RESEARCH ARTICLE

Derivative-Free Optimization With Proxy Models for Oil Production Platforms Sharing a Subsea Gas Network

CARLOS LUGUESI[1], EDUARDO CAMPONOGARA[1], (Member, IEEE), LAIO ORIEL SEMAN[1,2], JOSÉ TORREBLANCA GONZÁLEZ[3], AND VALDERI REIS QUIETINHO LEITHARDT[4,5], (Senior Member, IEEE)

[1]Department of Automation and Systems Engineering, Federal University of Santa Catarina, Florianópolis 88040-900, Brazil
[2]Graduate Program in Applied Computer Science, University of Vale do Itajaí (UNIVALI), Itajaí 88302-901, Brazil
[3]Expert Systems and Applications Laboratory, ETSII of Béjar (University of Salamanca), 37008 Salamanca, Spain
[4]COPELABS, University Lusófona (ULHT), 1749-024 Lisbon, Portugal
[5]VALORIZA—Research Centre for Endogenous Resource Valorization, Polytechnic Institute of Portalegre, 7301-110 Portalegre, Portugal

Corresponding author: Laio Oriel Seman (laio@univali.br)

ABSTRACT The deployment of offshore platforms for the extraction of oil and gas from subsea reservoirs presents unique challenges, particularly when multiple platforms are connected by a subsea gas network. In the Santos basin, the aim is to maximize oil production while maintaining safe and sustainable levels of CO_2 content and pressure in the gas stream. To address these challenges, a novel methodology has been proposed that uses boundary conditions to coordinate the use of shared resources among the platforms. This approach decouples the optimization of oil production in platforms from the coordination of shared resources, allowing for more efficient and effective operation of the offshore oilfield. In addition to this methodology, a fast and accurate proxy model has been developed for gas pipeline networks. This model allows for efficient optimization of the gas flow through the network, taking into account the physical and operational constraints of the system. In experiments, the use of the proposed proxy model in tandem with derivative-free optimization algorithms resulted in an average error of less than 5% in pressure calculations, and a processing time that was over up to 1000 times faster than the phenomenological simulator. These results demonstrate the effectiveness and efficiency of the proposed methodology in optimizing oil production in offshore platforms connected by a subsea gas network, while maintaining safe and sustainable levels of CO_2 content and pressure in the gas stream.

INDEX TERMS Oil production systems, subsea gas networks, proxy modeling, derivative-free optimization.

I. INTRODUCTION

One of the main problems faced by vertically integrated oil companies (*i.e.*, companies that control production, transport, storage and refining) is the supply of crude oil from offshore oilfields to refineries, that is, the management of the crude oil supply [1]. For offshore oil production, as in the Brazilian Pre-Salt layer [2], oil companies operate with floating production, storage and offloading vessels

(FPSOs), or simply platforms, to produce and store crude oil. After production, crude oil is then transferred to onshore terminals via submarine pipelines or special vessels equipped with dynamic positioning systems. As pipelines are not a viable alternative for oil transportation in deepwater oilfields, a fleet of vessels is deployed to transfer crude oil to terminals [3].

In addition to oil transportation logistics, such offshore production systems share an underwater gas flow network for transporting the produced gas to onshore terminals [4]. An example is the Santos Basin where several deep water

The associate editor coordinating the review of this manuscript and approving it for publication was Hao Wang.

FPSOs and fixed platforms are in operation. Due to the high content of contaminants in the produced gas, the platforms are equipped with dedicated units for sequestration of the CO_2, which is reinjected into the reservoirs, but a part remains in the exported gas. To meet the restrictions on the maximum concentration of CO_2 at the onshore terminal, gas exports from the Pre-Salt platforms must be coordinated and mixed with gas devoid of contaminants, produced by a limited number of platforms operating in Post-Salt reservoirs. Such restrictions lead to the complex problem of coordinating the production of the platforms, seeking to maximize the total oil production from the multi-reservoir asset while ensuring the CO_2 limits in the produced gas, among others.

A. OVERVIEW AND RELATED WORKS

The production optimization of multiple platforms sharing a gathering network and processing facilities is not a new topic in the literature. According with the multilevel control hierarchy [5], the related works fall in one of two layers depending on the time span of predictions and the nature of the models.

The first layer is the *long-term optimization* which concerns decisions extending over years of operation and which makes use of reservoir models. Regarding the long-term, [6] addressed the optimization of the production rates that maximize the recoverable oil volumes in the context of multiple reservoirs. In [7], the value chain of such oilfields was optimized in the long-run, while managing CO_2 injection sites and considering other shared facilities. In [8], strategies were presented to optimize gas transportation in order to honor contractual agreements. In [9], models were developed to optimize the transportation network of natural gas on the Norwegian Continental Shelf. Oilfield management in long-term operations has also been the focus of derivative-free optimization (DFO) [10], [11], [12], [13]. In [14], for instance, different DFO methods were discussed, including, but not limited to, Hooke-Jeeves direct search, genetic algorithms and particle swarm optimization. In [15], DFO was applied to gas-lift optimization in an integrated environment, including reservoirs, wells and facilities, by means of the Mesh Adaptive Direct Search (MADS) algorithm. In [16], a DFO algorithm based on the trust-region method was applied for robust well control optimization of oil reservoirs under geological uncertainty.

The second layer is the *short-term optimization* which regards the daily and weekly operations and which considers the production infrastructure, such as the submarine flow network and topside processes. Our work belongs to this layer, with a focus on production optimization of multiple offshore platforms that share a subsea gas pipeline network while considering composition constraints, such as limits on the CO_2 content in the gas delivered to onshore terminals. Besides being of academic interest, this problem has practical applications in offshore operations, such as the multi-reservoir oilfields in the Santos Basin. In [4],

a methodology was proposed to coordinate the production from multiple offshore platforms, which share a subsea gas pipeline network and are jointly limited by processing constraints at the onshore facilities. This methodology relies on surrogates models for production platforms and piecewise-linear approximation of fluid flow, yielding a mixed-integer linear programming (MILP) formulation to which standard algorithms can be applied. An application of this MILP-based methodology to the operations in the Santos Basin proved to be effective, but computationally costly. Aiming for locally optimal solutions with the direct use of multiphase flow simulators, an application of DFO algorithms was proposed in [17] for that problem. The DFO approach brings about flexibility and reduces the effort of model synthesis and maintenance. The application of a DFO algorithm to the Santos Basin obtained solutions of nearly-optimal quality, but the computational cost was exceedingly high due to the direct use of a complex phenomenological simulator. In [18], a methodology for hyper-tuning of oil well simulators was presented using DFO. Besides meeting demands from the regulatory agency, accurately adjusted well models empower engineers to plan production and derive piecewise-linear surrogates that can be used in MILP production optimization. In [19], a DFO exact penalty algorithm was proposed and applied to production optimization of an offshore platform that receives the production stream of gas-lifted oil wells. Being an exact penalty method, it can accept infeasible iterates and handle constraints not be given explicitly, allowing constraints to be modeled through the optimization procedure.

This work aims to develop proxy models [20], [21] for gas pipeline networks that enable fast computation of pressures as a function of flows and concentrations of components of interest. The proposed proxy models capture the structure of the governing equations of such networks, while at same time using system identification to yield polynomial models from simulated data and field measurements. As such, the proxy models can relieve algorithms from the need of using simulation software, promoting fast and reliable optimization using DFO algorithms.

B. PROBLEM STATEMENT

The problem being addressed in this paper is the optimization of production in offshore oil platforms that share a subsea gas network. This production system is modeled using a directed graph, with the nodes representing production platforms, terminals, and junctions where material and pressure balance are enforced. The edges in the graph represent pipelines, wells, valves, and equipment such as separators and electric submersible pumps. The set of vertices in the graph is divided into three subsets: source nodes representing production platforms, sink nodes representing terminals, and internal nodes that do not accumulate material. The set of edges is also divided into two subsets: equipment such as chokes and pumps, and pipelines. The goal of the optimization problem

is to maximize oil production, or an economic function, of an entire network of reservoirs and production platforms, while honoring various constraints such as capacity limits and pressure requirements. The oil is offloaded from the production platforms to the onshore terminal by shuttle tankers, whereas the produced gas is transferred by a subsea pipeline network.

In other words, the problem of interest consists in determining the optimal operation of an oilfield in order to maximize overall oil production or profits. This can be formulated as an optimization problem of the form:

$$\max_{\mathbf{x}} f(\mathbf{x})$$
$$\text{s.t. } g_i(\mathbf{x}) \leq 0, \quad i = 1, \ldots, m$$
$$h_j(\mathbf{x}) = 0, \quad j = 1, \ldots, n$$

where $\mathbf{x}$ is the vector of decision variables, $f(\mathbf{x})$ is the objective function representing the profit to be maximized, $g_i(\mathbf{x})$ are the inequality constraints, and $h_j(\mathbf{x})$ are the equality constraints that express physics relations, operational constraints, and limits that characterize the system behavior.

One challenge of the optimization problem being addressed is that the constraints may not be available in analytic form. This can make it difficult to solve the problem using traditional optimization methods, which typically rely on the ability to compute the gradient of the objective function and the constraints. In cases where the constraints are not available in analytic form, it may be necessary to employ alternative optimization approaches such as heuristics or metaheuristics, which do not rely on the availability of gradients. In this specific paper, we rely on DFO formulations for that matter.

C. CONTRIBUTIONS

The key contributions of this work to the state of the art are:

- A methodology based on systems identification for proxy modeling of gas flow networks, which takes into account the structure of the governing equations of gas flow in pipelines.
- An analysis showing that the proposed proxy models are extremely fast to compute and achieve good accuracy, in terms of flows and pressures, with respect to data obtained from a high-fidelity multiphase flow simulator.
- A computational analysis demonstrating that DFO combined with the proxy models is an effective optimization tool for multi-reservoir oilfields sharing a subsea gas network, which can maximize a joint objective and handle simulated constraints.

D. ORGANIZATION

Section II presents a conceptual formulation for production optimization of multiple platforms sharing a subsea gas pipeline network. Section III provides background on the physics of gas flow in pipelines and methodologies for system identification, which are key for the design of proxy

TABLE 1. Sets.

