, flow 301, or as engineered in Fig. 3B, flow 302. These additional non-temporal features can be processed separately, as shown in flow 632. The data can be formatted to be constant with the time-series data, as shown in flow 634. One aspect can utilize repeating the constant feature value across the length of the time-series data to maintain a constant time-series dataset. Another aspect can develop a multi-stage model where the features are separately provided to the next steps as input, as shown in output flow 636.

[0089] FIG. 6C is an illustration of a diagram of an example time-series predictive model flow 603. The data from the previous flows can be combined to make a single dataset. Production data can be collected as shown in flow 630a, the time-series data can be collected up to a time t, as shown in flow 630b, and the additional non-temporal data can be collected as shown in flow 630c. The combined dataset is shown in flow 632. The dataset can be divided into three separate classes or portions: (1) training data, (2) validation data, and (3) test data, as shown in flow 634. Training data can be used to build one or more models. Validation data can be used to evaluate the trained model and to assess different algorithms. The test data can be used as a blind dataset to check the model robustness and predictive power. The data portions can be created using a random partition selection or using a clustering algorithm based on specific features, such as location.

[0090] The data portions can then be used to develop models to predict well production or KPI output using statistical or machine learning techniques, such as deep learning, machine learning, Bayesian modeling, geospatial modeling, and other techniques, as shown in flow 634. Deep learning models can further utilize neural networks, such as convolutional neural networks, recurrent neural networks, other neural networks, or a combination of neural networks. In other aspects, the machine learning models can include random forest, gradient boost, decision trees, and other techniques. They can involve hierarchical and non-hierarchical techniques. The output can be a well production prediction at a specific time $t+n$, as shown in output flow 636.

[0091] FIG. 6D is an illustration of a diagram of an example time-series HF job design flow 604. A specific well, for which real-time job plan revisions are requested, can be selected, as shown in flow 650. A list of time-series features can be selected for optimization, as shown in flow 652. The features can be pre-defined or user selected. The pre-defined or custom constraints can be selected, as shown in flow 654.

[0092] The features from flow 652 and the constraints from flow 654 can be utilized in the optimization, *i.e.,* recommendation, model generation, as shown in flow 656. The optimization algorithm, such as a genetic algorithm, pattern search, differential evolution, and other types, can be applied to the collection of selected features within the provided constraints to generate multiple

iterations and scenarios to optimize the pre-defined KPI, as shown in flow 658. The process can then select the most optimized, or best fit, scenario given the constraints from a pool of top scenarios, to generate a predicted well production or other KPI at time $t+n$. This process can be represented by well production model flow 660, which can encapsulate the time-series predictive model flow 603. The output of the well production model flow 660 is the resultant of the output flow 636 of time-series predictive model flow 603.

[0093] The output can be a recommended job design revision that satisfies the constraints and KPIs, as shown in flow 662. For example, the model can recommend a change in proppant concentration from one treatment to another treatment based on the data collected on the location during the previous treatments. In some aspects, user input can be provided during the development phase of the model and the model can provide automated recommendations without further user input or feedback, once the process has been applied to the HF job plan.

[0094] FIG. 7 is an illustration of a flow diagram of an example time-series HF job design method 700. Method 700 demonstrates a way to adjust a HF job plan in real-time using treatment or stage data gathered from a well site. The changes can occur after a treatment or stage cycle is completed, or at another time point. Method 700 can build a model to assist in predicting what change to the HF job plan can move the HF job plan execution toward a better satisfaction of the identified KPIs.

[0095] Method 700 starts at a step 701 and proceeds to a step 705. In the step 705, the first data set is processed. The first data set is received, cleaned, and quality checks applied. For example, imputation can be used to fill in missing values and outlier data can be excluded from the first data set. The first data set is sourced from production values and data, reporting issues from one or more production wells, and one or more KPIs. Proceeding to a step 710, the time-series data set is received. The time-series data can be collected from HF pumping data from the well site and from historical data sources for the well site and nearby or similar well sites. The time-series data can be cleaned, processed, and have quality checks applied.

[0096] In a step 715, an optimized predictive model can be built. The predictive model can be built using a predictive data set including one or more elements of the first data set and one or more elements of the time-series data set. In a step 720, the HF job plan can be revised using the predictive model. Time-series data from the well site can be run through the predictive model to predict how the well site will react and produce over subsequent time intervals, such as treatment cycles. The method 700 ends at a step 750.

[0097] FIG. 8A is an illustration of a diagram of an example pressure response feature flow 800. FIG. 8B is an illustration of a diagram of an example flow distribution response feature flow 801. FIG. 8C is an illustration of a diagram of an example fracturing dimensions response feature flow 802. FIGs. 8A to 8C demonstrate various aspects of using response features as part of the time-

series data feature selection within the models developed in FIGs. 6A to 6D. The response features can be used as part of a model input to identify response features that can serve as a proxy for well production or other KPI predicted output. A second model input can be generated to establish a relationship between response features and on-location controllable features, such as flow rate, proppant concentration, and other controllable features. The response features thus can be controlled by means of the controllable features thereby leading to better modeling decisions to maximize KPI, such as production output.

[0098] Flow distribution feature flow 801 follows a similar path as described for pressure response feature flow 800. In place of the pressure response features, the flow distribution features can be used throughout the description of the pressure response feature flow 800. In one aspect, the flow distribution features can be extracted from data gathered from fiber optics data, for example, a distributed acoustic sensing (DAS) system.

[0099] Fracturing dimensions flow 802 follows a similar path as described for pressure response feature flow 800 and flow distribution feature flow 801. In place of the pressure response features, the fracturing dimensions features can be used throughout the description of the pressure response feature flow 800. In one aspect, the fracturing dimensions features can be extracted from data gathered from offset well-monitored microseismic data, in-hole-monitored microseismic data, offset well pressure monitoring data (*i.e.,* for fracture hits, such as from a pressure indication of communication between wells), offset well tiltmeter data, surface tiltmeter data, or physics-based modeling of pressure or net pressure response in real-time. In other aspects, other response features can be utilized in the algorithm.

[00100] In pressure response feature flow 800, flow 810 (and similar flows 830 and 850 for their respective response features) describes the extraction of pressure response features from time-series pumping data. The pressure response features can be identified from historic time-series pressure data or from real-time time-series pressure data. The features can be aggregated at different levels, for example over a treatment job, diversion cycle, pumping stage, or at a time interval, such as one second. The features can be extracted during the pumping operations or during a shutdown operation. An example feature for the pumping stage aspect can be first and second derivative of pressure with respect to time, non-linear spline fits of pressure responses, first, second, and third moment during pad phase, proppant pumping, diversion cycle properties, shape of entire treatments, and other types of features. During the shutdown operation, an example feature can be instantaneous shut-in pressure (ISIP), pressure decline slope post pumping, G-function (a dimensionless time function relating shut-in time to total pumping time) parameters, and water hammer amplitude, frequency, and decay. The features can be extracted using a statistical or machine learning method or can be based on SME engineering and domain knowledge. A combination of extraction techniques can also be used.

[00101] In the building of a model flow 812 (and similar flows 832 and 852 for their respective response features), the extracted pressure response features can be aggregated. The model can be built from the aggregated data using statistical or machine learning techniques that establish the relation of the pressure response features with the well production output. This can identify the statistically significant job response features. The pressure response features can then act as a proxy for the production output changes. Conventional machine learning models can be used to build the models, such as linear regression, regularization, ridge regression, lasso, support vector machine, random forest, gradient boosting, decision trees, and other types of algorithms.

[00102] In the determine optimal pressure response features flow 814 (and similar flows 834 and 854 for their respective response features), after the relevant response features are identified, the optimal combinations of the pressure response features can be determined. The goal is to determine the combination that will maximize the KPI.

[00103] In the build models using controllable features flow 816 (and similar flows 836 and 856 for their respective response features), the on-location features that can be modified during a job can be connected to the pressure responses. Conventional multivariate time-series forecasting models or machine learning or deep learning algorithms (such as recurrent neural network, long-short term memory (LSTM), and other models) can be used to build the models.

[00104] In the monitor pressure response features flow 818 (and similar flows 838 and 858 for their respective response features), the significant pressure response features identified in flows 812 and 814 can be extracted and monitored in real-time. A determination can be made whether the job response is favorable to the KPI, based on the analysis in model flow 812. If yes, then the job plan can continue as currently designed. If no, then one or more controllable features can be adjusted based on the models developed in controllable features flow 816 to attempt to cause the response features to move in a direction to improve the KPI results.

[00105] Multiple extracted response features can be combined in a flow. For example, pressure response and flow distribution response can be used in a single workflow, as input features to the model-building and subsequent steps. In another aspect of the method, flow distribution response and fracturing dimension response could be used together in a single workflow. In another aspect, pressure response and fracture dimension could be used together in a single workflow. In another aspect, pressure response, flow distribution response, and fracture dimension response could be used together in a single workflow.

[00106] FIG. 9 is an illustration of a diagram of an example general response feature flow 900. General response feature flow 900 is another view demonstrating an example flow for this aspect of the disclosure and is divided into sections, approximating the flows 800, 801, and 802 described in FIGs. 8A to 8C. The general response feature flow 900 begins with flow 910 where the historical time-series data is identified that is relevant for the well site job plan. The appropriate

response features are extracted in flow 915. The response features dataset is shown by flow 916. Flows 910, 915, and 916 are outlined by dash-dot box 961. Dash-dot box 961 represents the flows 810, 830, and 850 in FIGs. 8A to 8C respectively.

[00107] A model to predict the KPI is built in model building flow 920. It uses customer defined KPI parameters, as shown by flow 917. Model building flow 920 also uses uncontrollable features, as shown in flow 918. As part of the model building flow 920, a determination of which response features and their respective value ranges can better match the KPIs is made, as shown in flow 925. Flows 917, 918, 920, and 925 are boxed by dotted box 962. Dotted box 962 represents the model flows 812, 832, and 852 in FIGs. 8A to 8C respectively.

[00108] An evaluation can be made to determine whether the response features selected are significant in predicting the KPI parameters, as shown in decision flow 930. Should the resultant of decision flow 930 be 'No', then the response features can be discarded, as shown in flow 932. If the decision flow 930 is 'Yes', the response features are kept and used in further analysis and processing, as shown in flow 934. Flows 925, 930, 932, and 934 are boxed by a dashed line 963. Dashed line 963 represents the flows 814, 834, and 854 in FIGs. 8A to 8C respectively.

[00109] New models can be built to predict the significant response feature changes due to the controllable features, as shown in flow 940. The controllable features can be provided to flow 940, as shown in flow 938. A determination of how the controllable features impact the response features can be made based on the predictive model, as shown in flow 945. Flows 934, 938, 940, and 945 are boxed by a light gray line 964. Light gray line 964 represents the controllable feature flows 816, 836, and 856 in FIGs. 8A to 8C respectively.

[00110] FIG. 10 is an illustration of a diagram of an example job design adjustment flow 1000. Job design adjustment flow 1000 can be a real-time adjustment process for the job design parameters. Job design adjustment flow 1000 represents the flows 818, 838, and 858 in FIGs. 8A to 8C respectively. The model or models developed to predict the KPI from response features, from the flows described in FIG. 9, are retrieved for this process, as shown in flow 1018.

[00111] The flow process for the remaining flows form a partial loop mechanism where continuous monitoring, feedback, and adjustment is made to the HF job plan. Flow 1010 monitors the received data, such as from sensors in and near the well site system. The respective response features can be extracted from the received data, as shown in flow 1015. The new data set of extracted response features is shown by dataset flow 1016. The dataset flow 1016 is used as inputs to the model received from flow 1018.

[00112] A determination is made whether the response features are likely to have results favorable with respect to the KPIs, as shown in decision flow 1020. If the resultant from decision flow 1020 is 'Yes', then the HF job plan continues execution as designed, as shown in flow 1050. If the resultant from decision flow 1020 is 'No', then the models predicting response features from

controllable features is activated, as shown by flow 1024. One or more of the controllable features are adjusted, per the model predictions, as shown in flow 1026. The flow process from flow 1026 and from flow 1050 returns to flow 1010 for additional monitoring at the next time interval.

[00113] FIG. 11 is an illustration of a flow diagram of an example response feature method 1100. Method 1100 demonstrates a way to utilize extracted response features in conjunction with a prediction model. Response features can be received in real-time from a well site. Those response features can be run through as input into the prediction model. One or more features can be adjusted as input into the prediction model. The prediction model can output a recommendation on HF job plan changes in real-time to one or more features that can improve the well state with respect to the KPIs.

[00114] Method 1100 starts at a step 1101 and proceeds to a step 1105. In the step 1105, response features, such as pressure responses, flow distributions, and fracture dimensions, can be received. The response features can be gathered during pumping operations or during shutdown operations. In a step 1110, a prediction model can be built. The prediction model can use the response features as input. The prediction model can produce one or more scenarios so that as the response features are periodically updated, the predictions on the impact to the KPIs can be identified. Proceeding to a step 1115, a combination of features can be identified that could impact the KPIs. The features can be evaluated by how impactful they are to the KPIs. The features that are the most impactful can be identified and selected. The number of features can vary from one to the number of features gathered.

[00115] Proceeding to a step 1120, well site data from one or more treatment cycles can be received. In a step 1125 the HF job plan can be revised. The revision process can utilize the well site data and the prediction model. The revision process can keep the HF job plan as currently designed should the prediction model indicate, using the extracted response features, that the KPIs can be satisfied. If the KPIs may not be satisfied or could be improved upon, the revision process can select the features to adjust, based on the prediction model's output and extracted response features. The method 1100 ends at a step 1150.

[00116] FIG. 12A is an illustration of a diagram of an example graph 1200 demonstrating a selection of treatment and cycle start and end time points. A machine learning approach can be used to detect the presence or absence of an event, and the start and end of that event. In some aspects, the event properties can also be identified. An event can be a treatment, diversion cycle, sand slug, minifrac, rate step up sequence, rate step down sequence, initial shut-in pressure (ISIP), formation breakdown, screenout from time-series pumping data, and other event types. The pumping data can be treating pressure, slurry rate, proppant concentration, and other pumping data, along with available user defined data elements. The treatment and cycle information can be utilized by one or more of the time-series models described above, and provide additional analysis to the model generation flows.

[00117] Graph 1200 demonstrates a graph with three sections. Graph 1200 has an x-axis 1215 of time, which increases to the right. The three graph sections include the slurry rate 1220, the proppant concentration 1222, and the treatment pressure 1224. Graph 1200 has multiple vertical dashed lines, one of which is labeled as line 1218. Lines 1218 indicate the separation of each treatment cycle in the graph 1200.

[00118] Each treatment cycle includes a line graph indicating how each of the tracked data elements in each section correlate with the treatment cycle implementation. The slurry rate 1220 indicates a start of a detected event at point 1230. A maximum event response is shown by point 1232. The end of the event is indicated by point 1234. The proppant concentration 1222 demonstrates the same event data collection, shown by line graph portion 1240. The treatment pressure 1224 demonstrates the same event data collection, shown by line graph portion 1245.

[00119] A SME or an automated process can analyze the collected data, as shown in the graph 1200 to determine how the events correlate to the various treatment cycles. The information computed from this analysis can be used as input to the modeling phase of this aspect.

[00120] FIG. 12B is an illustration of a block flow diagram of an example treatment and cycle flow 1250. The treatment and cycle information 1260 can include three or more flows. For example, the treatment and cycle flow 1250 can have a data preparation flow 1270. Treatment and cycle information 1260 can be provided to data preparation flow 1270. The data preparation flow 1270 can identify events from a SME selection, as shown in flow 1272 or from other data sources, as shown in flow 1274. Next, events of interest can be identified, as shown in flow 1276, such as beginning and ending of treatments, which can be utilized to create a labeled dataset for training, validation, and test data for an event detection machine learning approach.

[00121] Following the data preparation flow 1270 can be a build model flow 1280 to automatically detect and identify such events. Using a subset of the data prepared in the data preparation flow 1270, one or more machine learning models can be trained, as shown in flow 1282. The models can then be validated against additionally collected data, as shown in flow 1284. Models can include deep learning, logistic regression, support vector machine, random forest, Bayesian, neural network, gradient boost, and other model algorithms. A model can be selected, as shown in flow 1286, based on the model that provides the results that best match the validation dataset, optimization capability and performance, ability to apply real world physical and operational constraints, ease of deployment, and other factors.

[00122] Following the build model flow 1280 can be a deployment of the selected model 1290. Events can be detected in unlabeled data using the model selected from the build model flow 1280, as shown in flow 1292. The model can incorporate available job data, as shown in flow 1294. Next, the deployment of the selected model 1290 can aggregate the resulting event types, times, and attributes in the data for further use, as shown in flow 1296.

[00123] FIG. 13 is an illustration of a diagram of an example graph 1300 demonstrating a machine learning selection of treatment cycles. Graph 1300 includes two portions, a graph 1305 demonstrating event data collected from a plug and perforate completion (plug-n-perf) sequence, and a graph 1306 demonstrating event data collected from a sliding sleeve completion sequence. Graph 1300 has an x-axis 1310 and 1330 of time, which increases to the right. Graphs 1305 and 1306 demonstrate graphs with three sections. The three sections include the slurry rate 1315, the proppant concentration 1316, and the treatment pressure 1317. Graph 1300 has multiple vertical lines, one of which is labeled as line 1319 on graph 1305, and line 1335 on graph 1306. Lines 1319 and lines 1335 indicate the separation of each treatment cycle in the graph 1300.

[00124] Example treatment cycles are labeled as treatment cycle 1320, treatment cycle 1322, and treatment cycle 1324 on graph 1305. Each treatment cycle includes a line graph indicating how each of the tracked data elements in each section correlate with the treatment cycle implementation. Similar to graph 1200, graph 1300 indicates how each of the events correlate to the treatment cycle implementation.

[00125] A SME or an automated process can analyze the collected data, as shown in the graphs 1200 and 1300, to determine how the events correlate to the various treatment cycles.

[00126] FIG. 14 is an illustration of a diagram of an example ensemble model flow 1400 with independent data sources. FIG. 15 is an illustration of a diagram of an example ensemble model flow 1500 with independent and combined data sources. FIG. 16 is an illustration of a diagram of an example ensemble model flow 1600 with independent data sources and data sources as direct input into the ensemble model. FIG. 17 is an illustration of a diagram of an example ensemble model flow 1700 with a two-step ensemble model. FIG. 18 is an illustration of a diagram of an example ensemble model flow 1800 with independent and combined data sources, and a two-step ensemble model.