Set	Description
$\mathcal{V}$	Set of vertices in the production network.
$\mathcal{V}^{src}$	Set of source nodes (*i.e.*, platforms). $\mathcal{V}^{src} \subset \mathcal{V}$.
$\mathcal{V}^{snk}$	Set of sink nodes (*i.e.*, terminals). $\mathcal{V}^{snk} \subset \mathcal{V}$.
$\mathcal{V}^{int}$	Set of internal nodes (*i.e.*, manifolds). $\mathcal{V}^{int} = \mathcal{V} \setminus (\mathcal{V}^{src} \cup \mathcal{V}^{snk})$.
$\mathcal{E}$	Set of edges in the production network.
$\mathcal{E}^{pipes}$	Set of edges denoting pipelines.
$\mathcal{E}^{eqp}$	Set of edges denoting equipment. $\mathcal{E}^{eqp} = \mathcal{E} \setminus \mathcal{E}^{pipes}$.
$\mathcal{P}$	Set of phases, typically oil, gas, and water.
$\mathcal{C}_p$	Set of components of interest in phase p, e.g. CO_2.

models. Section IV discusses how to convert the conceptual formulation into a concrete formulation using simulation and proxy-models, which can be optimized with standard algorithms. Section V introduces a production system in the Santos Basin as the case study, and further develops and analyzes proxy-models to approximate the simulation models of gas pipeline networks.

Section V-C reports results from the solution of concrete formulations for the case study with DFO algorithms. Finally, Section VI draws some final remarks.

II. PROBLEM FORMULATION

Based on [4], this section presents a formulation for production optimization of offshore oil platforms that share a subsea gas network. The formulation arises from a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ that models the production system. The nodes represent production platforms, terminals, and junctions where material balance is physically enforced. Connecting two nodes, arcs model pipelines that transfer fluids from source to sink and can also represent wells, valves, and equipment such as separators and electric submersible pumps (ESP) [22]. Because of the graph structure, the formulation becomes flexible and general to model heterogeneous production systems consisting of platforms, wells, subsea and topside facilities.

We follow the standard notation of network-flow models [23] which concerns material balance and specialized works [24], [25] that consider pressure, temperature and energy balance in oil fluid flow. The set of vertices $\mathcal{V}$ is split into three subsets: $\mathcal{V}^{src}$ are source nodes representing production platforms; $\mathcal{V}^{snk}$ are the sink nodes representing terminals; and $\mathcal{V}^{int}$ are the internal nodes that do not accumulate material, such as manifolds. The edge set $\mathcal{E}$ is divided in two subsets: $\mathcal{E}^{eqp}$ denotes equipment such as chokes and pumps; and $\mathcal{E}^{pipes}$ represents pipelines. Table 1 presents the notation for vertex and edge sets of the graph-based model for heterogeneous production networks. Sets are defined in calligraphic typeface, vectors in boldface, and scalar in normal typeface.

A pressure variable p_v and flow vector $\mathbf{q}_v$ are associated to each node $v \in \mathcal{V}$. Flow rates are characterized by a set $\mathcal{P}$ of phases, typically oil, gas, and water, and a set $\mathcal{C}_p$ of components of interest within phase p, such as CO_2 and

C_3 contaminants. The vector $\mathbf{q}_v = (q_{v,p}, q_{v,p}^c : p \in \mathcal{P}, c \in \mathcal{C}_p)$ tracks the phases and components, but the meaning depends on the node: $\mathbf{q}_v$ is the fluid flow traversing an intermediate node (manifold); the total production flow from a source (platform); or the flow entering a sink (terminal).

The pressure drop $\Delta p_e = p_u - p_v$ in a pipeline $e = (u, v) \in \mathcal{E}^{pipes}$ is a function of the flow q_e and physical properties of the pipeline, such as diameter, geometry, and roughness. The pressure drop is a function of the control variable when the arc $e \in \mathcal{E}^{eqp}$ models an equipment, such as a choke valve.

Table 2 lists the decision variables which play a part in the definition of physical and operating constraints.

- Flow balance must hold at internal nodes $v \in \mathcal{V}^{int}$:

$$\sum_{(u,v)\in\mathcal{E}} \mathbf{q}_{(u,v)} = \sum_{(v,u)\in\mathcal{E}} \mathbf{q}_{(v,u)}. \tag{1}$$

- The flows in the arcs (v, u) leaving a platform $v \in \mathcal{V}^{src}$ must add up to the platform's outlet production:

$$\sum_{(v,u)\in\mathcal{E}} \mathbf{q}_{(v,u)} = \mathbf{q}_v. \tag{2}$$

- The flows in the arcs (u, v) entering a terminal $v \in \mathcal{V}^{snk}$ must add up to the terminal's total inlet $\mathbf{q}_v$:

$$\sum_{(u,v)\in\mathcal{E}} \mathbf{q}_{(u,v)} = \mathbf{q}_v. \tag{3}$$

- Flows and pressures in arcs $(u, v) \in \mathcal{E}$ must be consistent with the pressure at upstream and downstream nodes:

$$\Delta p_{(u,v)} = p_u - p_v. \tag{4}$$

- The pressure drop in a pipeline $(u, v) \in \mathcal{E}^{pipe}$ is a function of flow and downstream pressure,

$$\Delta p_{(u,v)} = \widehat{\Delta p}_{(u,v)}(\mathbf{q}_{(u,v)}, p_v), \tag{5}$$

and a function of a control variable if (u, v) is an equipment.

The fraction of a component $c \in C_p$ is defined as:

$$z_{e,p}^c = \frac{q_{e,p}^c}{q_{e,p}}, \tag{6a}$$

$$z_{v,p}^c = \frac{q_{v,p}^c}{q_{v,p}}. \tag{6b}$$

for arc and node flows respectively. For an arc e, $q_{e,p}$ is the flow of phase p, $q_{e,p}^c$ is the flow of component c in phase p, and $z_{e,p}^c \in [0, 1]$ is the fraction. Variables for nodes are defined likewise.

Certain process limits at production platforms and terminals can be expressed in the form of constraints on component fractions. Additionally, compositional balance constraints must also hold for phase $p \in \mathcal{P}$ and component $c \in \mathcal{C}_p$ at network nodes:

TABLE 2. Variables in the conceptual formulation.

Variable	Description
p_v	Pressure at node $v \in \mathcal{V}$.
$q_{v,p}$	Flow rate of phase p from, through, or to a node v.
$q_{v,p}^c$	Component c of flow rate in phase p from, through, or into a node v.
$\mathbf{q}_v$	Flow rate vector associated to node v, $\mathbf{q}_v = (q_{v,p}, q_{v,p}^c : p \in \mathcal{P}, c \in \mathcal{C}_p)$.
$z_{v,p}^c$	Fraction of component c in phase p associated with node v, i.e. $z_{v,p}^c = q_{v,p}^c / q_{v,p}$.
$\mathbf{z}_v$	Vector of component fractions in flow related to node v, $\mathbf{z}_v = (z_{v,p}^c : p \in \mathcal{P}, c \in \mathcal{C}_p)$.
$\mathbf{x}_v$	Boundary condition for platform v, $\mathbf{x}_v = (\mathbf{q}_v, p_v, \mathbf{z}_c)$.
Δp_e	Pressure drop in edge $e = (u, v)$, $\Delta p_e = p_u - p_v$.
$q_{e,p}$	Flow rate of phase $p \in \mathcal{P}$ in arc $e \in \mathcal{E}$.
$q_{e,p}^c$	Flow rate of component c in phase p of arc e.
$\mathbf{q}_e$	Flow rate vector, $\mathbf{q}_e = (q_{e,p}, q_{e,p}^c : p \in \mathcal{P}, c \in \mathcal{C}_p)$.
$z_{e,p}^c$	Component c concentration in phase p of arc e.
$\mathbf{z}_e$	Vector of component fractions associated with flow in arc e, $\mathbf{z}_e = (z_{e,p}^c : p \in \mathcal{P}, c \in \mathcal{C}_p)$.
$\mathbf{q}$	System flow vector, $\mathbf{q} = (\mathbf{q}_v : v \in \mathcal{V}) \cup (\mathbf{q}_e : e \in \mathcal{E})$.
$\mathbf{p}$	System pressure vector, $\mathbf{p} = (\mathbf{p}_v : v \in \mathcal{V})$.
$\mathbf{z}$	System flow composition vector, $\mathbf{z} = (\mathbf{z}_v : v \in \mathcal{V}) \cup (\mathbf{z}_e : e \in \mathcal{E})$.

- For a production platform $v \in \mathcal{V}^{src}$, material balance must hold for component c of phase p,

$$q_{v,p}^c = \sum_{(v,u)\in\mathcal{E}} q_{(v,u),p}^c. \tag{7}$$

- In an internal node $v \in \mathcal{V}^{int}$, material balance must hold for the flow of component c in phase p,

$$\sum_{(u,v)\in\mathcal{E}} q_{(u,v),p}^c = \sum_{(v,u)\in\mathcal{E}} q_{(v,u),p}^c. \tag{8}$$

- Likewise, for a terminal node $v \in \mathcal{V}^{snk}$,

$$q_{v,p}^c = \sum_{(u,v)\in\mathcal{E}} q_{(u,v),p}^c. \tag{9}$$

- Compositions of flows leaving a source or internal $v \in \mathcal{V}^{src} \cup \mathcal{V}^{int}$ must be consistent,

$$z_{v,p}^c = z_{(v,u),p}^c, \quad \forall (v, u) \in \mathcal{E}, \ v \in \mathcal{V}^{src}, \tag{10a}$$

$$z_{(v,u),p}^c = z_{(v,w),p}^c, \quad \forall (v, u), (v, w) \in \mathcal{E}, \ v \in \mathcal{V}^{int}. \tag{10b}$$

From the first equation, the component fraction in streams emanating from a platform must be all the same, a common hypothesis in flow splitting [26]. For an internal node, the second equation ensures that streams of distinct compositions can reach the node, such as a manifold, where they get mixed into a homogeneous outlet.

A. CONSTRAINTS ON ARCS AND NODES
Bound constraints are imposed on each node $v \in \mathcal{V}$,

$$p_v^{min} \leq p_v \leq p_v^{max}, \tag{11a}$$

$$\mathbf{q}_v^{min} \leq \mathbf{q}_v \leq \mathbf{q}_v^{max}, \tag{11b}$$

and each arc $(u, v) \in \mathcal{E}$,

$$p_{(u,v)}^{\min} \leq p_{(u,v)} \leq p_{(u,v)}^{\max}, \tag{12a}$$

$$\Delta p_{(u,v)}^{\min} \leq \Delta p_{(u,v)} \leq \Delta p_{(u,v)}^{\max}, \tag{12b}$$

$$\mathbf{q}_{(u,v)}^{\min} \leq \mathbf{q}_{(u,v)} \leq \mathbf{q}_{(u,v)}^{\max}. \tag{12c}$$

Constraints are imposed on the composition of flows leaving production platforms and entering terminals $v \in \mathcal{V}^{snk} \cup \mathcal{V}^{src}$,

$$\mathbf{z}_v^{\min} \leq \mathbf{z}_v \leq \mathbf{z}_v^{\max}. \tag{13}$$

where $\mathbf{z}_v = (z_{v,p}^c : p \in \mathcal{P}, c \in \mathcal{C}_p\}$ is a vector with the compositions for all phases at node v.

B. BOUNDARY CONDITIONS AT PRODUCTION PLATFORMS

First proposed in [4], *boundary conditions* can model the interface of a production platform $v \in \mathcal{V}^{src}$ with the subsea gas network in terms of its outlet flow $\mathbf{q}_v$, pressure p_v, and vector $\mathbf{z}_v$ of compositions. Boundary conditions are a power tool to decouple a system into a set of subsystems, bringing about flexibility to connect heterogeneous subsystems and computational advantages to be explained later.

Let $\mathcal{X}_v = \{\mathbf{x}_v : \mathbf{x}_v \text{ is feasible}\}$ be the set of *feasible* boundary conditions, such that $\mathbf{x}_v = (\mathbf{q}_v, p_v, \mathbf{z}_v)$ holds all the relevant information to characterize the interface. If $\mathbf{x}_v \in \mathcal{X}_v$, there exist local control settings for wells, subsea and topside equipment that enable the platform to operate at the boundary condition $\mathbf{x}_v$. The feasible set $\mathcal{X}_v$ is assumed closed and bounded, a plausible assumption given that a platform v is restricted by physical limits.

Notice that multiple control settings might exist for a boundary condition $\mathbf{x}_v$, each potentially inducing a distinct objective value for the production unit. Whichever the objective function selected for a platform v, we assume that the optimal objective $f_v(\mathbf{x}_v)$ is known for the boundary condition $\mathbf{x}_v$, namely the maximum gain yielded by platform v when it operates at $\mathbf{x}_v$. This objective can be expressed in terms of production (cubic meters of oil per day) or any function modeling an economic gain.

The computation of control settings that maximize the objective for a given $\mathbf{x}_v$, *i.e.* calculating the optimal $f_v(\mathbf{x}_v)$, is a challenging task. It entails solving a potentially hard optimization problem of varying complexity that depends on the platform v. For instance, in offshore platforms operating with satellite wells, the optimization is relatively simple because surface operating conditions are determined by a fixed pressure at the separation unit [27]. In more complex environments, wells have subsea completion that demands the modeling of pressure, subsea and topside processing equipment, flow splitting, and routing to manifolds and headers, among other features [24], [25], [26].

C. CONCEPTUAL FORMULATION

This work focuses on the production optimization of large offshore oilfields, consisting of several platforms producing from multiple reservoirs that share a subsea gas pipeline and compete for limited resources. This production optimization problem is *conceptually* formulated as:

$$P : \max_{\boldsymbol{\theta}} f(\boldsymbol{\theta}) = \sum_{v \in \mathcal{V}^{src}} f_v(\mathbf{x}_v) \tag{14a}$$

$$\text{s.t.} : \mathbf{x}_v \in \mathcal{X}_v, \ v \in \mathcal{V}^{src}, \tag{14b}$$

$$\mathbf{H}(\mathbf{q}, \mathbf{p}, \mathbf{z}) = \mathbf{0}, \tag{14c}$$

$$\mathbf{G}(\mathbf{q}, \mathbf{p}, \mathbf{z}) \leq \mathbf{0}. \tag{14d}$$

where: (*i*) $\mathbf{q}$, $\mathbf{p}$, and $\mathbf{z}$ are vectors with the flows, pressures, and compositions for all network nodes and arcs; (*ii*) $\mathbf{x} = (\mathbf{x}_v : v \in \mathcal{V}^{src})$ is the vector with all boundary conditions; and (*iii*) $\boldsymbol{\theta} = (\mathbf{x}, \mathbf{q}, \mathbf{p}, \mathbf{z})$ collects all of the decision variables. The vector functions $\mathbf{H}$ and $\mathbf{G}$ model the network constraints (1)-(13), such as flow conservation, pressure balance, and composition restrictions. Further, the boundary conditions at sink nodes are embedded in the network constraints, such as their fixed operating pressure. The objective expresses the overall economic or production goal of the oilfield.

D. DISCUSSION

The production optimization problem P is regarded as *conceptual*, or *abstract*, because the sets $\mathcal{X}_v$ of boundary conditions, their respective objectives f_v, and pressure-drop relations are not explicitly known. Such relations are typically implemented in black-box simulators, or else need to be inferred from field measurements.

By generating proxy models, the conceptual problem is cast as a concrete optimization problem to which algorithms can be applied. The kinds of algorithms depend greatly on the nature of the proxy models, which are thus pivotal for the efficiency of a solution methodology.

We follow the modeling of platforms as sets $\mathcal{X}_v$ of boundary conditions as proposed in [4], using proxy models based on piecewise-linear approximation of the objective $f_v(\cdot)$ which enables the optimization of heterogeneous units. Explicitly accounting for the detailed model of production platforms would render an intractable problem P, given the complexity and sheer size of the MINLP problems for local optimization of platforms. This work contributes by proposing a methodology based on systems identification, and the properties of fluid flow in pipelines, that leads to the synthesis of proxy models for pipelines that are sufficiently accurate and extremely fast to compute. For one, the pipeline proxy models can be derived from field measurements or from data produced by high-fidelity fluid flow simulators. For another, the resulting proxy models can serve as extremely fast simulation tools in what-if analyses or embedded in P to obtain a concrete optimization problem.

In what follows, we propose a methodology for the synthesis of proxy models and the use of DFO for solving the resulting production optimization problem. The optimization algorithm will choose boundary conditions that ensure a feasible operation for the platforms and subsea network, while yielding an optimal system-wide operation.

III. BACKGROUND

This section presents basic concepts on the hydraulics of gas pipelines and networks, along with fundamentals on system identification that are key for developing proxy models.

A. GAS PIPELINES

The hydraulics of gas flow in pipelines is relevant for choosing a suitable structure for pipeline models, which will be identified from simulated data and real measurements. The focus is on equations and methods related to pressure calculation, given flows rates and CO_2 content. We start from equations of hydraulics for single-segment pipelines and then address multiple-segment networks.

1) HYDRAULICS

The theory of natural gas flowing in pipelines is in a very mature state. The Fritzsche equation that computes the flow rate for a given differential pressure dates back to 1908, for example. Several phenomena have been studied in detail in the past decades, leading to equations and models for a vast number of different flow scenarios.

For the sake of objectivity a scenario with modest complexity is chosen. It is not too complex where the physics intuition about the equations may be lost, yet not overly simple such that important information, which could potentially be used as prior knowledge in proxy modeling, is lost. The proposed scenario makes the following assumptions:

1) *Single-segment Pipeline*: Long offshore gas pipelines are often built of many connected different pieces of pipes with different physical properties; those smaller pieces are called segments. In order to compute flows given pressures, for example, the available methods use converging algorithms. Here, however, a single segment is considered to represent a pipeline and therefore equations can be used instead of algorithms.

2) *Single-phase Gas Flow*: Platforms can produce natural gas in purely gas phase, liquid phase or a mixture of them. Besides in the reservoir, natural gas can change its phase inside a pipeline depending on local pressure and temperature. With a mixture of phases, the so-called multi-phase flow, the equations become far more complex than the ones of single-phase. The hypothesis here is that single-phase equations can describe most behaviors satisfactorily, assuming there is a dominant flow regime in a particular segment.

3) *Real Conditions*: Under high pressure and at low temperature, which is the case in subsea pipelines, the ideal gas law is not applicable. A nonlinear term known as compressibility factor is defined as the ratio of the gas volume at a given temperature and pressure to the volume the gas would occupy if it were ideal at the same conditions. Given the relative magnitude of pressures and temperature variations in subsea pipelines this term can't be beforehand neglected.

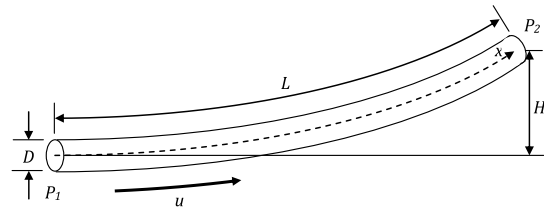4) *Pipeline Not Elevated*: Elevation is the difference of height between two points of a pipeline segment.

FIGURE 1. Pipeline segment with dimensions.

TABLE 3. Variables and parameters.

Parameter/ Variable	Description
u	fluid average velocity
P	absolute pressure
ρ_g	gas density
ρ_{air}	air density
g	gravitational acceleration
H	difference of height of points 1 and 2
f	Darcy friction factor
D	pipe inner diameter
m	mass
A	area
k_i	constants
q	gas flow rate
P_1	upstream pressure
P_2	downstream pressure
G	gas gravity
T_f	average gas flowing temperature
L	pipe segment length
ζ	gas compressibility factor at the flowing temperature
M_g	gas molecular weight
M_{air}	air molecular weight
M_i	molecular weight of a component i
y_i	molar fraction of component i
P_{avg}	average flowing pressure
Re	Reynolds number
μ	dynamic viscosity

The undersea elevation variation is normally small in comparison to the length of the pipeline itself, therefore is neglected.

5) *Isothermal Flow*: Under isothermal flow, the pressure drop can be calculated using constant gas properties such as the compressibility factor, instead of integrating the state equation for each infinitesimal piece of the pipeline. Because the whole purpose of the proxy is to avoid the use of algorithms, the assumption is that the flow is isothermal, in other words, that the gas temperature is constant along the pipeline.

6) *Turbulent Flow Regime in Smooth Pipes*: Among the categories of turbulent flow in pipes, this work chooses the turbulent flow in smooth pipes for better representing the scenario of interest. But other categories could be selected, such as turbulent flow in fully rough pipes, and transition flow between smooth and rough pipes.

Figure 1 illustrates a pipeline segment with dimensions. Before presenting the equations for gas flow in pipelines, let us introduce the nomenclature in Table 3.