[00127] Ensemble model flows 1400, 1500, 1600, 1700, and 1800 can represent one or more ensemble modeling aspects of this disclosure. Ensemble modeling can be utilized to support the flows previously described herein, utilized in any flow step wherein a predictive or analysis model is described. The ensemble modeling technique can be used to estimate or predict a well characteristic, for example, well production output or other KPI, completion efficiency, geospatial location, completion parameters, stimulation parameters, reservoir data, well logs, well directional surveys, drilling parameters, production data, fracture topology, fluid flow and dynamics, DAS, distributed temperature sensing (DTS) data, well header information, and other characteristics.

[00128] The ensemble model can consolidate a suite of modeling techniques and data sources, such as physics-based methods, numerical estimation methods, mathematical, statistical methods, earth models, geospatial, geological, reservoir-based models, data driven machine learning models, and other modeling techniques. Data sources can include information and parameters describing well characteristics. The ensemble modeling can combine some or all of the well characteristic

information and data sources in a systematic and scalable method. The ensemble modeling can be a mathematical or data driven statistical model, a weighted average, linear combination, a non-linear model, Bayesian hierarchical model, and other model types. The estimates from each of the models generated can vary and they may not converge towards the true value for the feature being examined. In addition, the data sources can share common underlying data. It can be beneficial to utilize the ensemble modeling techniques to combine the estimates from two or more models to obtain a single estimate that is reflective of the various models and data sources.

[00129] FIG. 14 ensemble model flow 1400 can build individual models for each of the data sources, parameters, and assumptions to obtain an estimate of well characteristics. The individual models can be incorporated together using a second stage modeling to obtain a combined estimate. The second stage modeling can develop a model that uses individual estimates as input parameters, and then build a new model to generate a single combined estimate. The second stage model can also be a weighted average, linear regression, or non-linear regression combination of the individual estimates. The second stage model can also be a Bayesian hierarchical graphical model where the individual estimates are nodes in the graph and the edges define relationships across different estimates or data sources.

[00130] There can be one or more data sources 1405, such as *n* number of data sources. These are labeled as data sources 1405-1 through 1405-n. The data sources can be public data sources, confidential data sources, proprietary data sources, and other data sources. The data sources can include well characteristics, *i.e.,* features. Well characteristics can include data and information about the wellbore, reservoir, and treatment. Each data source 1405 can have a respective modeling flow 1410. There can be *n* number of modeling flows 1410, matching the number of data sources 1405. The output from each respective modeling flow 1410 is an output 1415 of well characteristic estimates. There can be *n* number of outputs 1415, matching the number of data sources 1405. Data sources 1405, modeling flows 1410, and outputs 1415 can be grouped together to form an individual model layer step 1402.

[00131] The collected outputs 1415 can be used as inputs into an ensemble modeling flow 1422. The output 1424 of the ensemble modeling flow 1422 can be a set of final well characteristic estimates. The ensemble modeling flow 1422 and the output 1424 can be grouped together to form the ensemble model step 1420.

[00132] FIG. 15 ensemble model flow 1500 can be a variation of ensemble model flow 1400 where the first step can include building a model or a combination of models on a subset of data sources or parameters. The resulting output well characteristic estimates can be combined using an ensemble approach. The remaining parameters and data sources can be included with the combined estimate from the first step. This aspect can be applied to sources, methods, and parameters where

certain specific relationships should be taken care of prior to combining with the other independent sources.

[00133] There can be two or more data sources 1505, such as *m* number of data sources. These are labeled as data sources 1505-1, 1505-2, through 1505-m. As in ensemble model flow 1400, the data sources can be of various types and sourced from various data repositories. Each of the data sources 1505 are used as inputs into a modeling flow 1510. The output from the modeling flow 1510 is an output 1515 of well characteristic estimates.

[00134] In addition, there can be one or more data sources 1506, such as *n* number of data sources. These are labeled as data sources 1506-1 through 1506-n. The data sources, as previously described, can be of various types of data sources. Each data source 1506 can have a respective modeling flow 1511. There can be *n* number of modeling flows 1511, matching the number of data sources 1506. The output from each respective modeling flow 1511 is an output 1516 of well characteristic estimates. There can be *n* number of outputs 1516, matching the number of data sources 1506. Data sources 1505 and 1506, modeling flows 1510 and 1511, and outputs 1515 and 1516 can be grouped together to form an individual model layer step 1502.

[00135] The collected outputs 1515 and 1516 can be used as inputs into an ensemble modeling flow 1522. The output 1524 of the ensemble modeling flow 1522 can be a set of final well characteristic estimates. The ensemble modeling flow 1522 and the output 1524 can be grouped together to form the ensemble model step 1520.

[00136] FIG. 16 ensemble model flow 1600 can be an ensemble model where individual models can be built on a subset of data sources and parameters, and the estimates from the individual models can serve as input to the second layer model in addition to the other data sources and parameters as direct input. The model can then estimate from the first step, and raw parameters from the other data sources can be used as inputs into the second layer model to estimate the well characteristics.

[00137] There can be one or more data sources 1605, such as *n* number of data sources. These are labeled as data sources 1605-1 through 1605-n. The data sources, as previously described, can be of various types of data sources. Each data source 1605 can have a respective modeling flow 1610. There can be *n* number of modeling flows 1610, matching the number of data sources 1605. The output from each respective modeling flow 1610 can be an output 1615 of well characteristic estimates. There can be *n* number of outputs 1615, matching the number of data sources 1605.

[00138] In addition, there can be one or more data sources 1606-m. There can be *m* number of additional data sources. Data sources 1605 and 1606, modeling flows 1610, and outputs 1615 can be grouped together to form an individual model layer step 1602.

[00139] The collected outputs 1615 and the data sources 1606 can be used as inputs into an ensemble modeling flow 1622. The output 1624 of the ensemble modeling flow 1622 can be a set of

final well characteristic estimates. The ensemble modeling flow 1622 and the output 1624 can be grouped together to form the ensemble model step 1620.

[00140] FIG. 17 ensemble model flow 1700 extends the previously described ensemble model flows 1400, 1500, and 1600, by demonstrating that the two-layer model can be extended to multiple layer models encompassing various options, parameter combinations, and data or model relationships. Ensemble models can be built at different layers and then combined at the next layer using another ensemble model.

[00141] There can be one or more data sources 1705, such as *n* number of data sources. These are labeled as data sources 1705-1 through 1705-n. The data sources, as previously described, can be of various types of data sources. Each data source 1705 can have a respective modeling flow 1710. There can be *n* number of modeling flows 1710, matching the number of data sources 1705. The output from each respective modeling flow 1710 can be an output 1715 of well characteristic estimates. There can be *n* number of outputs 1715, matching the number of data sources 1705.

[00142] In addition, there can be one or more data sources 1706-m. There can be *m* number of additional data sources. Also, there can be one or more data sources 1707-p. There can be *p* number of additional data sources. Data sources 1705, 1706, and 1707, modeling flows 1710, and outputs 1715 can be grouped together to form a first individual model layer step 1702. The collected outputs 1715 and the data sources 1706 can be used as inputs into a first ensemble modeling flow 1722. The output 1724 of the first ensemble modeling flow 1722 can be a set of initial well characteristic estimates. The first ensemble modeling flow 1722 and the output 1724 can be grouped together to form the first ensemble model step 1720.

[00143] The output 1724 and the data sources 1707 can be used as inputs into a second ensemble modeling flow 1732. The output 1734 of the second ensemble modeling flow 1732 can be a set of final well characteristic estimates. The second ensemble modeling flow 1732 and the output 1734 can be grouped together to form the second ensemble model step 1730.

[00144] FIG. 18 ensemble model flow 1800 is a variation of ensemble model flow 1700 by demonstrating a variation of the multi-stage ensemble modeling. There can be two or more data sources 1805, such as *m* number of data sources. These are labeled as data sources 1805-1, 1805-2, through 1805-m. As in ensemble model flow 1400, the data sources can be of various types and sourced from various data repositories. Each of the data sources 1805 are used as inputs into a modeling flow 1810. The output from the modeling flow 1810 is an output 1815 of well characteristic estimates.

[00145] In addition, there can be one or more data sources 1806, such as *n* number of data sources. These are labeled as data sources 1806-1 through 1806-n. The data sources, as previously described, can be of various types of data sources. Each data source 1806 can have a respective modeling flow 1811. There can be *n* number of modeling flows 1811, matching the number of data

sources 1806. The output from each respective modeling flow 1811 is an output 1816 of well characteristic estimates. There can be *n* number of outputs 1816, matching the number of data sources 1806. Data sources 1805 and 1806, modeling flows 1810 and 1811, and outputs 1815 and 1816 can be grouped together to form an individual model layer step 1802.

[00146] The collected outputs 1816 can be used as inputs into a first ensemble modeling flow 1822. The output 1824 of the first ensemble modeling flow 1822 can be a set of initial well characteristic estimates. The first ensemble modeling flow 1822 and the output 1824 can be grouped together to form the first ensemble model step 1820.

[00147] The collected outputs 1815 (from the modeling flows 1810) and the outputs 1824 (from the ensemble model flow 1822) can be used as inputs into a second ensemble modeling flow 1832. The output 1834 of the second ensemble modeling flow 1832 can be a set of final well characteristic estimates. The second ensemble modeling flow 1832 and the output 1834 can be grouped together to form the second ensemble model step 1830.

[00148] Ensemble model flows 1400, 1500, 1600, 1700, and 1800 demonstrate some of the variations possible with this aspect of the disclosure. Additional variations are possible, such as increasing the number of layers, where each layer builds an ensemble model while joining additional data sources and analysis.

[00149] A portion of the above-described apparatus, systems or methods may be embodied in or performed by various digital data processors or computers, wherein the computers are programmed or store executable programs of sequences of software instructions to perform one or more of the steps of the methods. The software instructions of such programs may represent algorithms and be encoded in machine-executable form on non-transitory digital data storage media, e.g., magnetic or optical disks, random-access memory (RAM), magnetic hard disks, flash memories, and/or read-only memory (ROM), to enable various types of digital data processors or computers to perform one, multiple or all of the steps of one or more of the above-described methods, or functions, systems or apparatuses described herein.

[00150] Portions of disclosed embodiments may relate to computer storage products with a non-transitory computer-readable medium that have program code thereon for performing various computer-implemented operations that embody a part of an apparatus, device or carry out the steps of a method set forth herein. Non-transitory used herein refers to all computer-readable media except for transitory, propagating signals. Examples of non-transitory computer-readable media include, but are not limited to: magnetic media such as hard disks, floppy disks, and magnetic tape; optical media such as CD-ROM disks; magneto-optical media such as floptical disks; and hardware devices that are specially configured to store and execute program code, such as ROM and RAM devices. Examples of program code include machine code, such as produced by a compiler, and files containing higher level code that may be executed by the computer using an interpreter.

[00151] In interpreting the disclosure, all terms should be interpreted in the broadest possible manner consistent with the context. In particular, the terms "comprises" and "comprising" should be interpreted as referring to elements, components, or steps in a non-exclusive manner, indicating that the referenced elements, components, or steps may be present, or utilized, or combined with other elements, components, or steps that are not expressly referenced.

[00152] Those skilled in the art to which this application relates will appreciate that other and further additions, deletions, substitutions and modifications may be made to the described embodiments. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to be limiting, since the scope of the present disclosure will be limited only by the claims. Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this disclosure belongs. Although any methods and materials similar or equivalent to those described herein can also be used in the practice or testing of the present disclosure, a limited number of the exemplary methods and materials are described herein.

[00153] It is noted that as used herein and in the appended claims, the singular forms "a", "an", and "the" include plural referents unless the context clearly dictates otherwise.

[00154] Aspects disclosed herein include:

A. A method to revise a hydraulic fracturing (HF) job plan for directing operations of well site equipment for a well, including: (1) processing a final first data set, wherein the final first data set is computed from a preliminary first data set comprising estimated production values, reporting issues of a production well, and one or more key performance indicators (KPIs), (2) processing a time-series data set, wherein the time-series data set is generated from HF pumping data, (3) building a predictive model, utilizing a predictive data set comprising one or more of the final first data set and the time-series data set, and (4) revising the HF job plan, utilizing the predictive model.

B. A computer program product having a series of operating instructions stored on a non-transitory computer-readable medium that directs a data processing apparatus when executed thereby to perform operations to design a hydraulic fracturing (HF) job plan to direct operations of well site equipment of a well, having operations including: (1) processing a final first data set, wherein the final first data set is computed from a preliminary first data set comprising estimated production values, reporting issues of a production well, and one or more key performance indicators (KPIs), (2) processing a time-series data set, wherein the time-series data set is generated from HF pumping data, (3) building a predictive model, utilizing a predictive data set comprising one or more of the final first data set and the time-series data set, and (4) revising the HF job plan, utilizing the predictive model.

C. A system to revise a first hydraulic fracturing (HF) job plan for directing operations of well site equipment for a well, including: (1) a data analyzer, operable to generate cleaned data sets by analyzing, cleaning, correcting, and removing outlying data elements from each of received data sets, (2) a feature selector, operable to identify features utilizing a received job plan objective wherein the features affect the job plan objective, and wherein the features are identified in the cleaned data sets, (3) a modeler, operable to build predictive models utilizing the cleaned data sets and the features, and (4) a HF processor, operable to analyze each of the predictive models, operable to select a final predictive model, and operable to output a revised HF job plan.

Each of aspects A, B, and C can have one or more of the following additional elements in combination:

Element 1: communicating the HF job plan to a well site equipment of the well. Element 2: wherein the revising occurs after completion of a treatment cycle of the well or after a determined time interval. Element 3: wherein the building utilizes an ensemble model utilizing a single stage predictive model or a multiple stage predictive model, and wherein the ensemble model consolidates one or more modeling techniques and data sources. Element 4: wherein the modeling techniques are one or more of physics-based methods, numerical estimation methods, mathematical methods, statistical methods, earth models, geospatial models, geological models, reservoir models, and data-driven machine learning models. Element 5: wherein the data sources are one or more of geospatial location, completion parameters, stimulation parameters, reservoir data, well logs, well directional surveys, drilling parameters, production data, fracture topology, fluid flow and dynamics, distributed acoustic sensing, distributed temperature sensing, well header data, customer data, operator data, and service company data. Element 6: wherein the building utilizes one or more of a machine learning, neural network, deep learning, or statistical algorithm. Element 7: wherein the processing a final first data set further includes cleaning the preliminary first data set utilizing statistical techniques. Element 8: wherein the processing a final first data set further includes verifying the preliminary first data set utilizing statistical quality check techniques. Element 9: wherein the processing a final first data set further includes applying a smoothing technique to the preliminary first data set to generate a smoothed first data set, to realize one or more of reducing statistical noise, correcting erroneous data, and identifying unusable production well data. Element 10: wherein the processing a final first data set further includes computing a target production value utilizing the smoothed first data set, wherein the target production value is computed from a well production time interval. Element 11: wherein the processing a final first data set further includes generating the final first data set utilizing the smoothed first data set and the target production value. Element 12: wherein the processing the time-series data set further includes cleaning the HF pumping data utilizing statistical techniques. Element 13: wherein the processing the time-series data set further includes verifying the HF pumping data

utilizing statistical quality check techniques. Element 14: wherein the processing the time-series data set further includes sampling the HF pumping data utilizing up sampling or down sampling. Element 15: wherein the processing the time-series data set further includes shifting the HF pumping data aligning the HF pumping data with state of a bottom condition of the well. Element 16: wherein the processing the time-series data set further includes generating the time-series data set utilizing the HF pumping data. Element 17: wherein the shifting further includes scaling the HF pumping data, wherein the scaling utilizes determined value ranges for one or more HF pumping data elements. Element 18: wherein the predictive data set additionally utilizes a non-temporal data set, wherein the non-temporal data set is generated from non-temporal well data. Element 19: wherein the predictive data set additionally utilizes non-temporal well data set and further includes cleaning the non-temporal well data set utilizing statistical techniques. Element 20: wherein the predictive data set additionally utilizes non-temporal well data set and further includes verifying the non-temporal well data set utilizing statistical quality check techniques. Element 21: wherein the predictive data set additionally utilizes non-temporal well data set and further includes formatting the non-temporal well data set to align with the time-series data set. Element 22: wherein the building the predictive model further includes compiling the predictive data set by joining the final first data set and the time-series data set. Element 23: wherein the building the predictive model further includes partitioning the predictive data set, wherein the partitioning comprises a training data set and a validation data set. Element 24: wherein the building the predictive model further includes training a learning model utilizing the training data set and the validation data set. Element 25: wherein the building the predictive model further includes wherein the building a predictive model utilizes the learning model. Element 26: wherein the partitioning further comprises a test data set. Element 27: wherein the learning model is one or more of a deep learning algorithm, a machine learning algorithm, a Bayesian algorithm, a random forest algorithm, a gradient boost algorithm, or a decision tree algorithm. Element 28: wherein the deep learning algorithm is one or more of a convolutional neural network and recurrent neural network. Element 29: wherein the processing the time-series data set further includes receiving a first time-series pumping data set. Element 30: wherein the processing the time-series data set further includes receiving a second time-series pumping data set. Element 31: wherein the processing the time-series data set further includes identifying a first event set, wherein the first event set comprises event time intervals, utilizing the first time-series pumping data set. Element 32: wherein the processing the time-series data set further includes training a machine learning model utilizing the first event set. Element 33: wherein the processing the time-series data set further includes estimating a second event set, utilizing the second time-series pumping data set and the machine learning model. Element 34: wherein the processing the time-series data set further includes wherein the time-series data set utilizes the second event set. Element 35: wherein the first time-series pumping data set comprises one or more of treating pressure, slurry rate, and proppant

concentration, and wherein the second time-series pumping data set comprises one or more of treating pressure, slurry rate, and proppant concentration. Element 36: wherein the first event set and the second event set further comprise event property data. Element 37: wherein the first event set and the second event set comprise event types of one or more of treatments, diversion cycles, sand slug, minifrac, step-up, step-down, instantaneous shut-in pressure, breakdown, and screenouts. Element 38: wherein the first time-series pumping data set and the second time-series pumping data further comprise user defined data, and wherein the first event set and the second event set further comprise user defined data. Element 39: wherein the revising further includes selecting a target well for application of the HF job plan. Element 40: wherein the revising further includes selecting features from the final first data set or the time-series data set, where the features impact the KPIs. Element 41: selecting constraints, where the constraints impact the KPIs. Element 42: wherein the revising further includes generating one or more scenarios utilizing the features, the constraints, and the predictive model. Element 43: wherein the revising further includes building a well production model utilizing the scenarios. Element 44: wherein the revising further includes adjusting the HF job plan utilizing recommendations derived from the well production model. Element 45: a receiver, operable to receive the received data sets. Element 46: a communicator, operable to transmit the final predictive model, and the revised HF job plan. Element 47: a storage medium, operable to store the received data sets, the features, the predictive models, the first HF job plan, and the revised HF job plan. Element 48: a well controller, operable to receive the predictive model and the revised HF job plan, and execute the HF job plan. Element 49: wherein the received data sets comprise public, confidential, proprietary, well site location, and historical data sources. Element 50: a well pumping system, operable to provide pumping data to the data analyzer.