The momentum equation applied to a portion of a pipe of length dx, inside which flows a compressible fluid, for

example natural gas, in steady state conditions is:

$$\underbrace{u\,du}_{kinetic\,energy} + \underbrace{\frac{dP}{\rho_g}}_{pressure\,force\,work} + \underbrace{g\,dH}_{potential\,energy} $$

$$+ \underbrace{f\,\frac{dx}{D}\frac{u^2}{2}}_{viscous\,friction} = 0 \qquad (15)$$

The third term, potential energy, which depends on height is neglected according with the scenario characterization presented above. The last term, viscous dissipation, is conveniently simplified in order to ease integration as follows:

$$\rho\,u = \frac{\dot{m}}{A} = k_0 \qquad (16)$$

which is a constant and leads to the equation:

$$\underbrace{\rho^2 u\,du}_{kinetic\,energy} + \underbrace{\rho\,dP}_{pressure\,force\,work} + \underbrace{f\,\frac{dx}{D}\frac{C^2}{2}}_{viscous\,friction} = 0 \quad (17)$$

Integrating Eq. (17) results in the known general gas flow equation,

$$q = k_1 \sqrt{\left(\frac{P_1^2 - P_2^2}{G \cdot T_f \cdot L \cdot \zeta \cdot f}\right) D^{2.5}} \qquad (18)$$

in which the meaning of the constants k_i are not relevant for the purpose of system identification. From Eq. (18) the influence of flow rate on pressures is quite obvious. The influence of CO_2 however is hidden in G and ζ. The compressibility factor ζ can be further detailed as [28]:

$$\zeta = \frac{1}{1 + \left(\frac{P_{ag} k_2 10^{G \cdot k_3}}{T_f^{k_4}}\right)} \qquad (19)$$

It can be noticed that both Eqs. (18) and (19) depend on the natural gas specific gravity G, which is define as:

$$G = \frac{M_g}{M_{air}} \qquad (20)$$

where M_{air} is the air molecular weight, which is a constant equal to 28.9625 g/mol, and M_{air} is given by the mixture of gases:

$$M_g = \sum_{i=1}^{\mathcal{I}} M_i \cdot y_i \qquad (21)$$

where M_i is the molecular weight, y_i is the molar fraction of the ith component and $\mathcal{I}$ is the number of components. Table 4 shows the molecular weight value for the most important components of natural gases.

The first key to figure out the influence of CO_2 on pressures regards the natural gas specific gravity G. From Eq. (21), the molecular weights reported in Table 4, and assuming that the molar fraction of Methane is much greater than the other components, it is possible to conclude that the higher the CO_2 content the higher will be the specific gravity G. Thus,

TABLE 4. Molecular weight of natural gas components.

Component	Molecular Weight (g/mol)
Methane	16.043
Ethane	30.070
Propane	44.097
Carbon Dioxide	44.010

increasing CO_2 content increases G, which in turn decreases the flow rate q for the given pressure as seen in Eq. (18). Conversely, in Eq. (19) the compressibility factor ζ shrinks exponentially with the increase of the CO_2 content, which thereby increases the flow rate.

So far the clues do not lead to a conclusion simple enough to drive the design of the proxy model structure. Another try is the Darcy friction factor f present in the denominator of Eq. (18). For turbulent flow in smooth pipes the friction factor is given by the Collebrook-White equation:

$$\frac{1}{\sqrt{f}} = -2\log\left(\frac{e}{3.7D} + \frac{2.51}{Re\sqrt{f}}\right) \qquad (22)$$

This equation could have been the second key to try to figure out the influence of CO_2 on pressures, but it is far too complicated and associated to the influence of the first key the final conclusion is that, for this part of the pressure model, an approach based on data should be favored. Such an approach is called gray-box system identification.

2) PIPELINE NETWORK

So far the basic scenario has considered a single pipeline segment. A bit more complex scenario in which there are two different pipeline segments connected is shown in Figure 2. If the pressure $P_{1,1}$ and the flow rate q_1 are known, then it is possible to calculate $P_{1,2} = P_{2,1}$ with Eq. (18) and after that calculate $P_{2,2}$. As such, $P_{2,2}$ no longer comes from an equation, but from a cascade algorithm that uses this equation. Still this is a one-pass algorithm which, from a proxy system identification perspective, could be tackled as a hierarchically structured model.
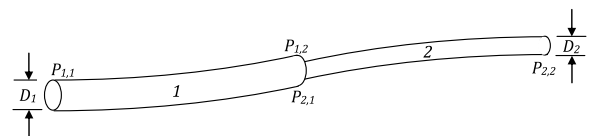

FIGURE 2. Two segments pipeline.

Considering gas pipeline networks found in scenarios of interest, such as oilfields in the Santos Basin, streams from multiple pipelines flow into a single pipeline as illustrated in the simple network shown in Figure 3. In this context, assuming that the pressure $P_{3,2}$ and flows q_1 and q_2 are known, the objective is to calculate pressures $P_{1,1}$ and $P_{2,1}$. If all simplifications are kept, the solution can also come from a one-pass algorithm. It computes first pressures $P_{1,2} = P_{2,2}$
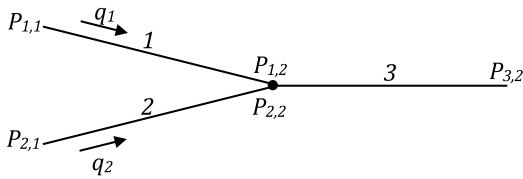
FIGURE 3. Small pipeline network.

that depend on $P_{3,2}$ and the sum of q_1 and q_2. With these pressure values, then $P_{1,1}$ and $P_{2,1}$ can be calculated.

If the assumptions from Section III-A1 no longer hold, then several other variables start to play a role in the computation process. For instance, the gas and liquid-phase rates change along the segment according to the pressure gradient, possibly involving complex behavior such as flashing. The change of phase in itself also influences the pressure gradient and the integrated equation such as (18) cannot be applied.

Simulators [29] handle such complex settings by iterative algorithms that integrate fluid property values along the pipe segment length. They also use another iterative method called nodal analysis that guesses flow rates and converges in an iterative fashion to a set of consistent pressures. This iterative nature is the reason for a high computational burden. Hundreds or even thousands of iterations are often needed to converge. Not only the number of iterations, but also the complexity of the equations involved in the calculations play a major role in the total computational load.

To be fast the chosen proxy model structure would need to avoid this iterative nature and simplify as much as possible the involved equations. The next sections detail how to achieve these objectives.

B. SYSTEMS IDENTIFICATION

System identification is a methodology for building mathematical models of dynamic systems using measurements of the system's input and output signals [30].

The field of system identification can be broadly divided in two main categories: linear and nonlinear. The field of system identification can be broadly divided into two main categories: linear and nonlinear. Linear models should be considered first; only when they do not yield a satisfactory performance should nonlinear models be used, which are significantly more complex. As will be seen later, the relation between pressures and flow rates of natural gas flowing in pipeline networks is nonlinear. Following, there is a brief description of the sequence of tasks performed in nonlinear system identification for static models.

1) MODEL INPUTS AND OUTPUTS

The inputs and outputs are straightforward for the problem of concern. The inputs consist of (i) the natural gas flow rate q_v from the offshore platforms $v \in \mathcal{V}^{src}$, assuming a single phase regime ($\mathcal{P} = \{gas\}$), (ii) fraction z_v^c of the component $c \in \mathcal{C}$ in the gas stream q_v, assuming CO_2 as the component

of interest. The output is (i) the exporting pressure p_v of the production platforms.

2) MODEL ARCHITECTURE

The selection of the architecture is a subjective decision that should be made based on three aspects.

The first aspect regards how the system's output changes along time once the input is set, to put in other words, if the model is static or dynamic. (i) *Static* is a kind of modeling to which the time between an excitation signal and its respective output response is not important. The outputs change instantaneously with input variation or reach a steady-state after lag time known as settle-time, which is a property of stable system. (ii) *Dynamic* modeling is when the temporal behavior of the system is taken into account. Such models tend to be more complex, given that the trajectory of the states of the system is also modeled during the settle-time [31].

The second aspect relates to the overall modeling approach. (i) *White-Box* is fully derived by first principles, such as physical and chemical laws, leading to a theoretical modeling of all equations and parameters [31]. (ii) *Black-Box* is based solely on measurement data. Both model structure and parameters are determined from experimental modeling. It is appropriate when no prior knowledge about the system is available, or the equations based on first principles are highly complex [31], [32]. (iii) *Grey-Box* represents a compromise between white- and black-box models by integrating various kinds of available information. Typically, the model's structure is determined from prior knowledge, whereas measurement data are used for parameter identification [32].

Finally, the third aspect consider the mathematical functions that define the model. (i) *Linear* models are the simplest and must be used whenever possible. (ii) *Polynomials* extend linear models, becoming more flexible as the degree increases [31]. (iii) *Look-up Tables* are essentially a large collection of samples combined with some interpolation and optimization technique. (iv) *Neural Networks* (NNs) are mathematical arrangements inspired in biological neurons with distinct structures that can be particularly interesting for static (multilayer perceptron) or dynamic models (recurrent neural networks). NNs with a large number of layers, known as deep learning, have achieved great success in complex problems. They are difficult and slow to train, and the incorporation of prior knowledge is limited. Recently, [33] proposed a regularization to bring prior knowledge of the physics involved into the training of NNs.

3) MODEL STRUCTURE

The basic principle here is to compose a complex model of many simpler low-dimension models. Some ways to organize the smaller models include the multiplicative correction model, refinement submodel, parameter scheduling, projection based structures, and additive and hierarchical structures. Hierarchical structures are particularly helpful to

model physical phenomena, mainly because they break the model down into physically meaningful pieces.

4) PARAMETER ESTIMATION

This step is not as subjective as the previous ones. It is the application of well know techniques over the model skeleton built so far, with defined inputs, architecture and structure.

A key concept is the *loss function*, which serves to assess the quality of the model parameters and guide optimization algorithms. The loss function gives the error value as a function of the parameters of the model. In a broad sense, error is the difference between the output of the model and the output of the process for a given input:

$$e_j = y_j - \widehat{y_j} \tag{23}$$

Arguably, the most common form of the loss function is the sum of squared errors:

$$I(\boldsymbol{\theta}) = \sum_{i=1}^{N} e_j^2 \tag{24}$$

where $\boldsymbol{\theta}$ is the vector with model parameters and N is the number of training data samples.

When the error between process and model output is linear on the parameters, and the sum of squares is applied as a loss function, the optimal parameters can be computed by means of *least squares*. For the sake of simplicity, we consider the case where the system input u and output y are scalars. Given an input u_j, the predicted model output $\widehat{y_j}$ is given by

$$\widehat{y_j} = \sum_{i=1}^{n} \theta_i x_{i,j} \tag{25}$$

where $x_{i,j}$ are called regressors given by

$$x_{i,j} = g_i(u_j) \tag{26}$$

with $g_i(\cdot)$ being problem dependent functions. For instance, if $g_i(u) = u^{i-1}$ is a monomial of ith degree, the prediction becomes a polynomial function of the inputs. The regressors can be arranged in matrix forms as

$$X = [\mathbf{x}_1 \ \mathbf{x}_2 \ \cdots \ \mathbf{x}_n] \tag{27}$$

where $\mathbf{x}_i = (x_{i,1}, x_{i,2}, \ldots, x_{i,m})$ is a column vector with the ith regressors for all inputs. Let $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_n)$ be the vector of model parameters. Rewriting Eq. (25) as

$$\widehat{\mathbf{y}} = X\boldsymbol{\theta} \tag{28}$$

and inserting it into Eq. (23), leads to the error vector,

$$\mathbf{e} = \mathbf{y} - X\boldsymbol{\theta}. \tag{29}$$

The least-squares problem for the loss function (24) becomes

$$\min_{\boldsymbol{\theta}} I(\boldsymbol{\theta}) = \frac{1}{2} \|\mathbf{y} - X\boldsymbol{\theta}\|^2 \tag{30}$$

which can be solved efficiently by means of Cholesky factorization and iterative methods.

5) MODEL VALIDATION

The aim here is to check whether or not all previous steps were carried out successfully. This is achieved in two steps. (*i*) The first step uses a criterion to measure the quality of the model on training data. The criteria can be a visual inspection of data plots, sum of squares due to error (SSE), coefficient of determination (R^2), degrees of freedom in the error (DFE), degree-of-freedom adjusted coefficient of determination, root mean squared error (RMSE), and finally expected error (EE). (*i*) If the chosen criterion calculated on training data achieves good enough results, the second task is to reevaluate the model on fresh data, also called test data.

The coefficient of determination is a measure of the system response variation that is explained by the model. An intuitive relation is given by:

$$R^2 = \frac{\text{Explained variation}}{\text{Total variation}} \tag{31}$$

For instance, $R^2 = 0.9$ means that 90% of system variation is explained by the model. In mathematical notation,

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} = 1 - \frac{\sum_{j=1}^{m}(y_j - \widehat{y_j})^2}{\sum_{j=1}^{m}(y_j - \overline{y})^2} \tag{32}$$

where $\overline{y}$ is the mean of training data output $\{y_j\}_{j=1}^{m}$.

The expected error is a measure of the model output error expectation. Because the value is expressed in the same units of the system's output, it is intuitively interpretable:

$$EE = \frac{\sum_{j=1}^{m}(y_j - \overline{y})}{m} \tag{33}$$

IV. SOLUTION FRAMEWORK

This section addresses the synthesis of a concrete formulation for the conceptual problem (14), which can be solved with standard algorithms for production optimization. The nature of the approximation of the objective and constraints is pivotal with respect to the class of algorithms that are applicable.

The equations (14c) that relate flows, pressures, and compositions of the gas pipeline network can be modeled directly with a phenomenological simulator, or with the proxy models developed in the previous section. Conversely, the objective function f_v is hard to compute since it entails solving a complex MINLP problem [25] for a given boundary condition $\mathbf{x}_v$, which depends on the particular structure of the platform v.

A. MODELING THE OBJECTIVE FUNCTION

Since it would impractical to solve f_v for several $\mathbf{x}_v$ iteratively, an alternative is to compute this objective offline for a finite, sample set $\widetilde{\mathcal{X}}_v \subset \mathcal{X}_v$ of feasible boundary conditions. Then, f_v can be approximated by a multidimensional piecewise-linear function $\widetilde{f}_v$ using the point set $S_v = \{(\mathbf{x}_v, f_v(\mathbf{x}_v)) : \mathbf{x}_v \in \widetilde{\mathcal{X}}_v\}$. Notice that the resulting piecewise-linear function is a proxy model for the objective. The modeling of a piecewise-linear function in mathematical programming is

a topic of considerable complexity though. The *first issue* regards the partitioning of the domain $\widetilde{\mathcal{X}}_v$ into a polyhedron set, which can be an arbitrary set of polyhedra, a triangulation, and a regular lattice, among others. The *second issue* is about the formulation of $\widetilde{f}_v$ in mathematical programming with the use of continuous and binary variables [34], [35]. A number of formulations are found in the literature that depend on the domain partition, including the convex combination, the disaggregated convex combination, the logarithmic convex combination, and the special order set of type 2 (SOS2) [34]. Instead, if the problem is solved with a DFO algorithm, then $\widetilde{f}_v(\mathbf{x}_v)$ can be obtained by convex combination of the closest points to $\mathbf{x}_v$ in S_v that form a simplex.

B. MODELING THE PIPELINE NETWORK

This work considers the use of the multiphase flow simulator Pipesim to solve the pipeline network equations (14c), and thereby compute the node pressures and pressure drops for a given vector $\mathbf{x} = (\mathbf{x}_v : v \in \mathcal{V}^{src})$ of boundary conditions. Together with a piecewise-linear approximation of the objective function f_v, this black-box model leads to a concrete problem that can be solved with DFO algorithms. This approach, which relies on the use of a black-box simulator in the optimization loop, was proposed in [17], which serves as a reference model and a means of comparison.

Conversely, our work proposes the use of the proxy model developed in the previous section in place of the black-box simulator. The concrete problem that results is explicit in the sense that network relations are given by algebraic equations, without simulation models that are computationally costly to run. Given the good accuracy of the proxy models identified for the pipelines in the Santos Basin, this alternative brings about efficiency gains. Actually, two solution methods may result from the pipeline proxy model. The first is obtained by combining the proxy models for the pipeline pressure drops (Eq. (5)), with a mixed-integer formulation of the objective functions f_v. Instead of a MILP formulation for f_v, the second uses the set S_v as a look-up table to compute $\widetilde{f}_v(\mathbf{x}_v)$ by convex approximation of the closest points to a given boundary condition $\mathbf{x}_v$, as in the case of direct use of a black-box simulator.

C. CONCRETE FORMULATIONS

Considering the models presented above for the objective function and pipeline network, three concrete formulations for the production optimization problem (14) are readily derived:

- *Simulation-Based (SB):* It consists of the use of Pipesim simulator to solve the pipeline network equations and algorithmic piecewise-linear modeling of the objective function in tandem with a DFO algorithm.
- *Proxy-Based (PB):* Akin to the simulation-based strategy, but instead the pipeline network equations are solved approximately and algebraically using the proxy-models identified from simulation and field data.
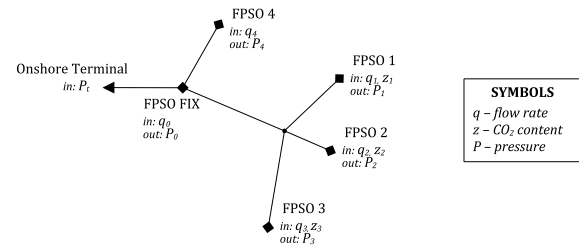

FIGURE 4. Schematic representation of a gas pipeline network in the Santos Basin. Model variables are indicated in the network.

The resulting concrete formulation can be optimized with DFOs.

- *Mixed Approach (MA):* The mixed approach combines both simulation and proxy-based strategies. Alternate searches can be performed between the two methods, or a sequential approach can be followed (*i.e.*, first optimizing the proxy model and using the results as a starting point for the simulation-based procedure).

V. CASE STUDY
A. SANTOS BASIN
The Santos Basin is a sedimentary basin on the Brazilian continental shelf, extending for hundreds of kilometers along the coast. Large oil and gas reserves were discovered in deep waters about 300 km off the coast, below the salt layer, the so-called Pre-salt. The Pre-salt carbonate reserves is probably the most important oil discovery in Brazil [36]. From the production point of view, this new frontier has technological challenges that are still being addressed [2]. Figure 4 gives the schematic representation of a gas pipeline network in the Santos Basin.

FIX is a fixed platform producing gas and condensate from wells with subsea completion. This platform is connected to the onshore terminal by a subsea pipeline of 167 km. FIX was designed to receive gas from the Floating Production Offloading and Storage (FPSO) platforms. The gas with high CO_2 content produced from Pre-salt reservoirs by FPSOs 1, 2 and 3 which is gathered in a manifold M_{PS}. The gas from the Pre-salt is then mixed with gas from the Pos-salt produced by FPSO 4 and FIX, which have low CO_2 content, and then transferred to the onshore terminal. The oil produced by the FPSOs is transported to refineries by special vessels.

The challenge for integrated operations rests on maximizing overall production from the oil reservoirs, while accounting for physical and operational constraints in the production platforms, pipelines, and terminals. Of concern is the concentration of contaminants in the gas delivered to the terminal, which must be within the operational range of the onshore facilities.

The set of network sources $\mathcal{V}^{src} = \{$FIX, FPSO-1, FPSO-2, FPSO-3, FPSO-4$\}$ consists of the platforms, the set of network sinks of intermediate nodes $\mathcal{V}^{snk} = \{$OS$\}$ models the onshore terminal (OS), and the set $\mathcal{V}^{int} = \{$M$_{PS}\}$ contains the Pre-salt manifold. The set of pipelines is

$\mathcal{E}^{pipes} = \{(\text{FPSO-1, M}_{\text{PS}}), (\text{FPSO-2, M}_{\text{PS}}), (\text{FPSO-3, M}_{\text{PS}}),$ $(\text{M}_{\text{PS}}, \text{FIX}), (\text{FPSO-4, FIX}), (\text{FIX, OS})\}$ and $\mathcal{E}^{eqp} = \emptyset$. The network extends over a large area, with pipelines reaching almost 200 km in length and depth varying from about 50 m to more than 2000 m.

The network transports gas from the platforms to the onshore terminal which imposes limits mainly on CO_2 content, but potentially others in the future as the reservoirs mature, such as C_3. Thus the phase set is $\mathcal{P} = \{\text{gas}\}$ and the content of interest is $\mathcal{C}_{gas} = \{CO_2\}$. The boundary condition $\mathbf{x}_v$ of a platform v consists of the gas outlet flow $q_{v,\text{gas}}$, the outlet pressure p_v, and the CO_2 content $z_{v,\text{gas}}^{CO_2}$ in the gas phase. The onshore terminal $t = \text{OS}$ operates at a fixed pressure $p_t = 45$ bar.

The case study and its associated characteristics including wells and pipelines are based on actual production systems operated by Petrobras. For this reason, further details of the processes are confidential and cannot be disclosed.

B. PROXY MODEL SYNTHESIS

Here, the methodology for systems identification is applied following the steps previously presented in Subsection III-B, with support from the theory on gas pipeline hydraulics. The aim is to design a proxy model for pressure calculation in the subsea gas pipeline of the Santos Basin.