WHAT IS CLAIMED IS:

1. A method to revise a hydraulic fracturing (HF) job plan for directing operations of well site equipment for a well, comprising:

processing a final first data set, wherein the final first data set is computed from a preliminary first data set comprising estimated production values, reporting issues of a production well, and one or more key performance indicators (KPIs);

processing a time-series data set, wherein the time-series data set is generated from HF pumping data;

building a predictive model, utilizing a predictive data set comprising one or more of the final first data set and the time-series data set; and

revising the HF job plan, utilizing the predictive model.

2. The method as recited in Claim 1, further comprising:

communicating the HF job plan to a well site equipment of the well.

3. The method as recited in Claim 1, wherein the revising occurs after completion of a treatment cycle of the well or after a determined time interval.

4. The method as recited in Claim 1, wherein the building utilizes an ensemble model utilizing a single stage predictive model or a multiple stage predictive model, and wherein the ensemble model consolidates one or more modeling techniques and data sources.

5. The method as recited in Claim 4, wherein the modeling techniques are one or more of physics-based methods, numerical estimation methods, mathematical methods, statistical methods, earth models, geospatial models, geological models, reservoir models, and data-driven machine learning models.

6. The method as recited in Claim 4, wherein the data sources are one or more of geospatial location, completion parameters, stimulation parameters, reservoir data, well logs, well directional surveys, drilling parameters, production data, fracture topology, fluid flow and dynamics, distributed acoustic sensing, distributed temperature sensing, well header data, customer data, operator data, and service company data.

7. The method as recited in Claim 1, wherein the building utilizes one or more of a machine learning, neural network, deep learning, or statistical algorithm.

8. The method as recited in Claim 1, wherein the processing a final first data set further comprises:

cleaning the preliminary first data set utilizing statistical techniques;

verifying the preliminary first data set utilizing statistical quality check techniques;

applying a smoothing technique to the preliminary first data set to generate a smoothed first data set, to realize one or more of reducing statistical noise, correcting erroneous data, and identifying unusable production well data;

computing a target production value utilizing the smoothed first data set, wherein the target production value is computed from a well production time interval; and

generating the final first data set utilizing the smoothed first data set and the target production value.

9. The method as recited in Claim 1, wherein the processing the time-series data set further comprises:

cleaning the HF pumping data utilizing statistical techniques;

verifying the HF pumping data utilizing statistical quality check techniques;

sampling the HF pumping data utilizing up sampling or down sampling;

shifting the HF pumping data aligning the HF pumping data with state of a bottom condition of the well; and

generating the time-series data set utilizing the HF pumping data.

10. The method as recited in Claim 9, wherein the shifting further comprises:

scaling the HF pumping data, wherein the scaling utilizes determined value ranges for one or more HF pumping data elements.

11. The method as recited in Claim 1, wherein the predictive data set additionally utilizes a non-temporal data set, wherein the non-temporal data set is generated from non-temporal well data.

12. The method as recited in Claim 11, wherein the predictive data set additionally utilizes the non-temporal well data set and further comprises:

cleaning the non-temporal well data set utilizing statistical techniques;

verifying the non-temporal well data set utilizing statistical quality check techniques; and

formatting the non-temporal well data set to align with the time-series data set.

13. The method as recited in Claim 1, wherein the building the predictive model further comprises:

compiling the predictive data set by joining the final first data set and the time-series data set;

partitioning the predictive data set, wherein the partitioning comprises a training data set and a validation data set;

training a learning model utilizing the training data set and the validation data set; and

wherein the building a predictive model utilizes the learning model.

14. The method as recited in Claim 13, wherein the partitioning further comprises a test data set.

15. The method as recited in Claim 13, wherein the learning model is one or more of a deep learning algorithm, a machine learning algorithm, a Bayesian algorithm, a random forest algorithm, a gradient boost algorithm, or a decision tree algorithm.

16. The method as recited in Claim 15, wherein the deep learning algorithm is one or more of a convolutional neural network and recurrent neural network.
17. The method as recited in Claim 1, wherein the processing the time-series data set further comprises:

    receiving a first time-series pumping data set;

    receiving a second time-series pumping data set;

    identifying a first event set, wherein the first event set comprises event time intervals, utilizing the first time-series pumping data set;

    training a machine learning model utilizing the first event set;

    estimating a second event set, utilizing the second time-series pumping data set and the machine learning model; and

    wherein the time-series data set utilizes the second event set.
18. The method as recited in Claim 17, wherein the first time-series pumping data set comprises one or more of treating pressure, slurry rate, and proppant concentration, and wherein the second time-series pumping data set comprises one or more of treating pressure, slurry rate, and proppant concentration.
19. The method as recited in Claim 17, wherein the first event set and the second event set further comprise event property data.
20. The method as recited in Claim 17, wherein the first event set and the second event set comprise event types of one or more of treatments, diversion cycles, sand slug, minifrac, step-up, step-down, instantaneous shut-in pressure, breakdown, and screenouts.
21. The method as recited in Claim 17, wherein the first time-series pumping data set and the second time-series pumping data further comprise user defined data, and wherein the first event set and the second event set further comprise user defined data.
22. The method as recited in Claim 1, wherein the revising further comprises:

    selecting a target well for application of the HF job plan;

    selecting features from the final first data set or the time-series data set, where the features impact the KPIs;

    selecting constraints, where the constraints impact the KPIs;

    generating one or more scenarios utilizing the features, the constraints, and the predictive model;

    building a well production model utilizing the scenarios; and

    adjusting the HF job plan utilizing recommendations derived from the well production model.

23. A computer program product having a series of operating instructions stored on a non-transitory computer-readable medium that directs a data processing apparatus when executed thereby to perform operations to design a hydraulic fracturing (HF) job plan to direct operations of well site equipment of a well, having operations comprising:

processing a final first data set, wherein the final first data set is computed from a preliminary first data set comprising estimated production values, reporting issues of a production well, and one or more key performance indicators (KPIs);

processing a time-series data set, wherein the time-series data set is generated from HF pumping data;

building a predictive model, utilizing a predictive data set comprising one or more of the final first data set and the time-series data set; and

revising the HF job plan, utilizing the predictive model.

24. The computer program product as recited in Claim 23, further comprising:

communicating the HF job plan to a well site equipment of the well.

25. The computer program product as recited in Claim 23, wherein the building utilizes an ensemble model utilizing a single stage predictive model or a multiple stage predictive model, and wherein the ensemble model consolidates one or more modeling techniques and data sources.

26. The computer program product as recited in Claim 23, wherein the processing a final first data set further comprises:

cleaning the preliminary first data set utilizing statistical techniques;

verifying the preliminary first data set utilizing statistical quality check techniques;

applying a smoothing technique to the preliminary first data set to generate a smoothed first data set, to realize one or more of reducing statistical noise, correcting erroneous data, and identifying unusable production well data;

computing a target production value utilizing the smoothed first data set, wherein the target production value is computed from a well production time interval; and

generating the final first data set utilizing the smoothed first data set and the target production value.

27. The computer program product as recited in Claim 23, wherein the processing the time-series data set further comprises:

cleaning the HF pumping data utilizing statistical techniques;

verifying the HF pumping data utilizing statistical quality check techniques;

sampling the HF pumping data utilizing up sampling or down sampling;

shifting the HF pumping data aligning the HF pumping data with state of a bottom condition of the well; and

generating the time-series data set utilizing the HF pumping data.

28. The computer program product as recited in Claim 27, wherein the shifting further comprises:

scaling the HF pumping data, wherein the scaling utilizes determined value ranges for one or more HF pumping data elements.

29. The computer program product as recited in Claim 23, wherein the predictive data set additionally utilizes a non-temporal well data set and further comprises:

cleaning the non-temporal well data set utilizing statistical techniques;

verifying the non-temporal well data set utilizing statistical quality check techniques; and

formatting the non-temporal well data set to align with the time-series data set.

30. The computer program product as recited in Claim 23, wherein the building the predictive model further comprises:

compiling the predictive data set by joining the final first data set and the time-series data set;

partitioning the predictive data set, wherein the partitioning comprises a training data set and a validation data set;

training a learning model utilizing the training data set and evaluating the learning model utilizing the validation data set; and

wherein the building a predictive model utilizes the learning model.

31. The computer program product as recited in Claim 23, wherein the building the predictive model further comprises:

receiving a first time-series pumping data set, wherein the first time-series pumping data set comprises one or more of treating pressure, slurry rate, and proppant concentration;

receiving a second time-series pumping data set, wherein the second time-series pumping data set comprises one or more of treating pressure, slurry rate, and proppant concentration;

identifying a first event set, wherein the first event set comprises event time intervals, utilizing the first time-series pumping data set;

training a machine learning model utilizing the first event set;

estimating a second event set, utilizing the second time-series pumping data set and the machine learning model; and

wherein the building a predictive model utilizes the second event set.

32. The computer program product as recited in Claim 23, wherein the revising further comprises:

selecting a target well for application of the HF job plan;

selecting features from the final first data set or the time-series data set, where the features impact the KPIs;

selecting constraints, where the constraints impact the KPIs;

generating one or more scenarios utilizing the features, the constraints, and the predictive model;

building a well production model utilizing the scenarios; and

adjusting the HF job plan utilizing recommendations derived from the well production model.

33. A system to revise a first hydraulic fracturing (HF) job plan for directing operations of well site equipment for a well, comprising:

a data analyzer, operable to generate cleaned data sets by analyzing, cleaning, correcting, and removing outlying data elements from each of received data sets;

a feature selector, operable to identify features utilizing a received job plan objective wherein the features affect the job plan objective, and wherein the features are identified in the cleaned data sets;

a modeler, operable to build predictive models utilizing the cleaned data sets and the features; and

a HF processor, operable to analyze each of the predictive models, operable to select a final predictive model, and operable to output a revised HF job plan.

34. The system as recited in Claim 33, further comprising:

a receiver, operable to receive the received data sets;

a communicator, operable to transmit the final predictive model, and the revised HF job plan; and

a storage medium, operable to store the received data sets, the features, the predictive models, the first HF job plan, and the revised HF job plan.

35. The system as recited in Claim 33, further comprising:

a well controller, operable to receive the predictive model and the revised HF job plan, and execute the HF job plan.

36. The system as recited in Claim 33, wherein the received data sets comprise public, confidential, proprietary, well site location, and historical data sources.

37. The system as recited in Claim 33, further comprising:

a well pumping system, operable to provide pumping data to the data analyzer.

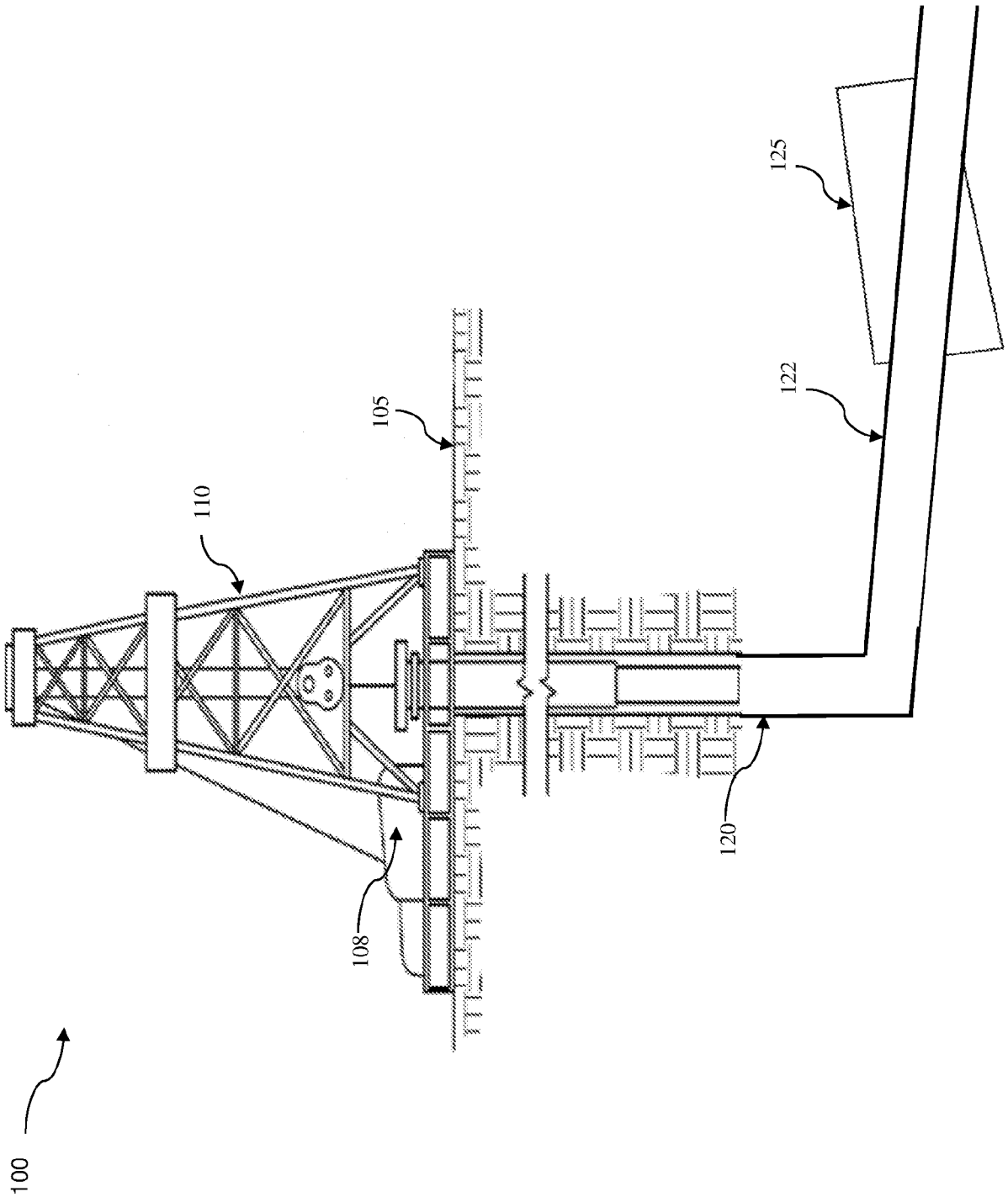

FIG. 1

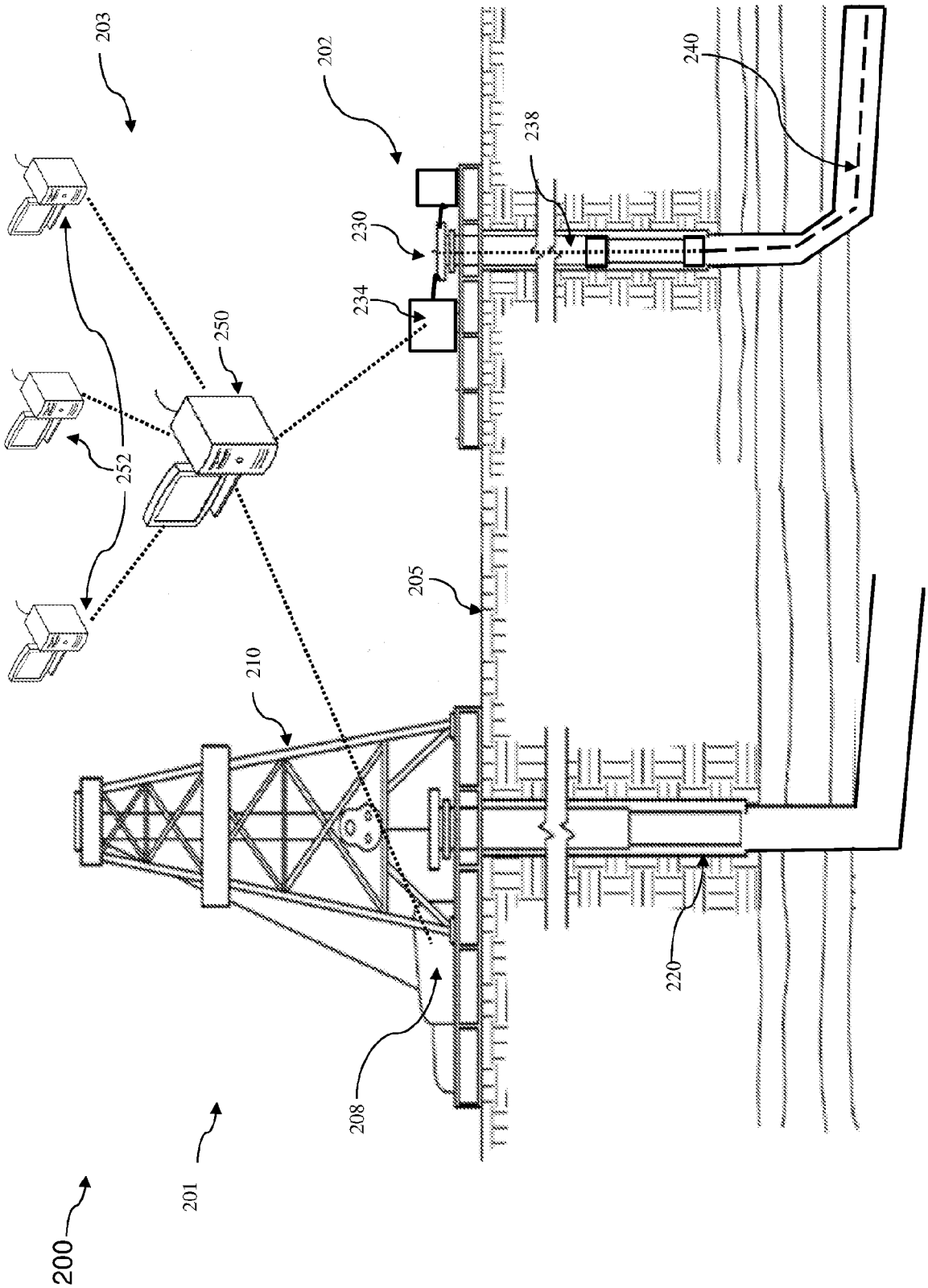

FIG. 2

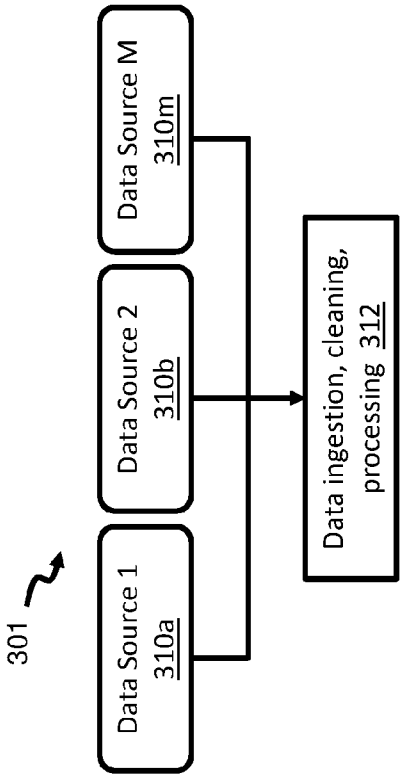
301
Data Source 1 310a
Data Source 2 310b
Data Source M 310m
Data ingestion, cleaning, processing 312

FIG. 3A

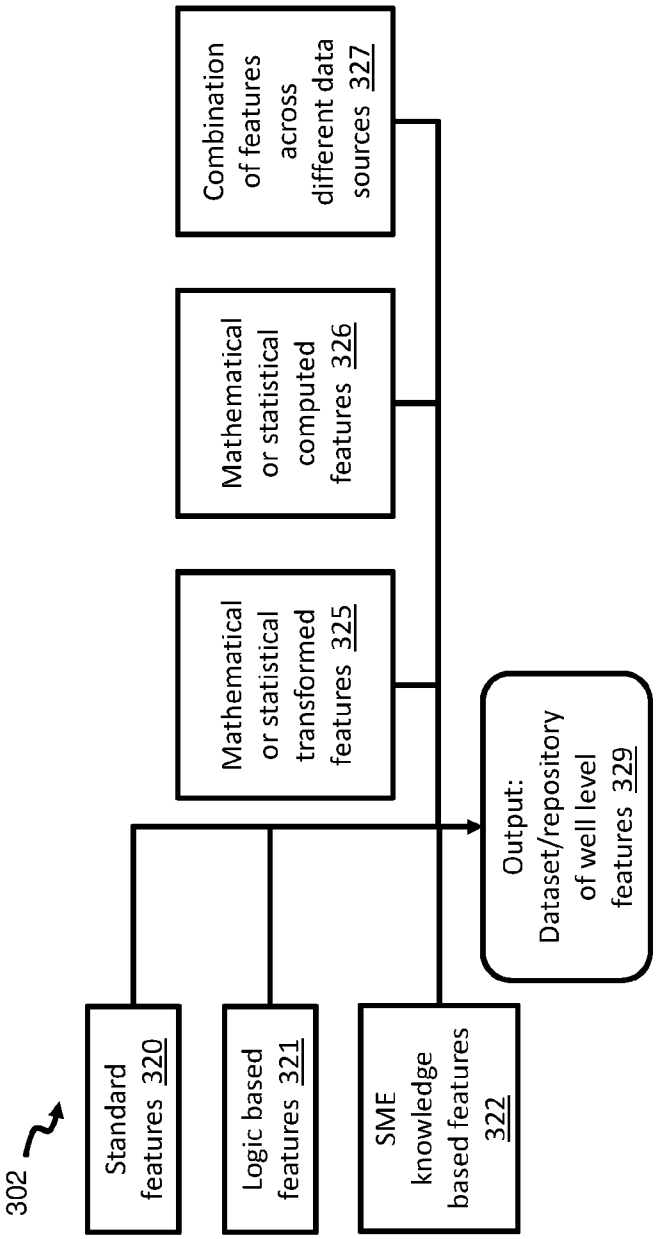
302
Standard features 320
Logic based features 321
SME knowledge based features 322
Mathematical or statistical transformed features 325
Mathematical or statistical computed features 326
Combination of features across different data sources 327
Output: Dataset/repository of well level features 329

FIG. 3B

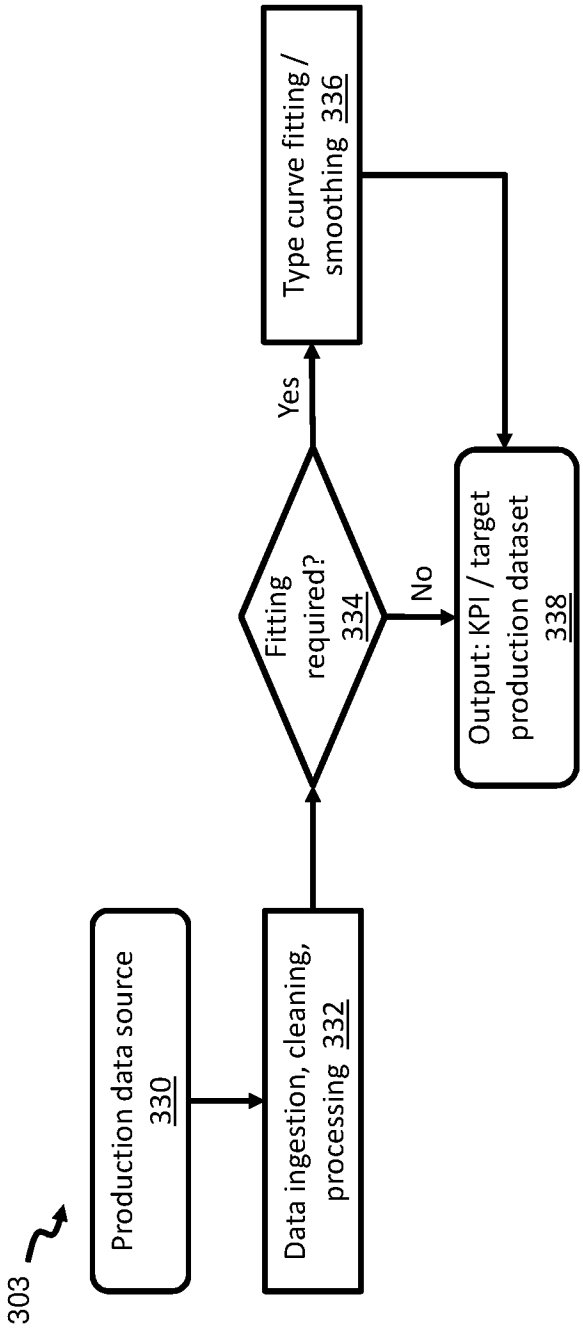
303
Production data source 330
Data ingestion, cleaning, processing 332
Fitting required? 334
Yes
No
Type curve fitting / smoothing 336
Output: KPI / target production dataset 338

FIG. 3C

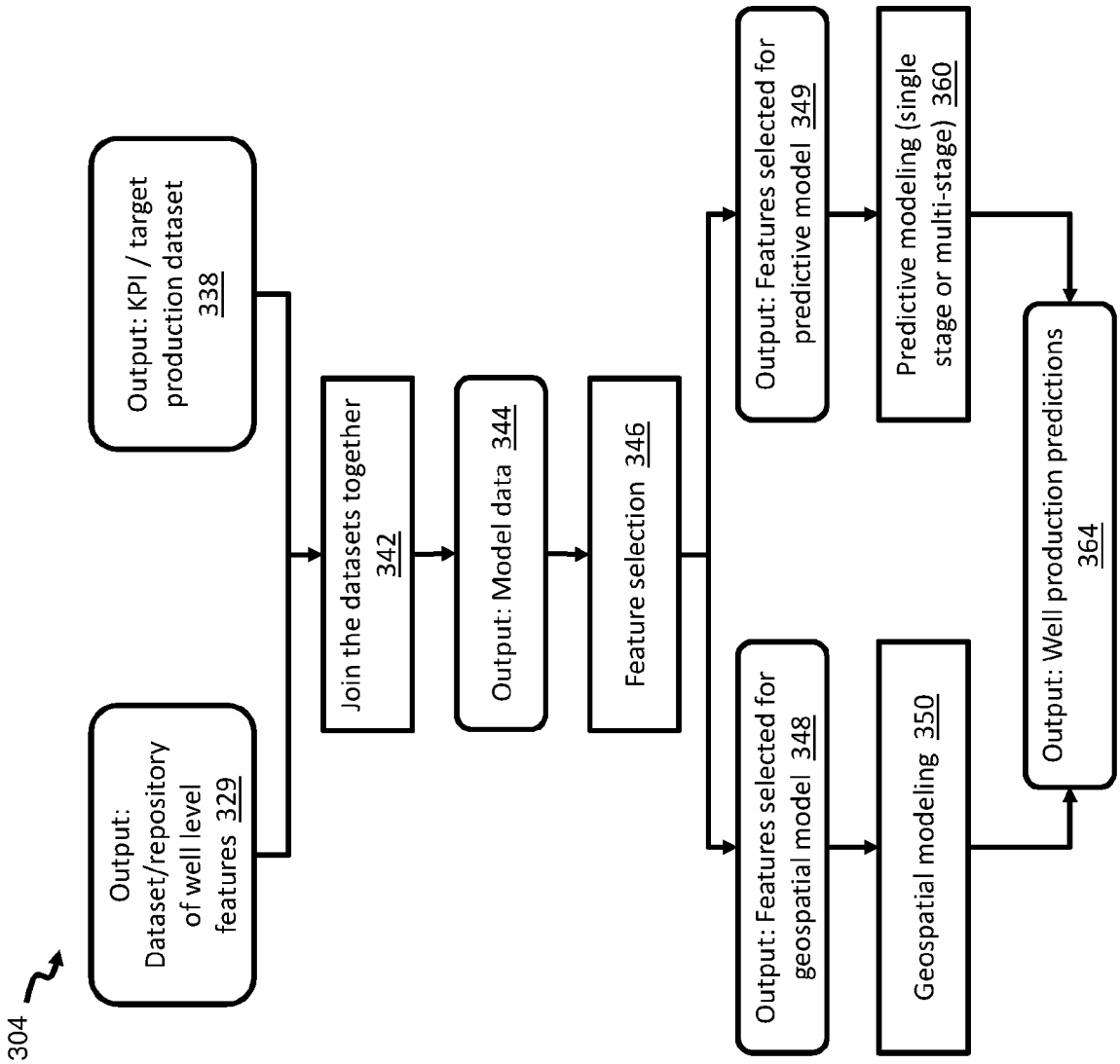
304
Output: Dataset/repository of well level features 329
Output: KPI / target production dataset 338
Join the datasets together 342
Output: Model data 344
Feature selection 346
Output: Features selected for geospatial model 348
Output: Features selected for predictive model 349
Geospatial modeling 350
Predictive modeling (single stage or multi-stage) 360
Output: Well production predictions 364

FIG. 3D

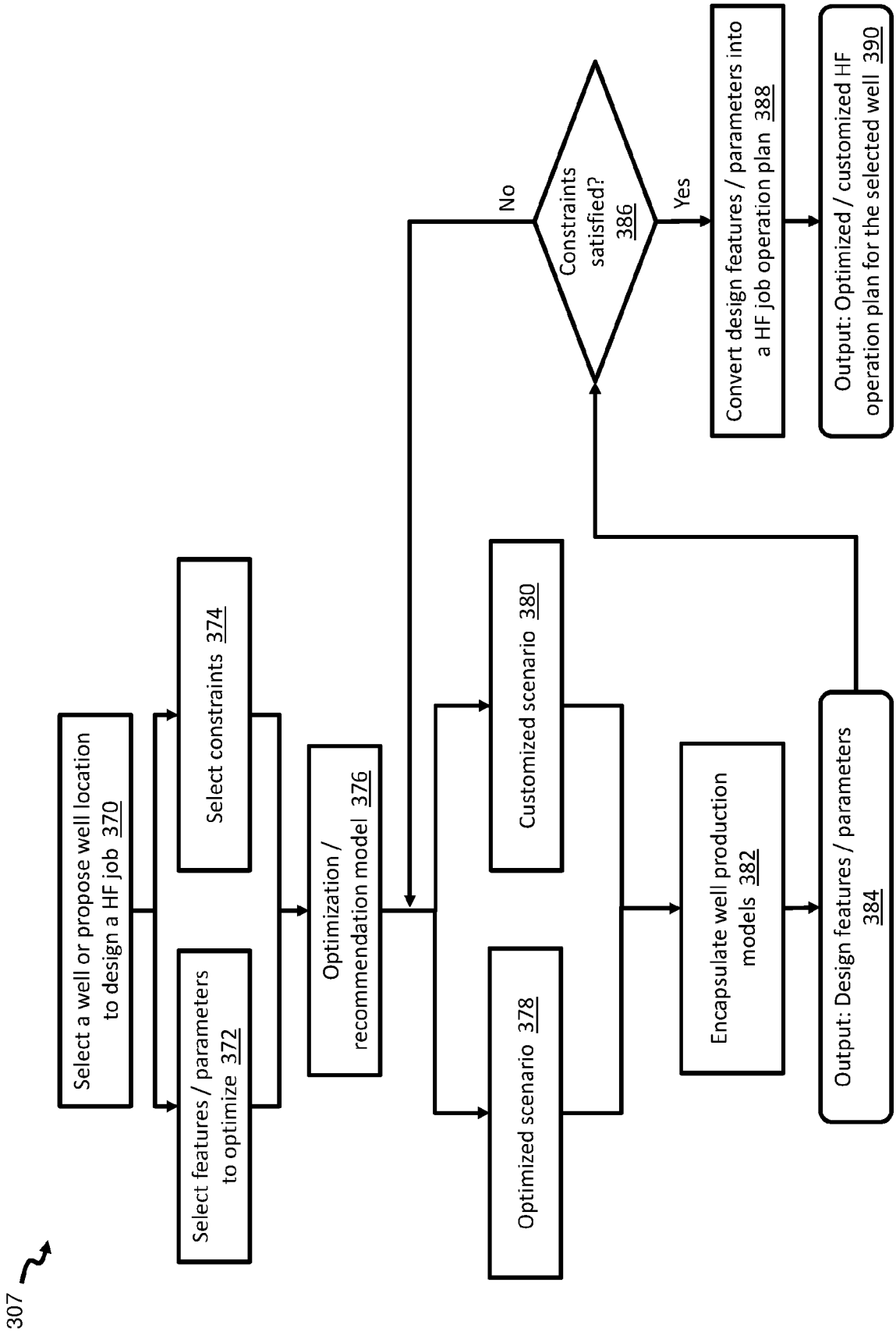
307
Select a well or propose well location to design a HF job 370
Select features / parameters to optimize 372
Select constraints 374
Optimization / recommendation model 376
Optimized scenario 378
Customized scenario 380
Encapsulate well production models 382
Output: Design features / parameters 384
Constraints satisfied? 386
No
Yes
Convert design features / parameters into a HF job operation plan 388
Output: Optimized / customized HF operation plan for the selected well 390

FIG. 3E

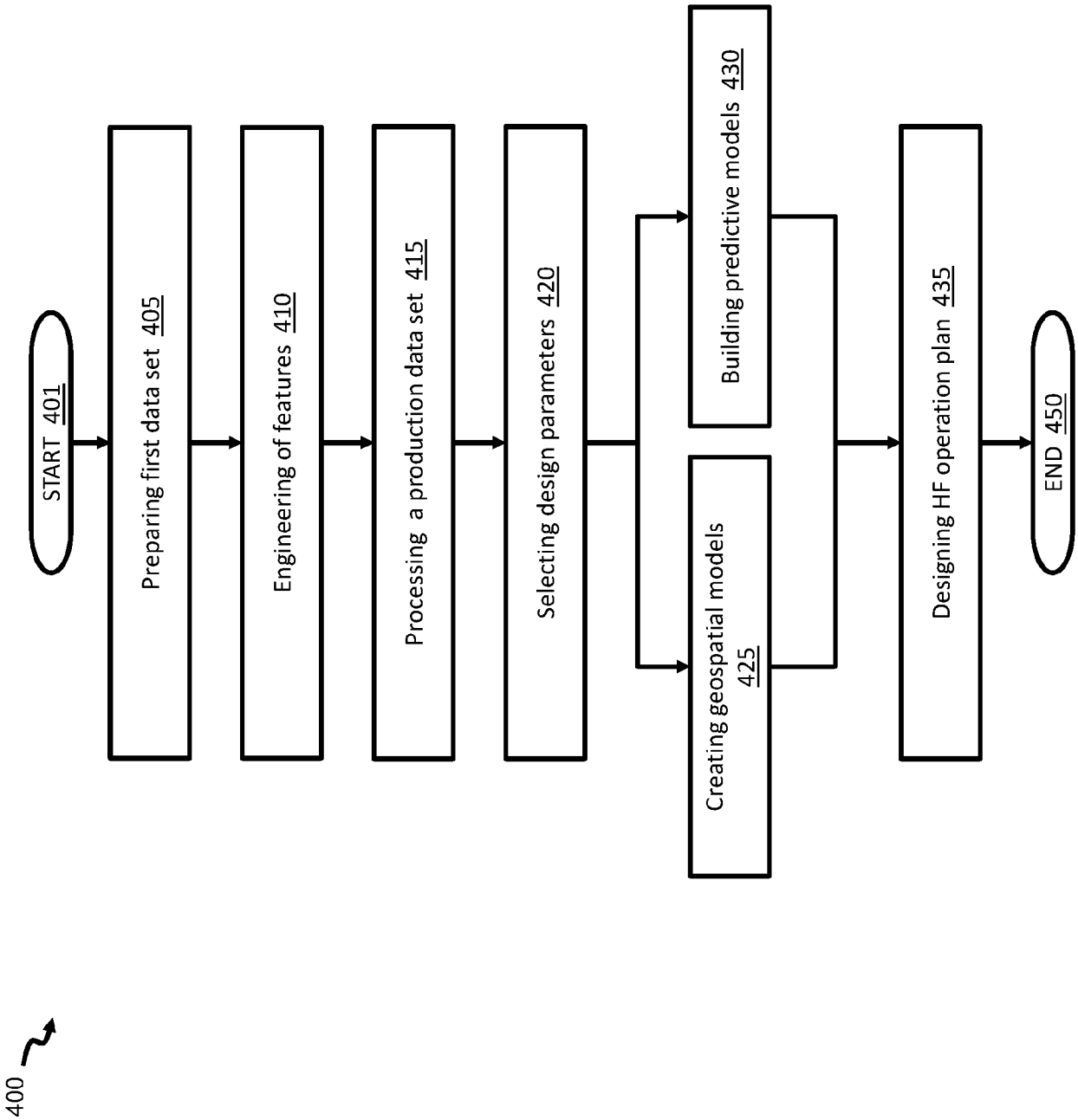
400
START 401
Preparing first data set 405
Engineering of features 410
Processing a production data set 415
Selecting design parameters 420
Creating geospatial models 425
Building predictive models 430
Designing HF operation plan 435
END 450