1) MODEL INPUTS

As illustrated in Figure 4, the boundary conditions with the network inputs are:

1) $\mathbf{x}_v = (p_v)$: the operating pressure at the terminal $v = \text{OS}$.
2) $\mathbf{x}_v = (q_v, z_{v,\text{gas}}^{CO_2})$: the exporting gas rate and CO_2 content from the platforms $v \in \{\text{FPSO-1, FPSO-2, FPSO-3}\}$ that operate in the Pre-salt reservoirs.
3) $\mathbf{x}_v = (q_v)$: the exporting rate of the platforms $v \in \{\text{FIX, FPSO-4}\}$ that produce gas without contaminants.

2) MODEL OUTPUTS

The output consists of total gas delivered to the terminal, its CO_2 content, and pressures at the source and intermediate nodes of the network. Mathematically, the output vector is $\mathbf{y} = (q_t, z_{t,\text{gas}}^{CO_2}) \cup (p_v : v \in \mathcal{V}^{src} \cup \mathcal{V}^{int})$.

For the proxy model formulation, we assume that there is no loss in the gas flows (i.e., the total flow is the sum of the individual platform flows) and no phase transition. These assumptions allow the approximation of the gas delivered to the terminal and its CO_2 content with the following algebraic equations:

$$q_{\text{t}} = \sum_{v \in \mathcal{V}^{src}} q_v \tag{34a}$$

$$z_{\text{t,gas}}^{CO_2} = \frac{\sum_{v \in \mathcal{V}^{src}} q_v \cdot z_{v,\text{gas}}^{CO_2}}{\sum_{v \in \mathcal{V}^{src}} q_v} \tag{34b}$$

As for pressure outputs, a proxy model is presented below based on the system identification methodology presented above. Notice that when the experiments employ the phenomenological simulator for benchmark purposes, no such assumptions are made and the pressures are taken from the resulting simulations.

3) MODEL ARCHITECTURE

Regarding the model architecture, three decisions are made:

1) *Temporality*: because the settling-time is much shorter than setpoint changes in daily production operations, a static model is chosen rather than a dynamic model.
2) *Overall Modeling Approach*: white-box modeling is overly complex and slow for optimization purposes, which motivates the design of a proxy model. The gray-box is favored in regards to black-box modeling because of the prior knowledge seen thus far in terms of physical principles, which can determine the model structure.
3) *Mathematical Base Functions*: considering the general flow equation (18), polynomial functions seem appropriate to model the relation between pressures and flow rates. Conversely, the equations that relate pressures and CO_2 content are far more complicated than the relation between flows and pressures. For these reasons, polynomial functions of flow rate and CO_2 content will be used as proxy models for pressures.

4) MODEL STRUCTURE

The model structure is based on two decisions:

1) *Polynomial Degrees*: Equation (18) shows a quadratic relation between pressures and flow rate in pipelines, therefore a polynomial of second degree is chosen. Given the complexity of the relation between pressure and CO_2 content based on prior knowledge, a second degree polynomial is initially chosen to represent the nonlinear behavior. The degree may be increased depending on performance.
2) *Submodel Structure*: With 9 inputs for 5 outputs (pressures at network nodes), the model dimensionality is relatively low to be handled without submodel composition. Nevertheless, it is worth looking into the behavior of a single variable to assess how well a polynomial model would fit the process. Figure 5 shows how the exporting pressure at FPSO 1 behaves according to its flow rate and CO_2 content variation, keeping constant the boundary conditions of the remaining platforms.

From Figure 5 one concludes that the raw data is not a good fit for a polynomial model. This highly nonlinear behavior arises from the superposition of several complex equations and algorithms used in white-box simulators, which jointly generate data that is difficult to interpret.

This work proposes a simple submodel defined as a string of segments without junctions that preserves enough homogeneity such that the integrated
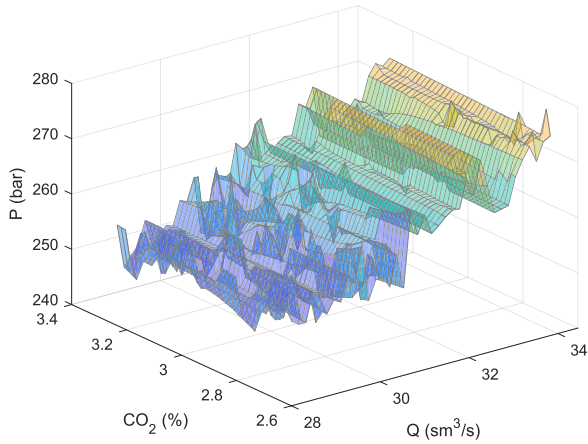
FIGURE 5. Data collected from a phenomenological simulator showing FPSO 1 pressure as a function of its own gas flow and CO_2 content, keeping fixed the boundary conditions of the other platforms at their nominal operating point.

equation (18) is still applicable. In such a composition, the original network must be dismantled in as few pieces as possible, as shown in Figure 6, otherwise too many submodels would have to be identified.

The pipeline segments from a to f, highlighted in red, are modeled individually. Each submodel has only one output, being the difference between its upstream and downstream pressure, called differential pressure ΔP. The inputs are the flow rate q, CO_2 content z, and downstream pressure P_{ds}. The use of downstream pressure as input variable decouples a submodel from the adjacent ones.

Figure 7 shows the behavior of one of these pieces (submodel) with data gathered from the simulator. The curve is much smoother than in Figure 5 and apparently fits quite well to a polynomial model due to its curvature.

Considering the developments above, the second-degree polynomial equation (35) is used as a proxy model for the pressure drop ΔP in a pipeline segment, with the number of inputs being equal to three. This submodel equation is used for all six pipeline segments, from a to f, as follows:

$$\begin{aligned}
\Delta P = {} & \theta_0 + \theta_1 \cdot q + \theta_2 \cdot z_{CO_2} + \theta_3 \cdot P_{ds} \\
& + \theta_4 \cdot q^2 + \theta_5 \cdot z_{CO_2}^2 + \theta_6 \cdot P_{ds}^2 \\
& + \theta_7 \cdot q \cdot z_{CO_2} + \theta_8 \cdot q \cdot P_{ds} + \theta_9 \cdot P_{ds} \cdot z_{CO_2}
\end{aligned}$$

$$(35)$$

The final composite model able to calculate the pressures of the platforms, instead of pieces, is organized hierarchically and the set of equations, one for each platform, is:

$$P_{\text{FIX}} = P_t + \Delta P_f \qquad (36a)$$

$$P_{\text{FPSO1}} = P_t + \Delta P_f + \Delta P_d + \Delta P_a \qquad (36b)$$

$$P_{\text{FPSO2}} = P_t + \Delta P_f + \Delta P_d + \Delta P_b \qquad (36c)$$

$$P_{\text{FPSO3}} = P_t + \Delta P_f + \Delta P_d + \Delta P_c \qquad (36d)$$

$$P_{\text{FPSO4}} = P_t + \Delta P_e \qquad (36e)$$

C. COMPUTATIONAL ANALYSIS

In this section, the outcome of proxy models for estimating pressure in gas pipeline networks is evaluated in relation to the optimization of production in a multi-reservoir oilfield situated in the Santos Basin. Three distinct formulations are solved using a DFO algorithm, and the accuracy of these models is subsequently assessed.

1) PROXY MODELING

The values for θ in each $\Delta P_{a...f}$ of Equation (36) is now estimated. In order to train the proxy model, a set of data containing inputs u_i and outputs y_i are supplied by the phenomenological simulator. For each submodel a to f, 8000 samples were collected from the combination of $20 \times 20 \times 20$ inputs (q, z and P_{ds}), except for the gas stream in the segment e which does not carry CO_2. From this database, the values of all θ are calculated by least squares and the result is shown in Table 5.

TABLE 5. Values of proxy model parameters.

	a	b	c	d	e	f
θ_0	9.0457	12.1601	16.6467	84.0775	20.3603	-17.0242
θ_1	0.1925	0.1691	-0.0113	0.5305	1.2749	0.3399
θ_2	0.3493	0.3590	0.3395	-0.0844	0.0000	-0.0042
θ_3	0.2866	0.2501	0.1234	-0.3413	-0.4231	0.0000
θ_4	0.0074	0.0039	0.0050	0.0077	0.0099	-0.0003
θ_5	0.0035	0.0009	-0.0004	-0.0105	0.0000	0.0132
θ_6	-0.0003	-0.0002	-0.0002	-0.0003	0.0047	0.0000
θ_7	0.0011	0.0017	0.0011	0.0125	0.0000	0.0010
θ_8	-0.0009	-0.0006	-0.0006	-0.0028	-0.0094	0.0000
θ_9	0.0020	0.0020	0.0001	0.0020	0.0000	0.0000

Based only on data from simulator, the criteria R^2 and EE as defined by Equations (32) and (33) are applied to each submodel a to f. The results are shown in Table 6.

TABLE 6. Model quality indexes.

	a	b	c	d	e	f
R^2 [-]	0.9993	0.9999	0.9996	0.9885	0.9859	0.9807
EE [bar]	0.0753	0.0211	0.0179	2.1456	1.5416	1.3314

As far as R^2 is concerned, the worst fitting result is in submodel f. The interpretation is that the model explains 98% of ΔP response. Regarding EE, submodel d was outperformed by the others. The value 2.1456 is the expected error in pressure units for this pipeline segment.

2) TEST CASES

The nominal case is denoted by BASE. PRE_{gas} is a variation of the nominal case in which the total gas coming from the Pre-Salt is increased by 15%, while the same ratio
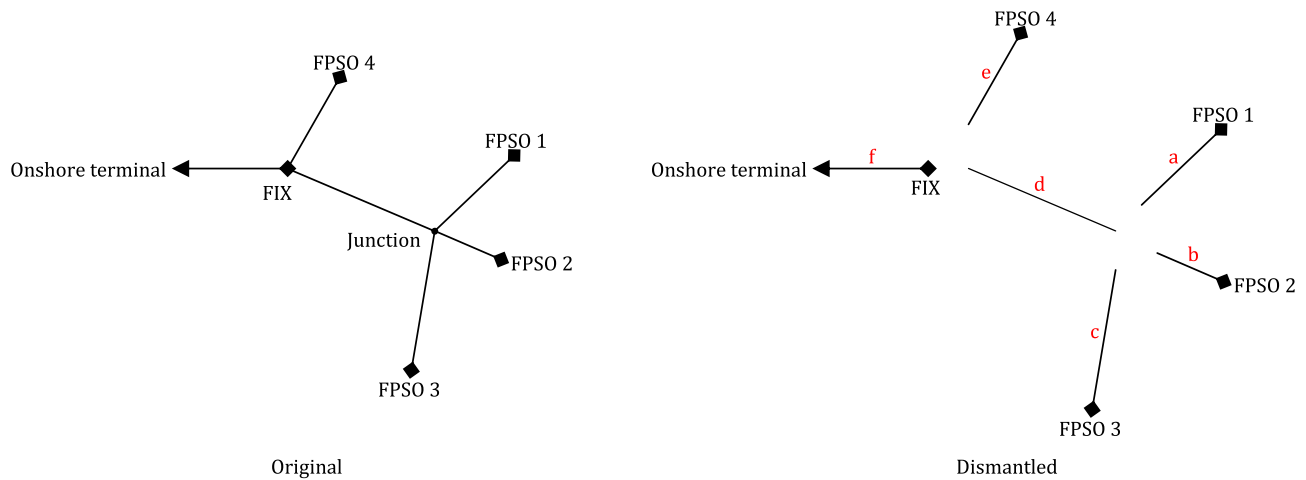
FIGURE 6. Santos Basin gas pipeline network representation. On the left-hand side the original network, on the right-hand side the dismantled network.

reduces the Post-Salt gas. The initial guess POS_{gas} represents a boundary condition in which the gas coming from the Pos-Salt is increased by 15%, and the gas with high content of contaminants is decreased in the same proportion. The same analogy holds for PRE_{CO2} and POS_{CO2}, i.e., these guesses represent boundary conditions in which the CO_2 concentration is increased and decreased by 15% with respect to the nominal case, respectively.

However, since the CO_2 concentration is constant for the Pos-Salt platforms, only a variation in CO_2 content of the Pre-Salt platforms is considered. It should be mentioned that the values of physical quantities are given in the International System of Units (SI). Gas flow is measured in million standard cubic meters per day (MSm^3/d), CO_2 concentration in percent (%), and time in seconds. Three test cases are considered in the experiments:

- C1: the first case does not impose constraints on the flow rate ($q_{t,gas}^{max}$) and CO_2 concentration (z_{t,CO_2}^{max}) at the terminal, only pressure constraints are in place.
- C2: the second one limits the total gas reaching the onshore terminal with respect to the first case.
- C3: the third case constrains the maximum CO_2 content allowed in the gas stream reaching the terminal, while relaxing the upper bound on the total gas delivered.

All three cases impose a maximum pressure of 250 bar at the production platforms. Overall, the problem to be optimized consist of:

$$\max_{\boldsymbol{\theta}} f(\boldsymbol{\theta}) = \sum_{v \in \mathcal{V}^{src}} f_v(\mathbf{x}_v) \tag{37a}$$

$$\text{s.t.} : \begin{cases} \mathbf{x}_v \in \mathcal{X}_v, \\ p_v(\mathbf{q}, \mathbf{p}, \mathbf{z}) \le 250, \end{cases} v \in \mathcal{V}^{src}, \tag{37b}$$

$$\mathbf{H}(\mathbf{q}, \mathbf{p}, \mathbf{z}) = \mathbf{0}, \tag{37c}$$

$$\mathbf{G}(\mathbf{q}, \mathbf{p}, \mathbf{z}) \le \mathbf{0}, \tag{37d}$$

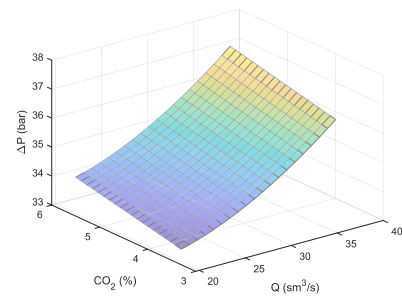$$q_t(\mathbf{q}, \mathbf{p}, \mathbf{z}) \le q_{gas,t}^{max}, \tag{37e}$$

FIGURE 7. FPSO 1 differential pressure as a function of its gas flow rate and CO_2 content in an isolated segment with fixed downstream pressure.

$$z_{t,gas}^{CO_2}(\mathbf{q}, \mathbf{p}, \mathbf{z}) \le z_{t,CO_2}^{max}, \tag{37f}$$

$$p_t = p_t^{fix}. \tag{37g}$$

where $f_v(\mathbf{x}_v)$ represents the total oil production from platform v (Sm^3/d), when it operates according to boundary condition $\mathbf{x}_v$; and $p_t^{fix} = 45$ bar. However, other objectives can be readily accommodated such as the maximization of an economic gain considering the revenue from hydrocarbon sales, discounted costs regarding gas compression and water treatment before discharge.

3) OPTIMIZATION RESULTS
For all test cases and concrete formulations, the resulting production optimization problem (37) was solved with the OrthoMADs algorithm [37], which is part of a class of algorithm known as MADS (Mesh Adaptive Direct Search) [38]. MADS employs a strategy that seeks to find an optimal point of a function by evaluating and comparing its value at different points. In Mesh Adaptive methods, all points are on top of a mesh which is refined iteratively. Overall, the goal of each iteration k is to find a point $\mathbf{x}$ that improves the best point $\mathbf{x}_k$ known thus far (i.e., $f(\mathbf{x}) < f(\mathbf{x}_k)$).

TABLE 7. Simulation-based approach using the pipesim simulator (benchmark).

	$q_{t,\text{gas}}^{\max}$	$z_{t,\text{CO}_2}^{\max}$	x_0	$q_{t,\text{gas}}$	$z_{t,\text{gas}}^{\text{CO}_2}$	$f(\theta)$	f. eval.	time [s]
	5.663	2.50	BASE	5.27	1.62	57425	251	2209.72
	5.663	2.50	PRE_{gas}	5.54	1.66	59728	320	2835.46
C1	5.663	2.50	POS_{gas}	5.56	1.41	57496	128	1178.59
	5.663	2.50	PRE_{CO2}	5.19	1.45	57425	356	3157.83
	5.663	2.50	POS_{CO2}	5.65	1.65	57412	228	2132.25
	3.681	2.50	BASE	3.68	2.20	59370	367	4049.32
	3.681	2.50	PRE_{gas}	3.67	1.96	59378	550	6057.91
C2	3.681	2.50	POS_{gas}	3.67	2.21	59443	414	4713.02
	3.681	2.50	PRE_{CO2}	3.67	1.94	57178	415	4761.96
	3.681	2.50	POS_{CO2}	3.67	2.25	57086	341	3909.63
	5.663	1.25	BASE	5.66	1.06	56436	266	2918.41
	5.663	1.25	PRE_{gas}	5.56	1.06	56136	242	2645.74
C3	5.663	1.25	POS_{gas}	5.65	1.06	56736	217	2414.33
	5.663	1.25	PRE_{CO2}	5.07	1.05	39876	198	2132.70
	5.663	1.25	POS_{CO2}	5.61	1.05	40420	316	3326.17

Gas flow rates are given in $MSm^3/d.$
Oil flow rates ($f(\theta)$) are given in $Sm^3/d.$

OrthoMADS provides an improvement in the Poll stage of the MADS algorithm, considering polling directions that are deterministic and orthogonal to each other [39]. The algorithm uses Halton's pseudo-random sequence, which covers the space more evenly than a real random sequence, to generate tentative solution vectors. Our work used the NOMAD package [40], which implements the OrthoMADS algorithm in C++. The experiments were implemented in Python 3.6 using OrthoMADS and performed in a Windows virtualized environment, with 2 cores and 8 GB of RAM.

SIMULATION-BASED METHOD

Table 7 presents the results obtained by optimizing problem (37) with direct use of Pipesim as the black-box simulator (*Simulation Based*). These results define the benchmark, against which the results to be obtained with the support of proxy models will be compared.

PROXY-BASED METHOD

For the proxy model optimization (*Proxy-Based*), a safety threshold variable (γ) is defined on the pressure constraints to counter for the cumulative error in the network pressures, which are calculated according with ΔP regressions given in Eq. (36). After optimizing problem (37) with the proxy model, the obtained solution is tested with the simulator for more accurate pressure measurements, i.e., checking solution feasibility. Figure 8 presents the workflow of the proposed proxy-based strategy. With the addition of γ, the pressure constraints become:

$$P \leq (1 - \gamma)250, \tag{38}$$

with $\gamma \in (0, 1)$. Next, we will illustrate the impact of the parameter γ on solution feasibility.

Table 8 shows the results for the proxy-based optimization following the workflow from Figure 8, with γ ranging from 0% to 5%. It is possible to notice that, given some initial scenarios and conditions, the feasibility of the solution with
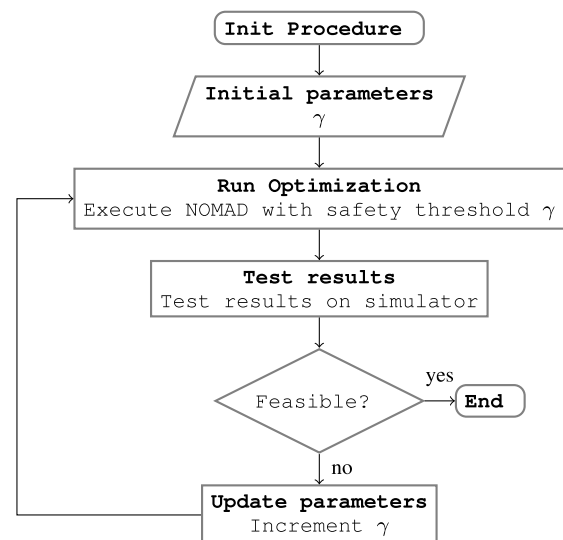
FIGURE 8. Workflow of the proxy model optimization.

respect to the simulator is achieved by increasing the value of γ.

An interesting aspect is that the optimization with the proxy model produced better results than the benchmark, arguably because of the smoothing characteristic resulting from the regression. When using the simulator itself, sometimes the optimizer can get stuck at certain points on the grid, being prevented from leaving the region due to the non-smoothness of the simulations. The non-smoothness of the simulation output can be observed in Figure 5.

HYBRID METHODS

Considering the optimization based jointly on simulation and proxy models, we can highlight two types of model combination:

- The first is a sequential method. First a good starting point x_0 for the boundary conditions is sought through

TABLE 8. Proxy model approach.