FIG. 4

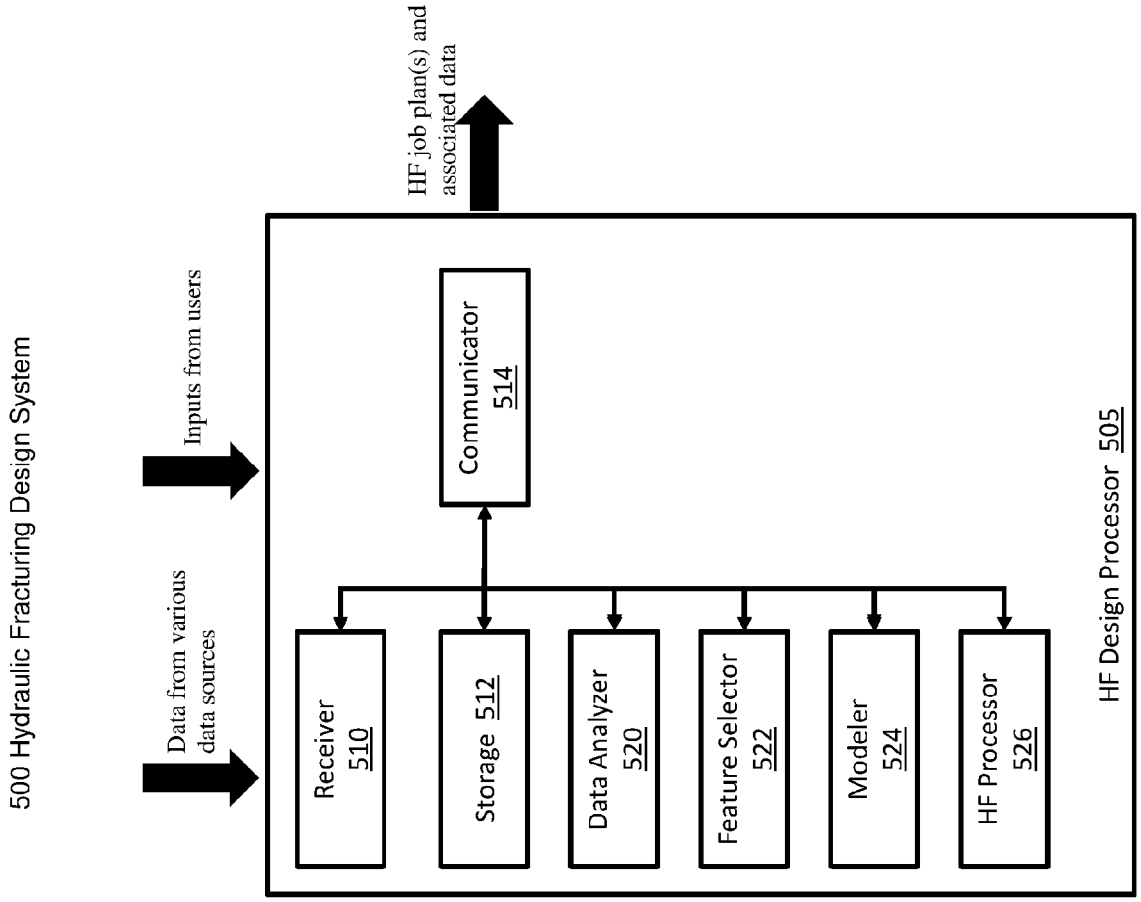
500 Hydraulic Fracturing Design System
Data from various data sources
Inputs from users
Receiver 510
Storage 512
Communicator 514
Data Analyzer 520
Feature Selector 522
Modeler 524
HF Processor 526
HF Design Processor 505
HF job plan(s) and associated data

FIG. 5

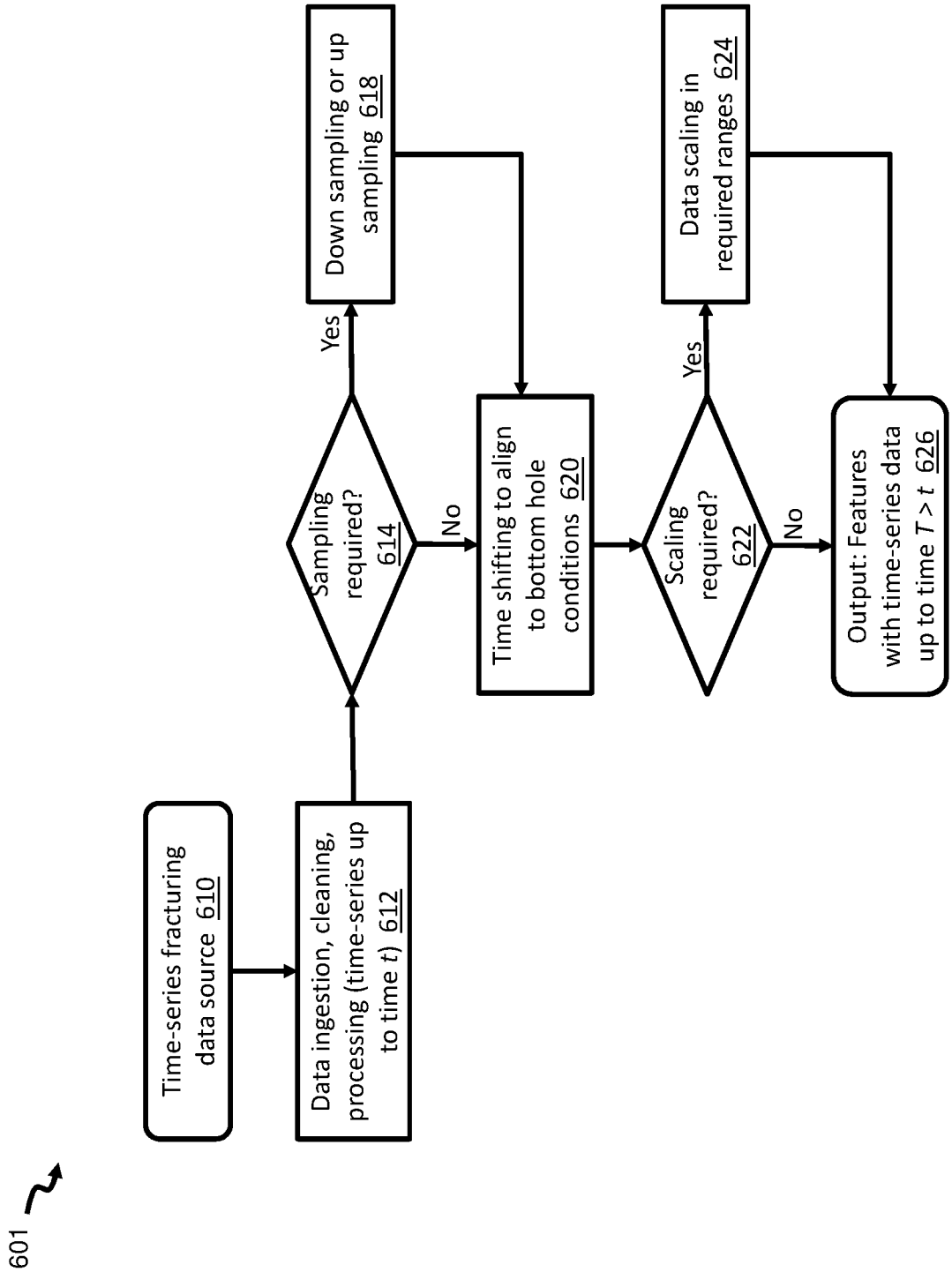
601
Time-series fracturing data source 610
Data ingestion, cleaning, processing (time-series up to time t) 612
Sampling required? 614
Yes
Down sampling or up sampling 618
No
Time shifting to align to bottom hole conditions 620
Scaling required? 622
Yes
Data scaling in required ranges 624
No
Output: Features with time-series data up to time T > t 626

FIG. 6A

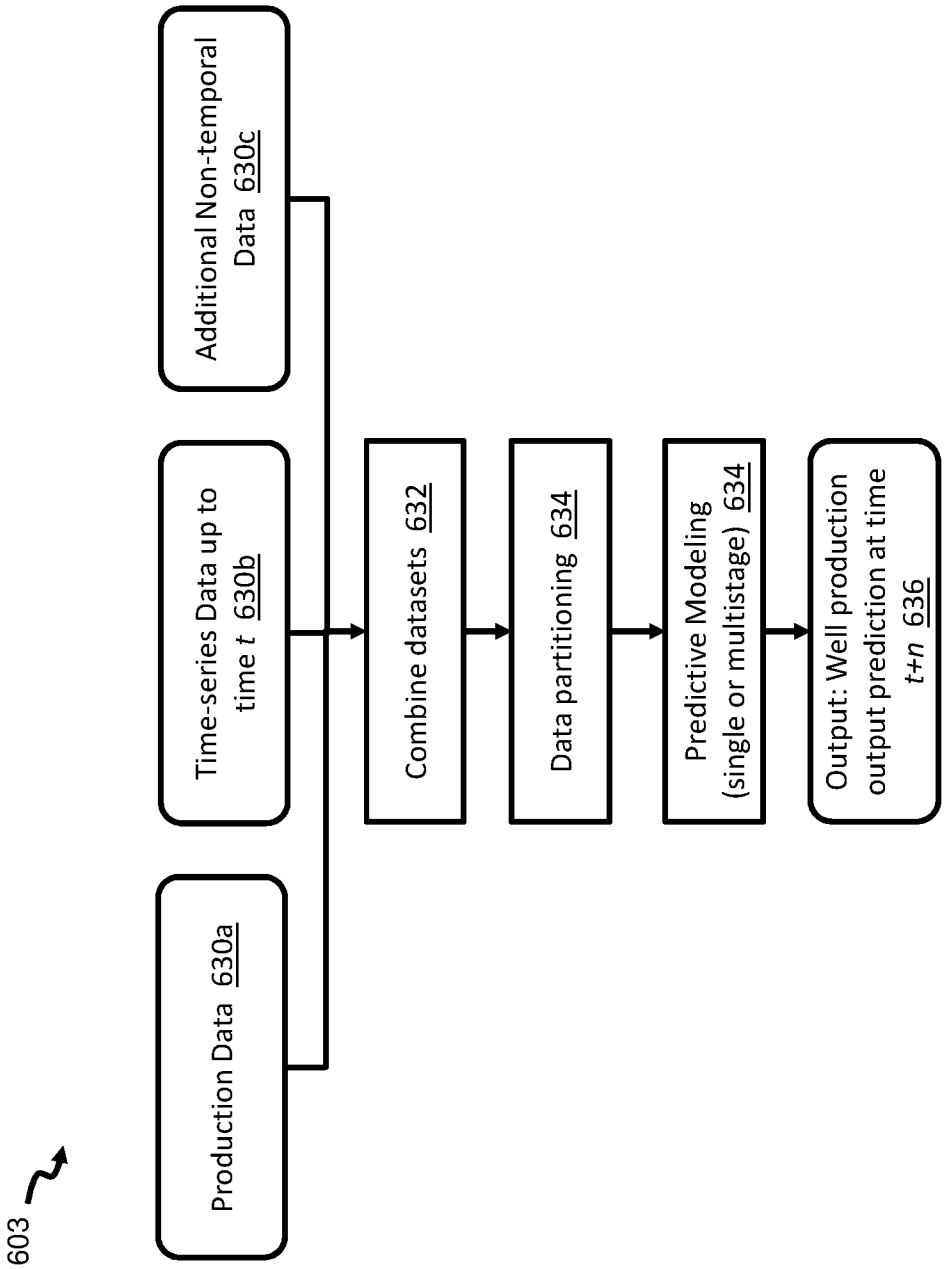
603
Production Data 630a
Time-series Data up to time t 630b
Additional Non-temporal Data 630c
Combine datasets 632
Data partitioning 634
Predictive Modeling (single or multistage) 634
Output: Well production output prediction at time t+n 636

FIG. 6C

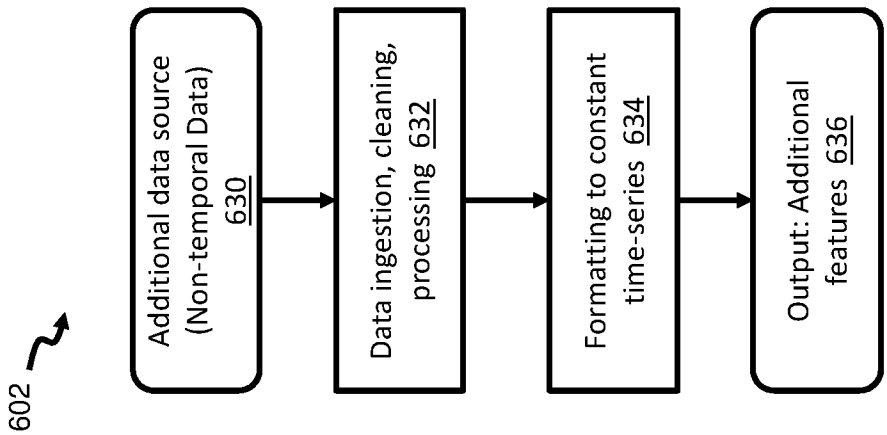
602
Additional data source (Non-temporal Data) 630
Data ingestion, cleaning, processing 632
Formatting to constant time-series 634
Output: Additional features 636

FIG. 6B

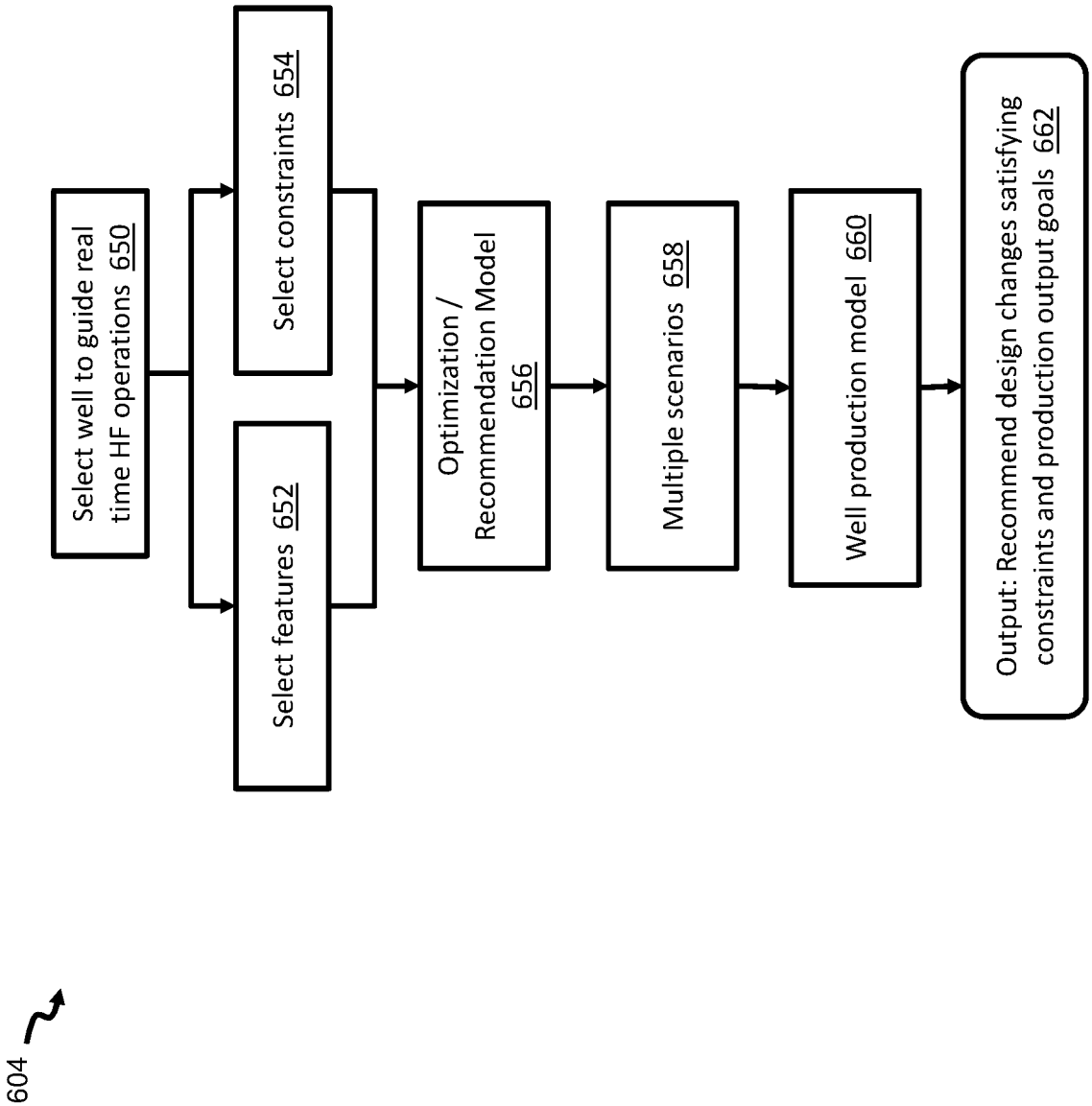
604
Select well to guide real time HF operations 650
Select features 652
Select constraints 654
Optimization / Recommendation Model 656
Multiple scenarios 658
Well production model 660
Output: Recommend design changes satisfying constraints and production output goals 662

FIG. 6D

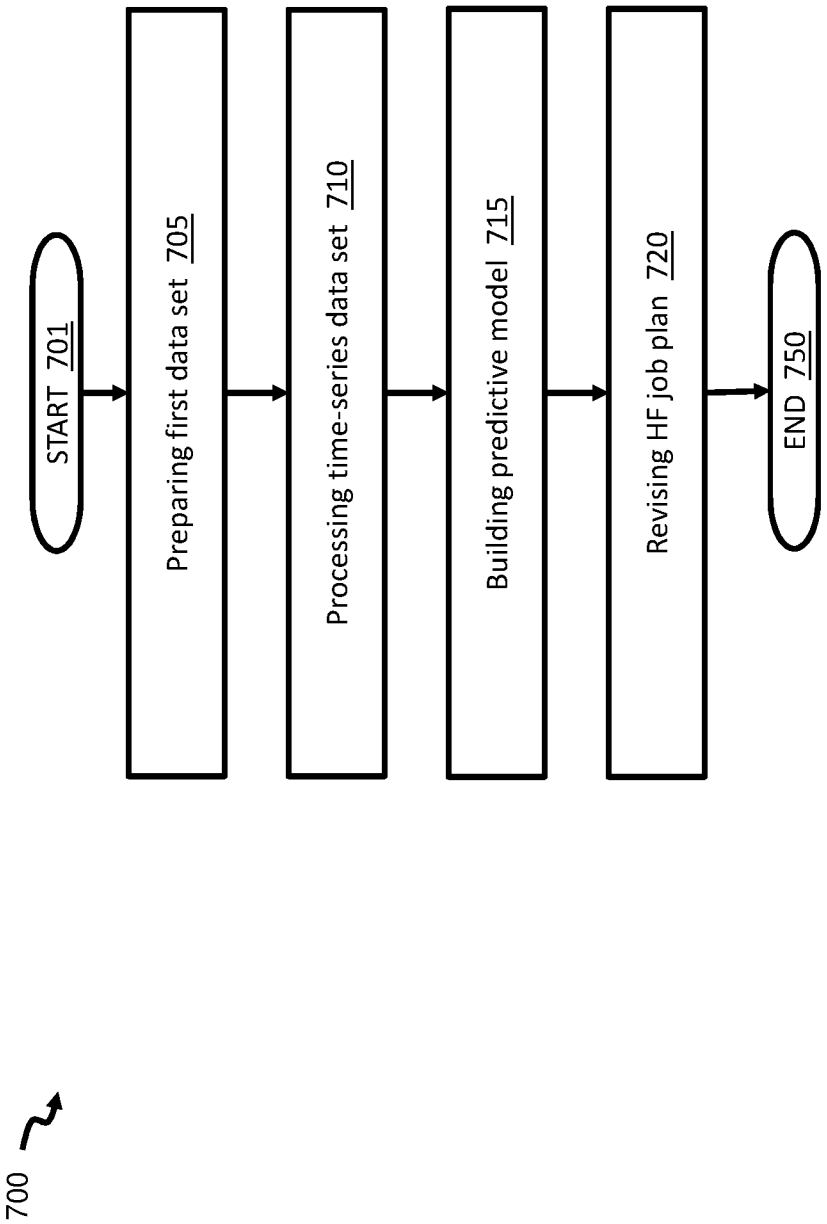
700
START 701
Preparing first data set 705
Processing time-series data set 710
Building predictive model 715
Revising HF job plan 720
END 750

FIG. 7

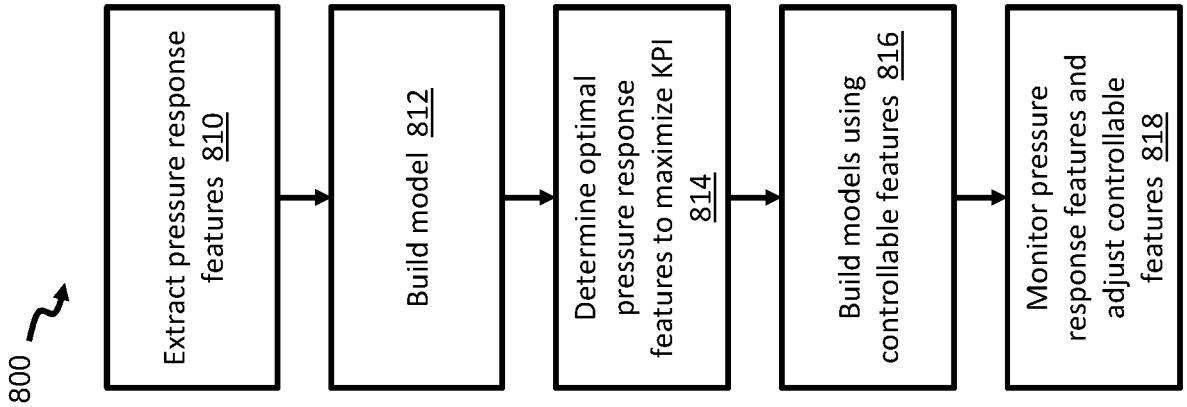

FIG. 8A

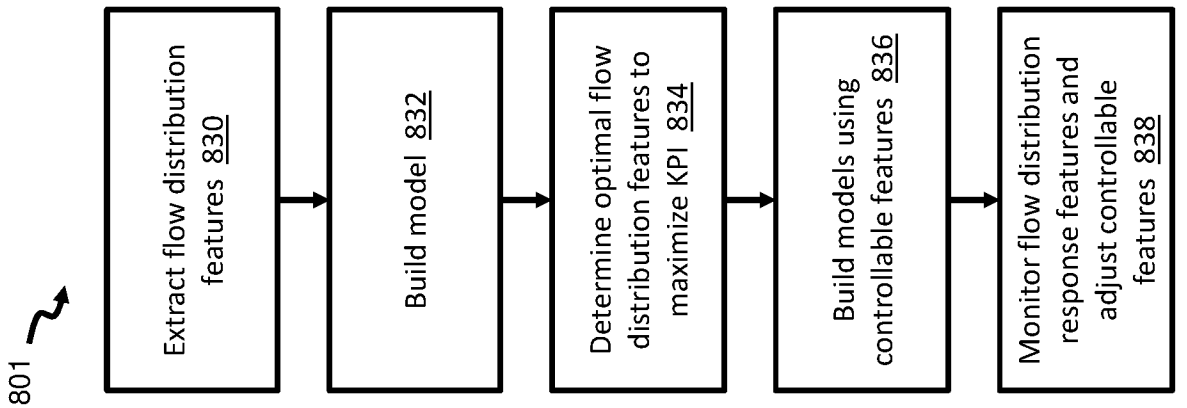

FIG. 8B

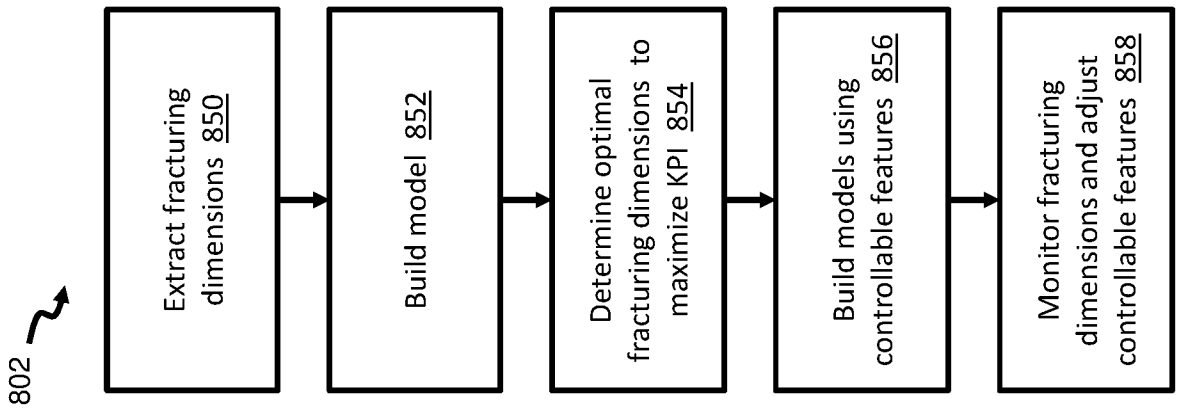

FIG. 8C

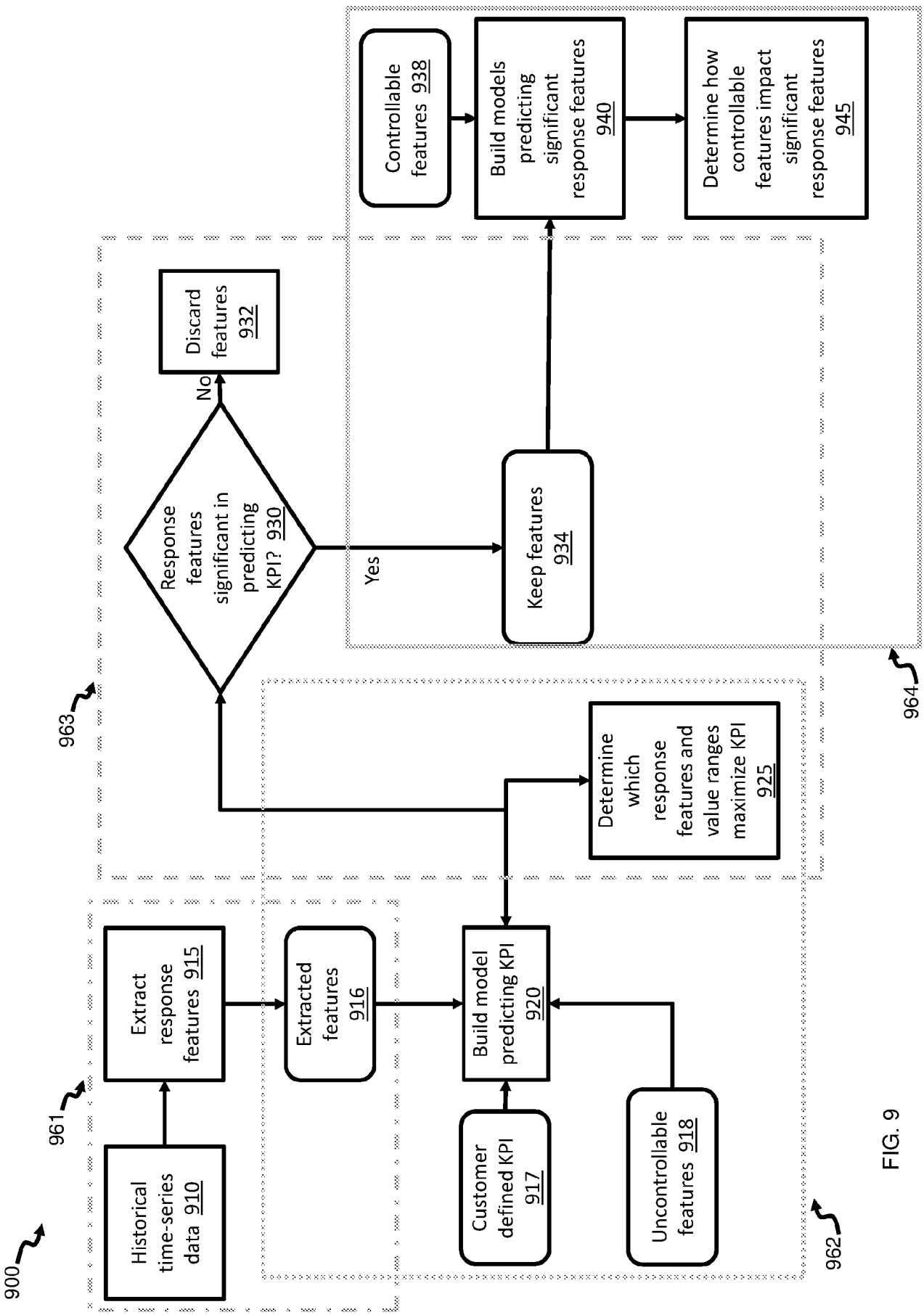
900
Historical time-series data 910
961
Extract response features 915
Extracted features 916
Customer defined KPI 917
Build model predicting KPI 920
Uncontrollable features 918
962
963
Response features significant in predicting KPI? 930
No
Discard features 932
Yes
Keep features 934
Determine which response features and value ranges maximize KPI 925
964
Controllable features 938
Build models predicting significant response features 940
Determine how controllable features impact significant response features 945

FIG. 9

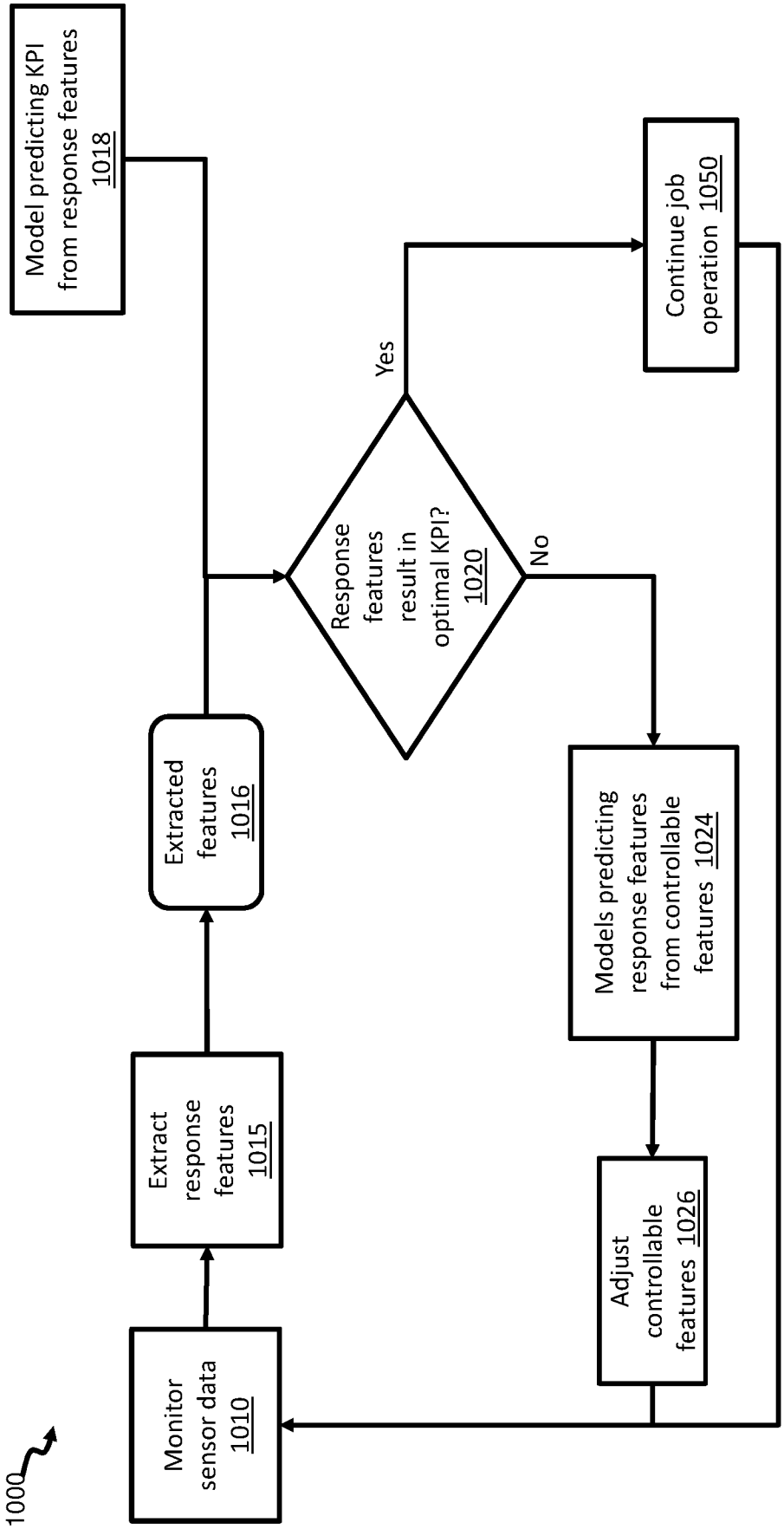
1000
Monitor sensor data 1010
Extract response features 1015
Extracted features 1016
Model predicting KPI from response features 1018
Response features result in optimal KPI? 1020
Yes
No
Continue job operation 1050
Models predicting response features from controllable features 1024
Adjust controllable features 1026

FIG. 10

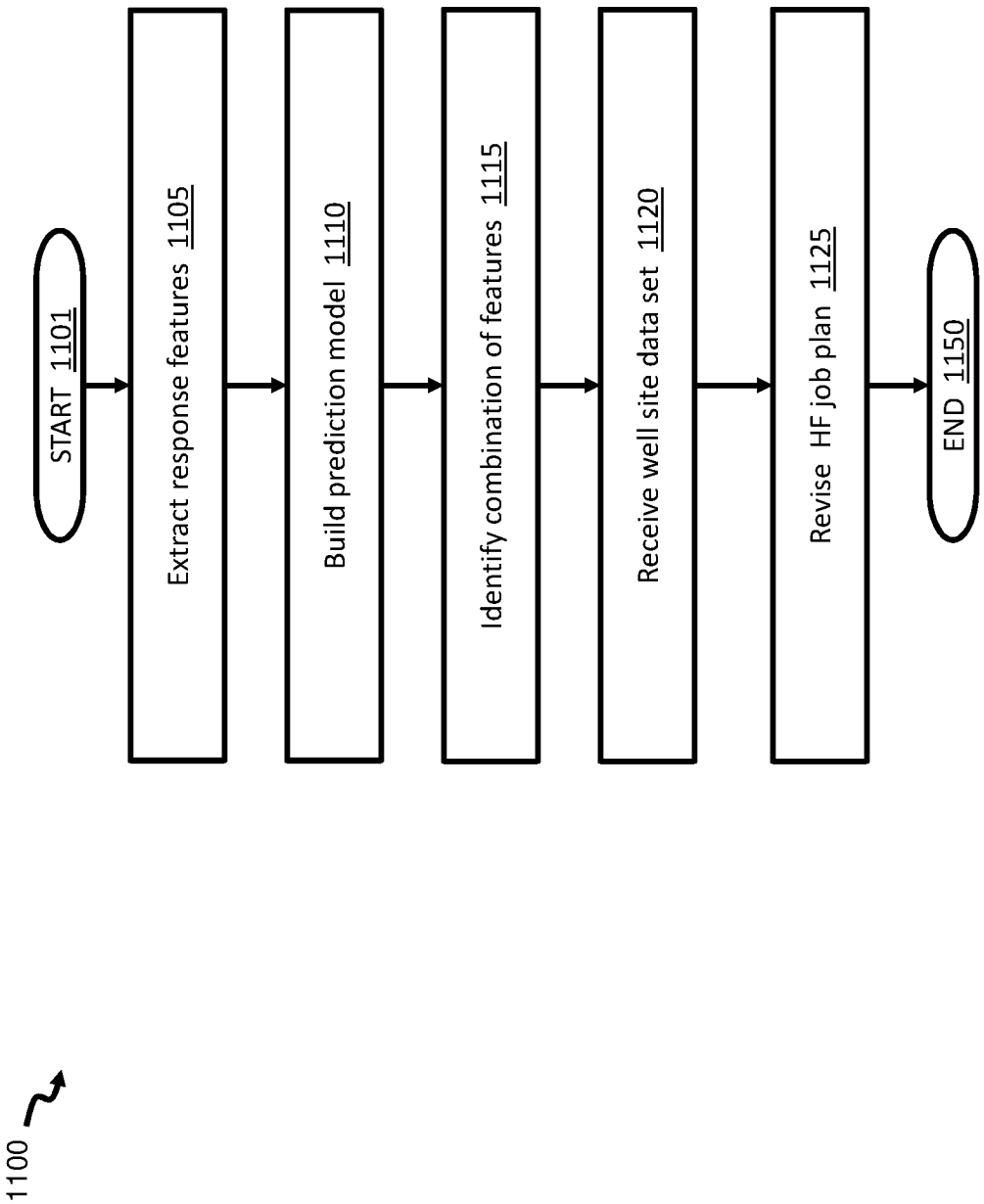
1100
START 1101
Extract response features 1105
Build prediction model 1110
Identify combination of features 1115
Receive well site data set 1120
Revise HF job plan 1125
END 1150

FIG. 11

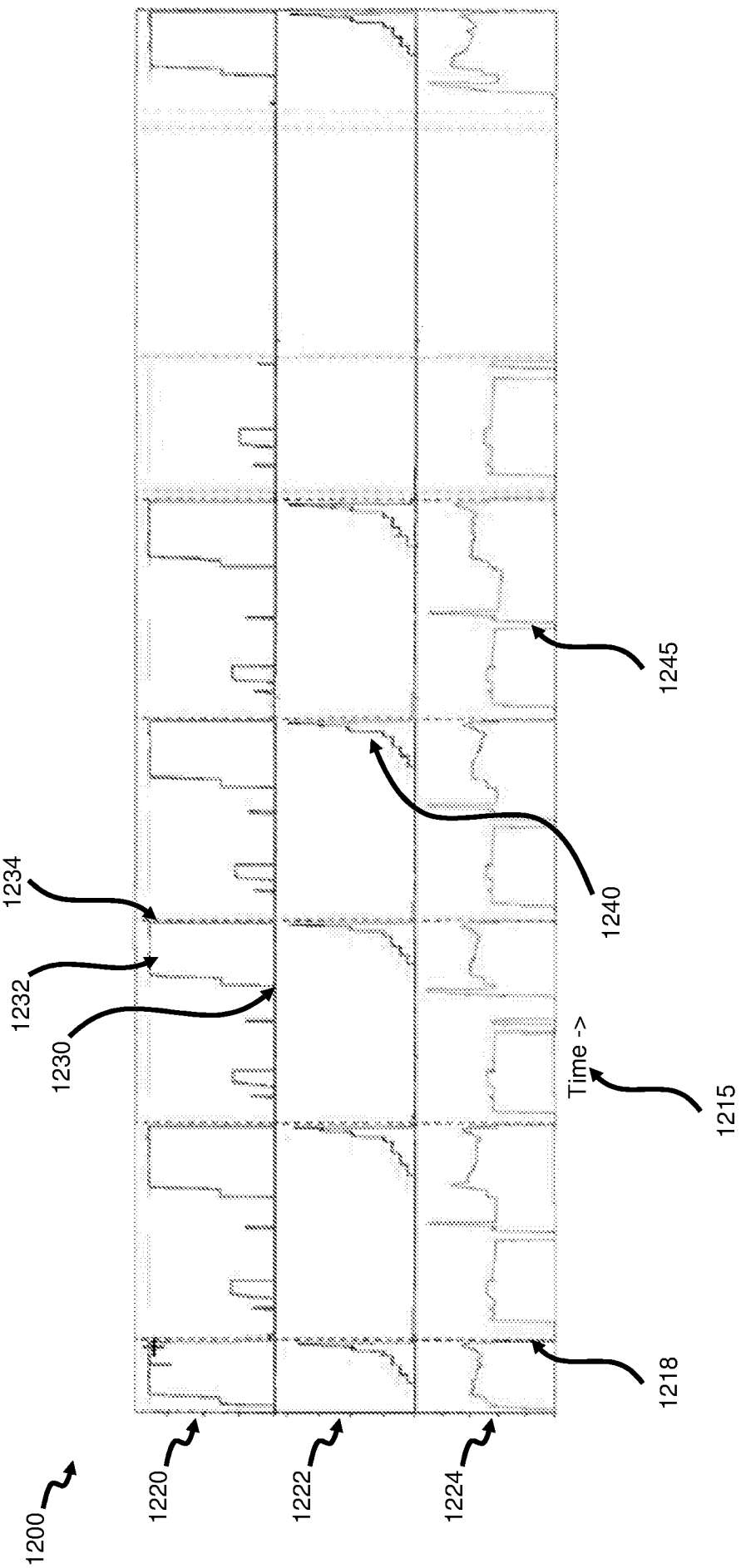

FIG. 12A

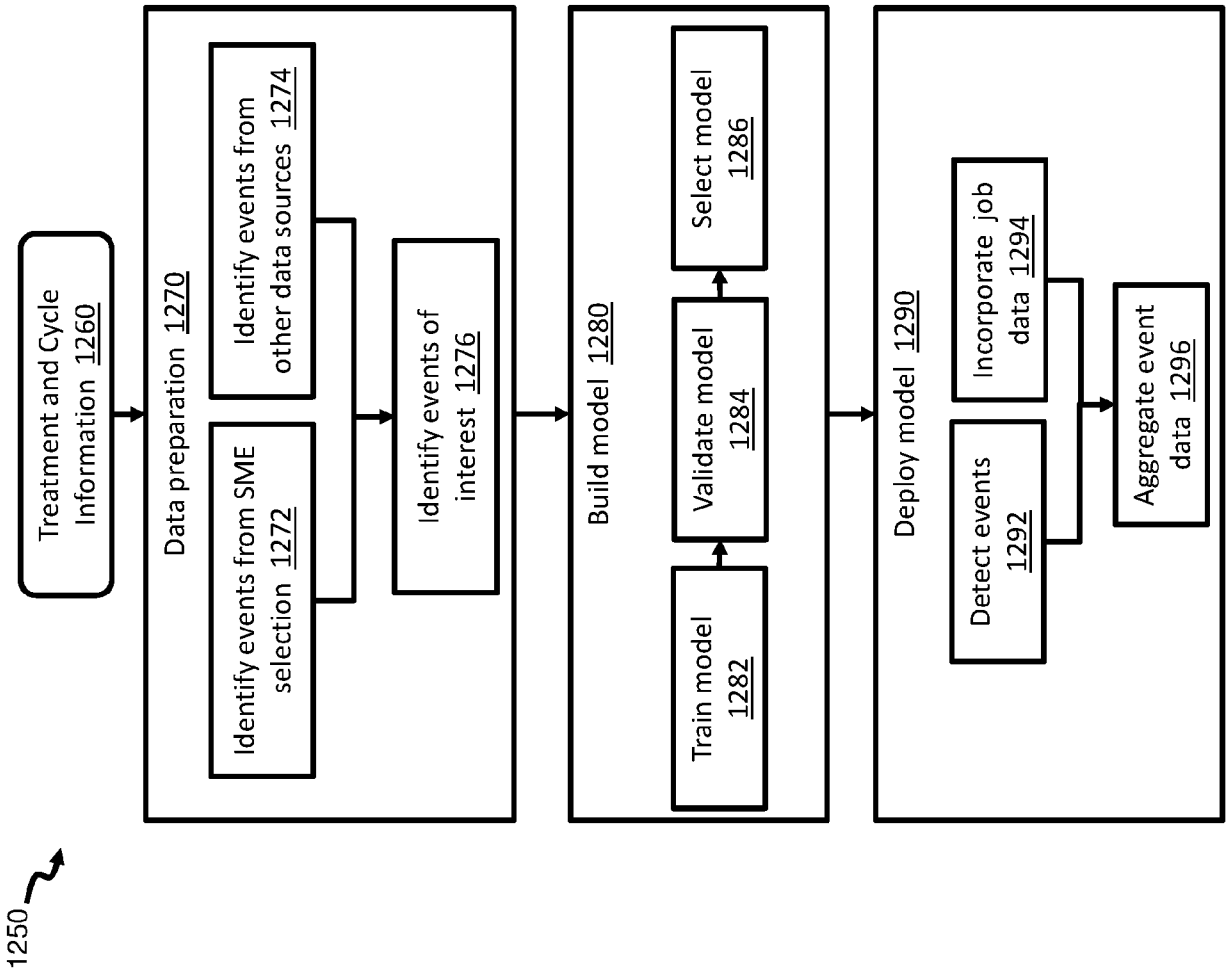
1250
Treatment and Cycle Information 1260
Data preparation 1270
Identify events from SME selection 1272
Identify events from other data sources 1274
Identify events of interest 1276
Build model 1280
Train model 1282
Validate model 1284
Select model 1286
Deploy model 1290
Detect events 1292
Incorporate job data 1294
Aggregate event data 1296

FIG. 12B

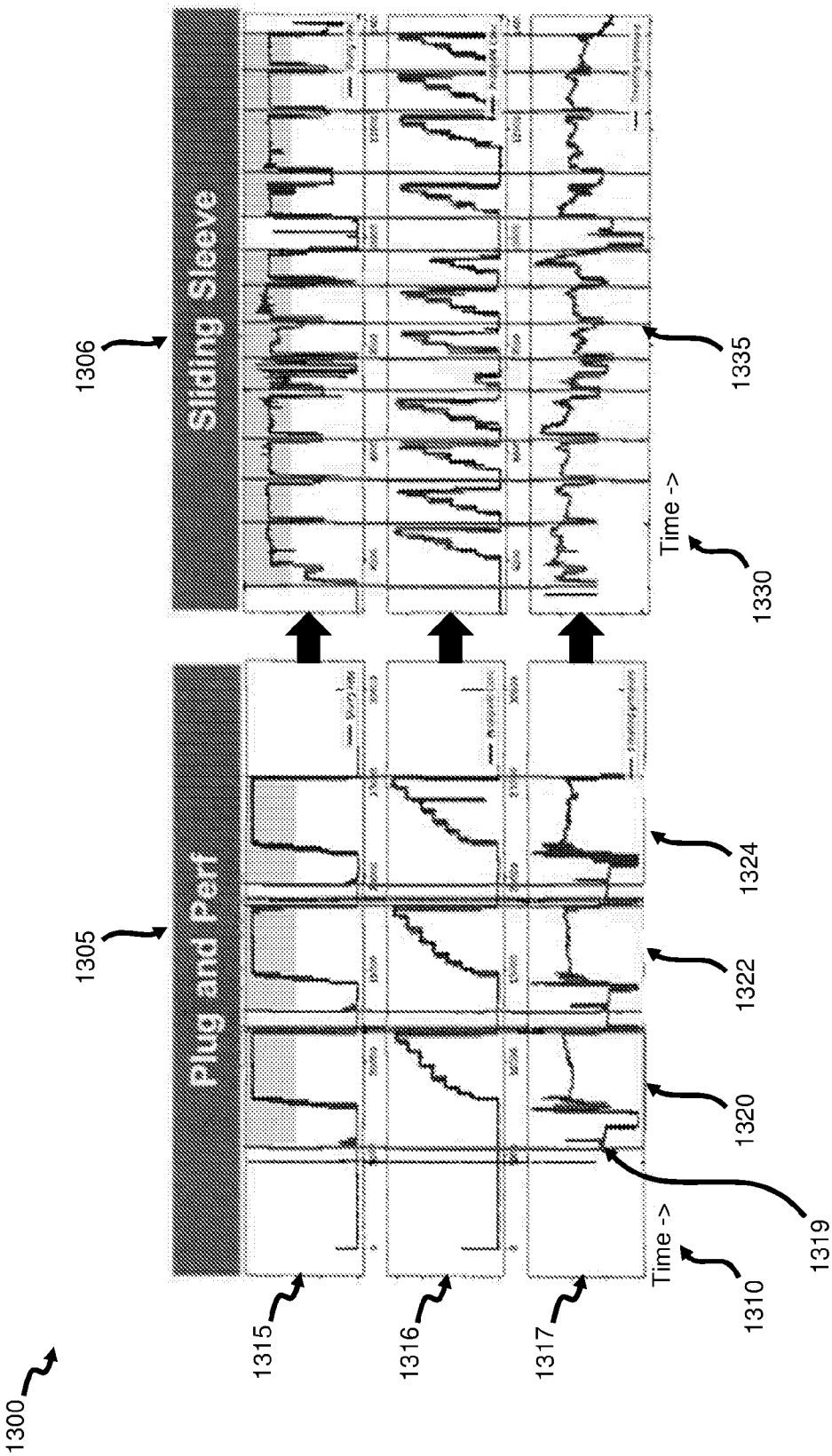
1300
1315
1316
1317
Plug and Perf
1305
Time ->
1310
1319
1320
1322
1324
Sliding Sleeve
1306
Time ->
1330
1335

FIG. 13

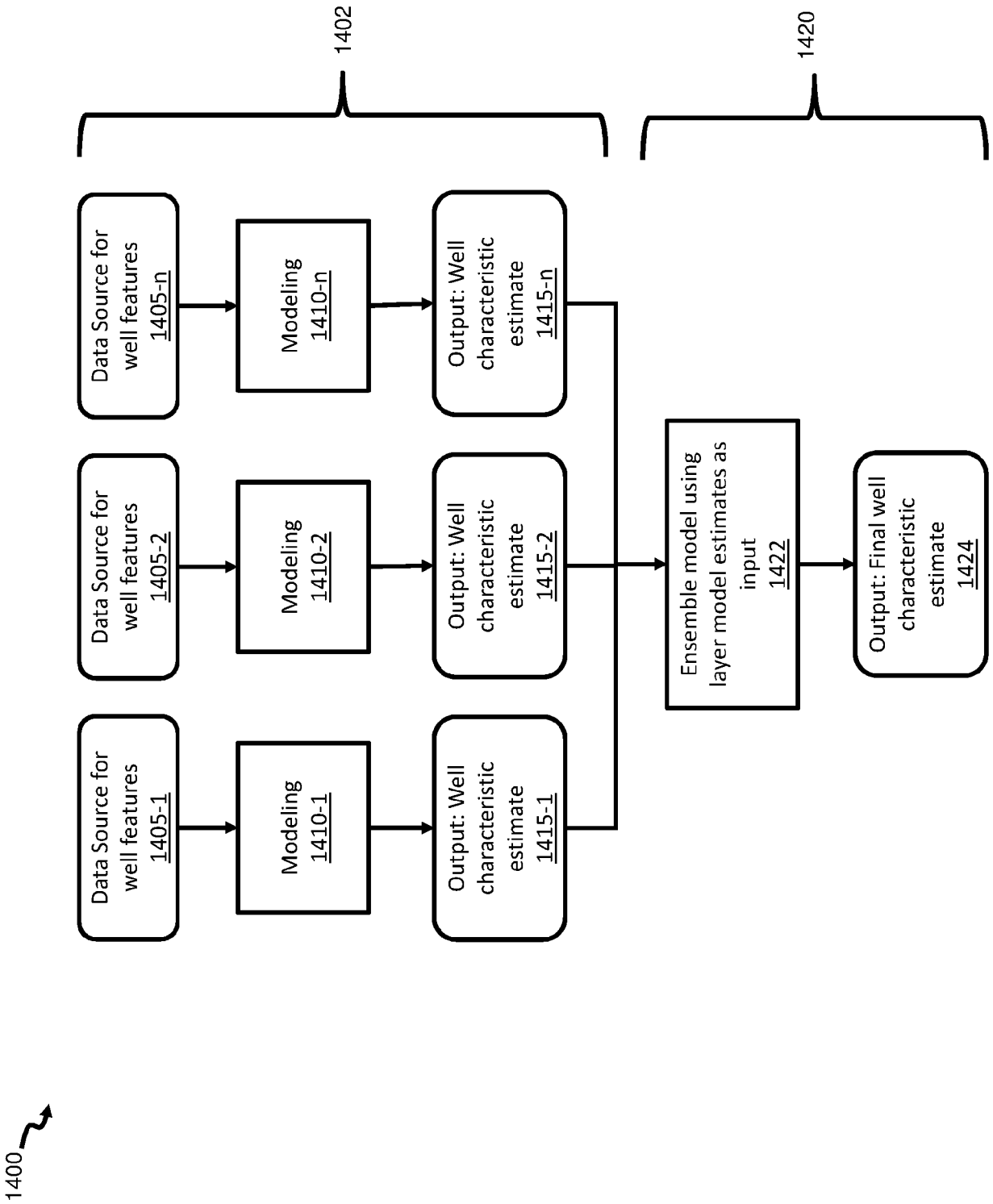
1400
1402
1420
Data Source for well features 1405-1
Data Source for well features 1405-2
Data Source for well features 1405-n
Modeling 1410-1
Modeling 1410-2
Modeling 1410-n
Output: Well characteristic estimate 1415-1
Output: Well characteristic estimate 1415-2
Output: Well characteristic estimate 1415-n
Ensemble model using layer model estimates as input 1422
Output: Final well characteristic estimate 1424

FIG. 14

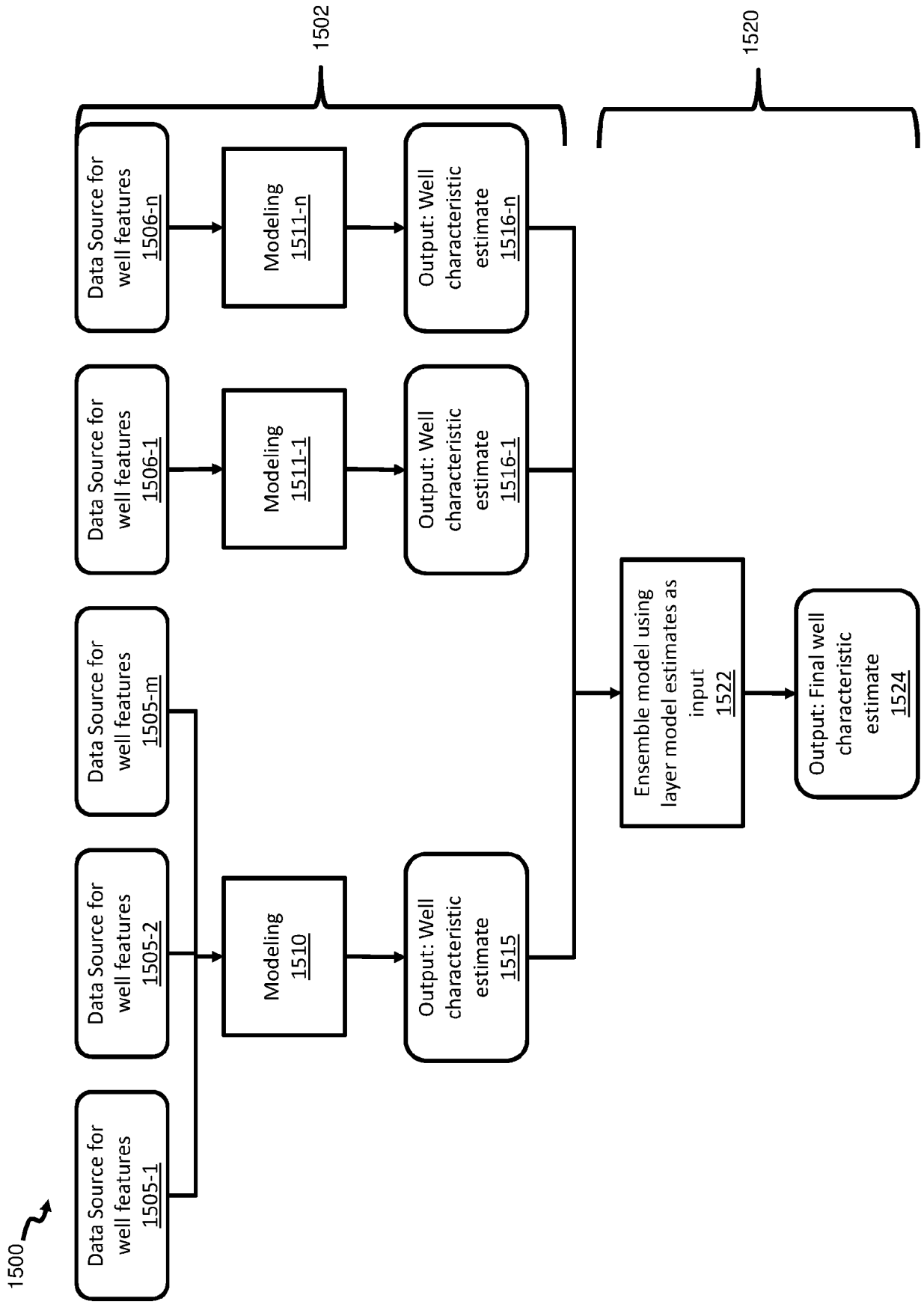
1500
1502
1520
Data Source for well features 1505-1
Data Source for well features 1505-2
Data Source for well features 1505-m
Data Source for well features 1506-1
Data Source for well features 1506-n
Modeling 1510
Modeling 1511-1
Modeling 1511-n
Output: Well characteristic estimate 1515
Output: Well characteristic estimate 1516-1
Output: Well characteristic estimate 1516-n
Ensemble model using layer model estimates as input 1522
Output: Final well characteristic estimate 1524

FIG. 15

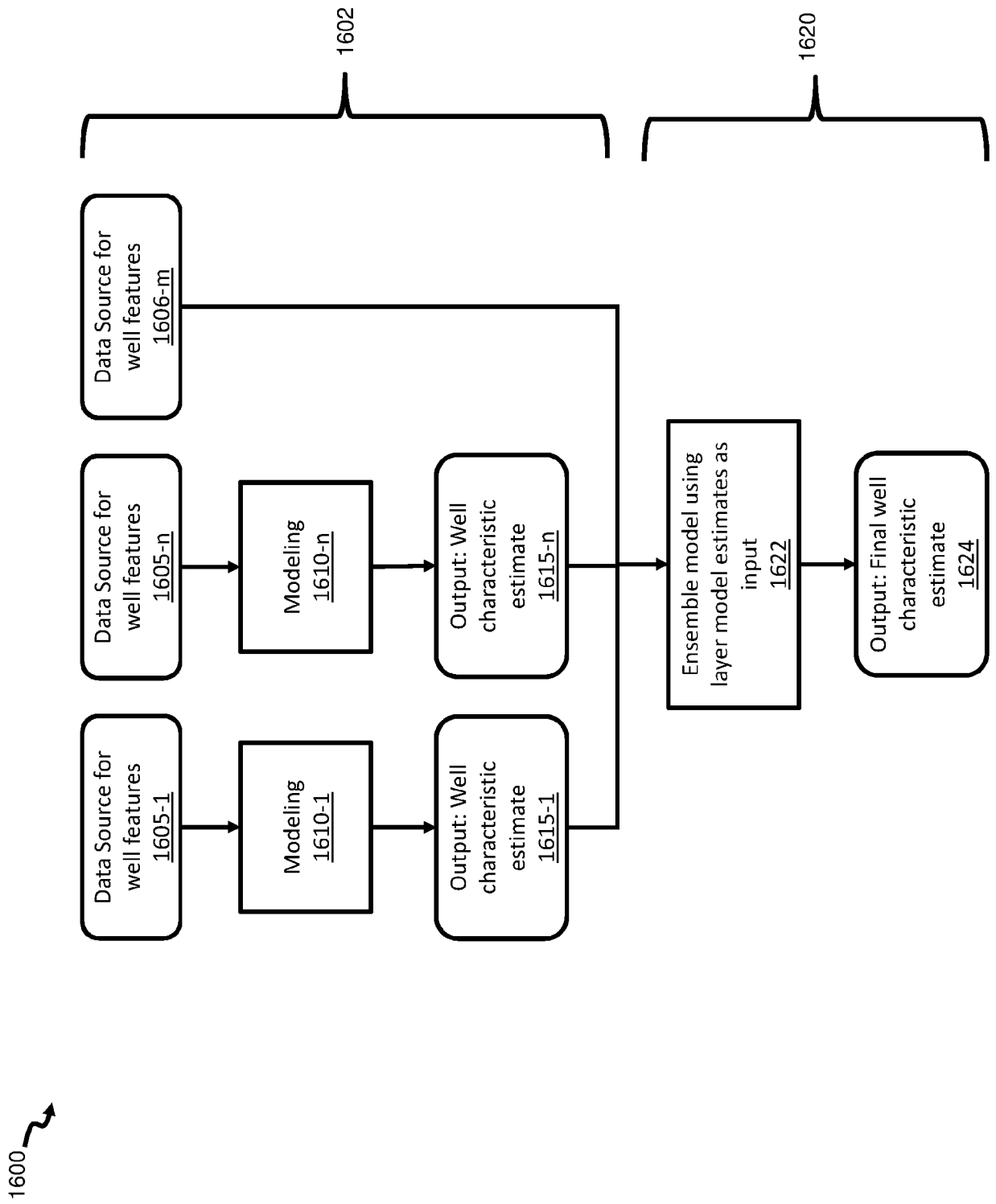
1600
1602
1620
Data Source for well features 1605-1
Data Source for well features 1605-n
Data Source for well features 1606-m
Modeling 1610-1
Modeling 1610-n
Output: Well characteristic estimate 1615-1
Output: Well characteristic estimate 1615-n
Ensemble model using layer model estimates as input 1622
Output: Final well characteristic estimate 1624

FIG. 16

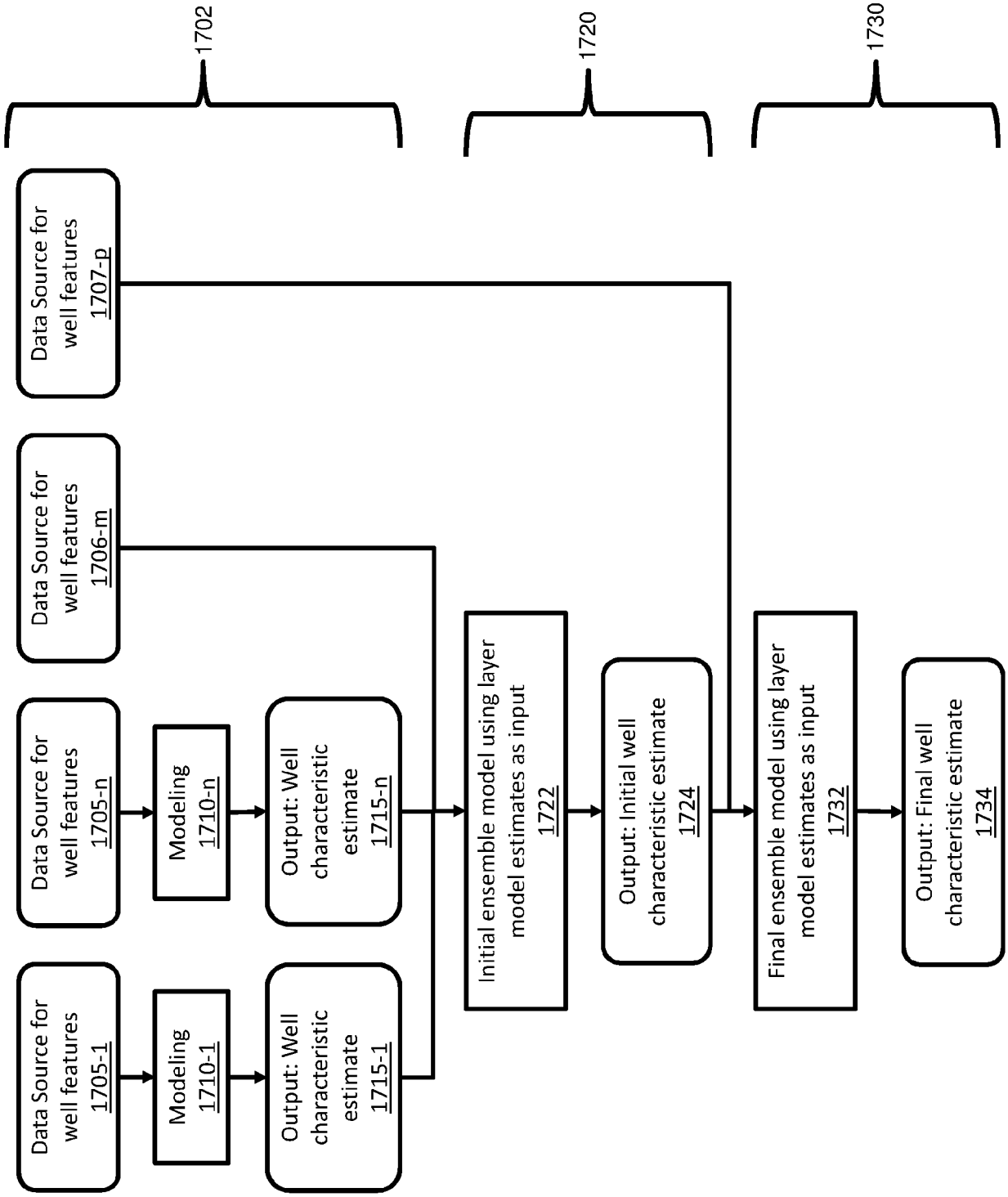
1700
1702
1720
1730
Data Source for well features 1705-1
Data Source for well features 1705-n
Data Source for well features 1706-m
Data Source for well features 1707-p
Modeling 1710-1
Modeling 1710-n
Output: Well characteristic estimate 1715-1
Output: Well characteristic estimate 1715-n
Initial ensemble model using layer model estimates as input 1722
Output: Initial well characteristic estimate 1724
Final ensemble model using layer model estimates as input 1732
Output: Final well characteristic estimate 1734

FIG. 17

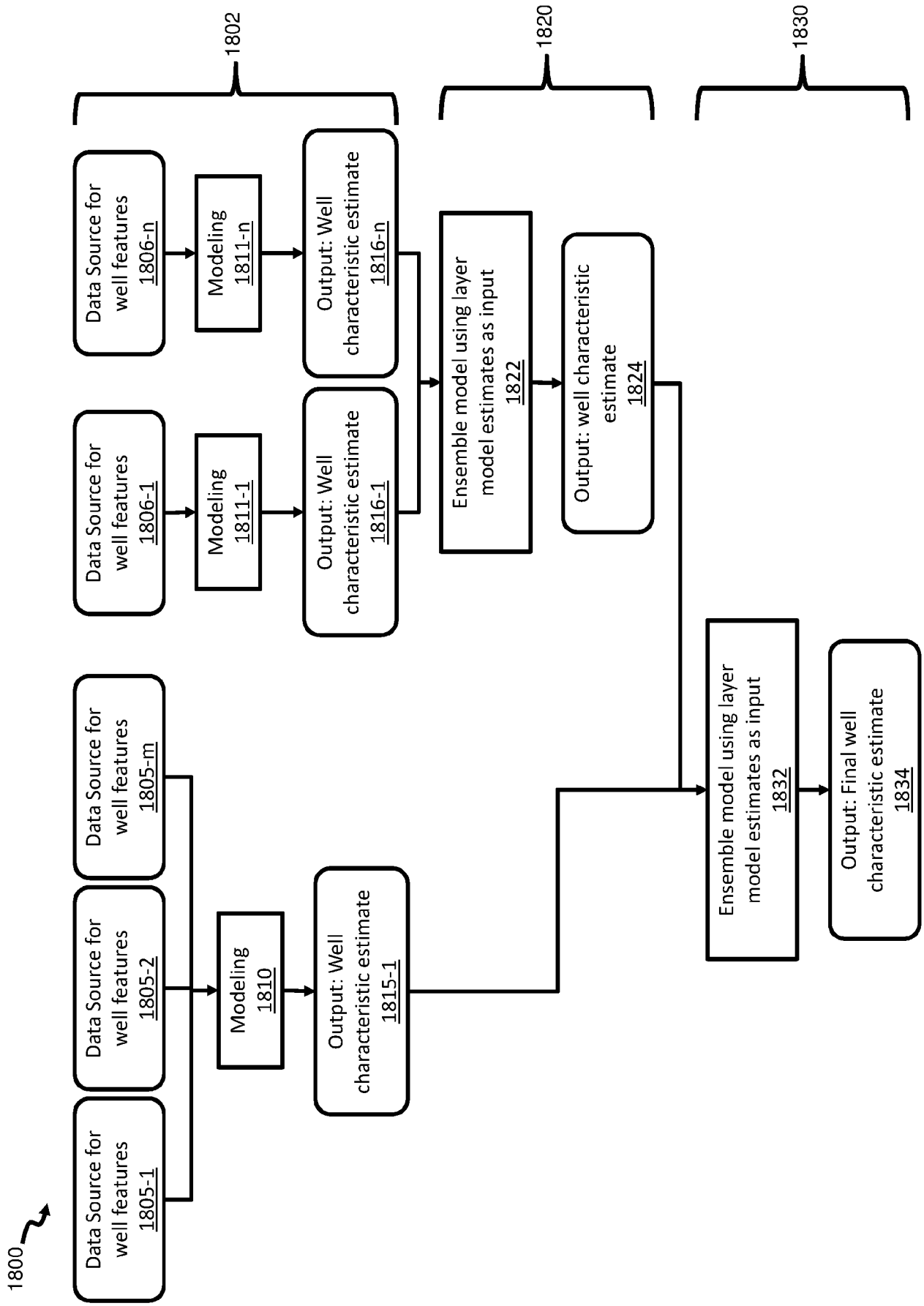
1800
1802
1820
1830
Data Source for well features 1805-1
Data Source for well features 1805-2
Data Source for well features 1805-m
Modeling 1810
Output: Well characteristic estimate 1815-1
Data Source for well features 1806-1
Modeling 1811-1
Output: Well characteristic estimate 1816-1
Data Source for well features 1806-n
Modeling 1811-n
Output: Well characteristic estimate 1816-n
Ensemble model using layer model estimates as input 1822
Output: well characteristic estimate 1824
Ensemble model using layer model estimates as input 1832
Output: Final well characteristic estimate 1834

FIG. 18

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2018/067688

A. CLASSIFICATION OF SUBJECT MATTER

E21B 43/26(2006.01)i, E21B 41/00(2006.01)i, G06N 20/00(2019.01)i, G06N 3/08(2006.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
E21B 43/26; E21B 21/08; E21B 41/00; E21B 44/00; G01V 1/28; G01V 1/50; G05B 13/04; G06F 17/18; G06N 3/02; G06N 20/00; G06N 3/08

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
Korean utility models and applications for utility models
Japanese utility models and applications for utility models

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
eKOMPASS(KIPO internal) & keywords: hydraulic fracturing, final first data, time-series pumping data, key performance indicator, prediction model

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2018-0335538 A1 (SCHLUMBERGER TECHNOLOGY CORPORATION) 22 November 2018 See paragraphs [0052]-[0059], [0065], [0069], [0073]-[0076], [0128], [0131] and figures 3-4.	1-7,9-16,23-25,27-30,33-37
A		8,17-22,26,31-32
A	US 2018-0259668 A1 (HALLIBURTON ENERGY SERVICES, INC.) 13 September 2018 See paragraphs [0041]-[0071] and figure 2.	1-37
A	US 10012748 B2 (SCHLUMBERGER TECHNOLOGY CORPORATION) 03 July 2018 See column 6, lines 3-54 and figure 3.	1-37
A	US 2017-0328179 A1 (HALLIBURTON ENERGY SERVICES, INC.) 16 November 2017 See claim 1 and figure 10.	1-37
A	US 2016-0312552 A1 (EARLY et al.) 27 October 2016 See claim 1.	1-37

☐ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"D" document cited by the applicant in the international application
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents,such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
26 September 2019 (26.09.2019)	**27 September 2019 (27.09.2019)**

Name and mailing address of the ISA/KR	Authorized officer
International Application Division Korean Intellectual Property Office 189 Cheongsa-ro, Seo-gu, Daejeon, 35208, Republic of Korea	PARK, Tae Wook
Facsimile No. +82-42-481-8578	Telephone No. +82-42-481-3405

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.

PCT/US2018/067688

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2018-0335538 A1	22/11/2018	EP 3406844 A1	28/11/2018
US 2018-0259668 A1	13/09/2018	WO 2017-074869 A1	04/05/2017
US 10012748 B2	03/07/2018	MX 2014009200 A	27/05/2015
		MX 359360 B	26/09/2018
		US 2015-0039544 A1	05/02/2015
US 2017-0328179 A1	16/11/2017	WO 2016-108872 A1	07/07/2016
US 2016-0312552 A1	27/10/2016	US 10345764 B2	09/07/2019
		US 10370941 B2	06/08/2019
		US 2016-0312608 A1	27/10/2016
		US 2017-0177761 A1	22/06/2017
		WO 2017-106513 A1	22/06/2017