	$q_{t,gas}^{max}$	z_{t,CO_2}^{max}	$\mathbf{x}_0$	$q_{t,gas}$	$z_{t,gas}^{CO_2}$	$f(\boldsymbol{\theta})$	f. eval.	time [s]	γ [%]	feasible
C1	5.663	2.50	BASE	5.51	2.3	59728	315	300.87	0	no
				5.65	1.91	59799	233	314.98	1	no
				5.48	1.66	**59728◊**	329	356.45	2	**yes**
			PRE$_{gas}$	5.38	1.48	59728	141	203.56	0	no
				5.48	2.18	59728	271	318.56	1	no
				5.51	1.42	59728	153	227.89	2	no
				4.89	1.61	59653	472	425.18	3	no
				5.41	2.01	**59728◊**	494	494.09	4	**yes**
			POS$_{gas}$	5.26	2.21	59728	130	261.86	0	no
				5.41	1.67	59799	232	233.32	1	no
				5.45	1.76	59728	334	355.99	2	no
				5.53	1.77	**59728◊**	480	155.33	3	yes
			PRE$_{CO2}$	5.45	1.86	59799	204	257.07	0	no
				5.42	2.11	59728	228	264.24	1	no
				5.27	1.73	59728	242	257.47	2	no
				4.93	1.87	59709	405	386.71	3	no
				5.47	1.63	**59728◊**	318	393.44	4	yes
			POS$_{CO2}$	5.4	2.11	59799	247	316.43	0	no
				5.59	1.64	59728	444	419.09	1	no
				5.38	2.19	59728	226	256.86	2	no
				5.32	1.99	59728	227	333.68	3	no
				5.45	1.95	59715	354	369.34	4	no
				5.11	1.89	59728	230	305.39	5	no
C2	3.681	2.50	BASE	3.62	2.23	**59346**	427	368.42	0	**yes**
			PRE$_{gas}$	3.68	2.27	**59329**	447	380.78	0	**yes**
			POS$_{gas}$	3.66	2.12	59385	377	349.62	0	no
				3.67	2.24	**59388**	439	373.20	1	**yes**
			PRE$_{CO2}$	3.68	2.28	58820	562	472.02	0	no
				3.68	2.28	58820	562	483.27	1	no
				3.67	2.20	**59388◊**	326	293.86	2	**yes**
			POS$_{CO2}$	3.67	2.3	**59392◊**	566	477.35	0	**yes**
C3	5.663	1.25	BASE	5.63	1.05	**55196**	267	282.34	0	**yes**
			PRE$_{gas}$	5.61	1.06	**56733◊**	443	397.74	0	**yes**
			POS$_{gas}$	5.61	1.06	**56436**	100	178.23	0	**yes**
			PRE$_{CO2}$	5.58	1.04	**56136◊**	202	238.72	0	**yes**
			POS$_{CO2}$	5.33	1.06	**56136◊**	136	231.95	0	**yes**

◊ - feasible solution better than benchmark
Gas flow rates are given in MSm^3/d.
Oil flow rates ($f(\boldsymbol{\theta})$) are given in Sm^3/d.

TABLE 9. Sequential model optimization (proxy models are used to obtain an initial guess $\mathbf{x}_0$ for simulation-based optimization).

	$q_{t,gas}^{max}$	z_{t,CO_2}^{max}	$\mathbf{x}_0$	$q_{t,gas}$	$z_{t,gas}^{CO_2}$	$f(\boldsymbol{\theta})$	f. eval.	time [s]
	5.663	2.50	BASE	5.45	1.65	59728◊	124	1202.52
	5.663	2.50	PRE$_{gas}$	5.41	2.01	59728◊	67	643.19
C1	5.663	2.50	POS$_{gas}$	5.53	1.77	59728◊	104	998.59
	5.663	2.50	PRE$_{CO2}$	5.47	1.63	59728◊	80	771.73
	5.663	2.50	POS$_{CO2}$	*no feasible start point*				
	3.681	2.50	BASE	3.68	2.12	59374◊	228	2179.61
	3.681	2.50	PRE$_{gas}$	3.68	2.34	59357	272	2566.66
C2	3.681	2.50	POS$_{gas}$	3.68	2.24	59388	122	1211.24
	3.681	2.50	PRE$_{CO2}$	3.67	2.29	59395◊	165	1651.57
	3.681	2.50	POS$_{CO2}$	3.67	2.34	59392◊	186	1760.86
	5.663	1.25	BASE	5.52	1.04	55670	130	1252.98
	5.663	1.25	PRE$_{gas}$	5.61	1.06	56733◊	109	997.80
C3	5.663	1.25	POS$_{gas}$	5.65	1.06	56736	175	1558.91
	5.663	1.25	PRE$_{CO2}$	5.65	1.06	56736◊	146	1298.47
	5.663	1.25	POS$_{CO2}$	5.33	1.06	56136◊	196	1738.20

◊ - better than benchmark
Gas flow rates are given in MSm^3/d.
Oil flow rates ($f(\boldsymbol{\theta})$) are given in Sm^3/d.

TABLE 10. Combined proxy and simulation based approach.

	$q_{t,\text{gas}}^{\max}$	$z_{t,\text{CO}_2}^{\max}$	$\mathbf{x}_0$	$q_{t,\text{gas}}$	$z_{t,\text{gas}}^{\text{CO}_2}$	$f(\boldsymbol{\theta})$	f. eval.	time [s]
	5.663	2.50	BASE	5.47	1.43	59645$^\diamond$	273	3684.74
	5.663	2.50	PRE$_{\text{gas}}$	4.91	1.76	59727	254	3379.24
C1	5.663	2.50	POS$_{\text{gas}}$	5.46	1.73	57496	204	2701.63
	5.663	2.50	PRE$_{\text{CO2}}$	5.26	1.38	59342$^\diamond$	458	5905.39
	5.663	2.50	POS$_{\text{CO2}}$	5.33	1.86	59728$^\diamond$	252	3384.36
	3.681	2.50	BASE	3.67	2.00	59312	395	5451.44
	3.681	2.50	PRE$_{\text{gas}}$	3.67	2.27	59438$^\diamond$	404	5816.03
C2	3.681	2.50	POS$_{\text{gas}}$	3.68	2.27	58505	269	3718.81
	3.681	2.50	PRE$_{\text{CO2}}$	3.66	2.24	57161	263	3451.77
	3.681	2.50	POS$_{\text{CO2}}$	3.66	2.23	59440$^\diamond$	535	7617.83
	5.663	1.25	BASE	5.60	1.02	55670	237	3107.56
	5.663	1.25	PRE$_{\text{gas}}$	5.50	1.05	54723	316	4252.38
C3	5.663	1.25	POS$_{\text{gas}}$	5.62	1.05	56436	209	2596.08
	5.663	1.25	PRE$_{\text{CO2}}$	5.60	1.05	56136$^\diamond$	247	3274.78
	5.663	1.25	POS$_{\text{CO2}}$	5.64	1.04	55670$^\diamond$	286	4016.61

$\diamond$ - better than benchmark

Gas flow rates are given in MSm^3/d.

Oil flow rates ($f(\boldsymbol{\theta})$) are given in Sm^3/d.

the proxy model, which has a relatively shorter computational time. Then, a simulation-based optimization is invoked from this starting point.

- The second is a concurrent method, in which the proxy model is used to assist the simulation-based optimization. The surrogates are used by the algorithm to dynamically adapt parameters, perform local search, and identify promising candidate solutions [41], [42].

Table 9 reports the results of the sequential method, whereby the solution obtained with the proxy model formulation is used as an initial guess for the simulation-based approach. Despite being initialized with a relatively good starting point $\mathbf{x}_0$, the algorithm performs several iterations until converging to a local solution. In this approach, optimal points from the proxy model that violate the base restrictions are also considered, leaving the task to recover feasibility to the DFO algorithm coupled with the simulation model.

Table 10 shows the results obtained with the DFO algorithm using jointly proxy and simulation models. It can be seen that the proxy model can aid the optimizer to achieve better results than the benchmark, taking advantage of the smoothness of the proxy model to get out of locally optimal points on the grid. An advantage of this approach is the guarantee of feasibility at termination, since the algorithm optimizes the simulation model. However, the combined method incurs additional computational time.

Although this approach can improve the benchmark results, it can have a detrimental effect on the objective in some cases. Adding this fact to the higher computational time, the search for a starting point using the pure proxy model may become more appropriate than the simultaneous mode in this particular case study.

VI. CONCLUSION

This work developed proxy models of gas pipeline networks for fast computation of pressures at inlet nodes, and gas flow rate and CO_2 content at the outlet node. For pressure, the proxy model was derived from the system identification methodology, while taking into account the governing equations of gas flow into the model's structure. For CO_2 and gas flow, simple algebraic equations based on mass conservation and proportions provided good approximations. Using the resulting proxy models, the variables of interest can be calculated far more efficiently when compared to a high-fidelity simulator, given the required boundary conditions at the nodes that represent production platforms and terminals.

The use of the proxy models with a DFO algorithm proved to be effective. Characterized by the gas export rate and CO_2 content, the operating conditions for production platforms were optimized to maximize overall oil production, while ensuring pressure bounds at platforms, and gas rate and CO_2 limits at the terminal. However, a suitable safety margin on pressure bounds was introduced to compensate for the prediction error of the proxy model. Besides optimizing exceedingly faster than with the direct use of the simulator, the DFO algorithm converged more consistently to solutions of higher production.

For future works, different methods of DFO could be employed and compared such as trust-region and augmented Lagrangian algorithms. Additionally, other characteristics could be considered in the model, such as moisture, contaminants besides CO_2, and multiphase flow.

REFERENCES

[1] L. S. Assis, E. Camponogara, B. C. Menezes, and I. E. Grossmann, "An MINLP formulation for integrating the operational management of crude oil supply," *Comput. Chem. Eng.*, vol. 123, pp. 110–125, Apr. 2019.

[2] C. T. da Costa Fraga, A. C. Capeleiro Pinto, C. C. M. Branco, and C. A. da Silva Paulo, "Brazilian pre-salt: An impressive journey from plans and challenges to concrete results," in *Proc. Offshore Technol. Conf.*, Houston, TX, USA, May 2015, p. 25710.

[3] R. Rocha, I. E. Grossmann, and M. V. S. P. de Aragão, "Petroleum supply planning: Reformulations and a novel decomposition algorithm," *Optim. Eng.*, vol. 18, no. 1, pp. 215–240, Mar. 2017.

[4] E. Camponogara, A. F. Teixeira, E. O. Hulse, T. L. Silva, S. Sunjerga, and L. K. Miyatake, "Integrated methodology for production optimization from multiple offshore reservoirs in the Santos Basin," *IEEE Trans. Autom. Sci. Eng.*, vol. 14, no. 2, pp. 669–680, Apr. 2017.

[5] B. A. Foss and J. P. Jensen, "Performance analysis for closed-loop reservoir management," *SPE J.*, vol. 16, no. 1, pp. 183–190, Mar. 2011.

[6] A. B. Huseby and N. F. Haavardsson, "Multi-reservoir production optimization," *Eur. J. Oper. Res.*, vol. 199, no. 1, pp. 236–251, Nov. 2009.

[7] Ø. Klokk, P. F. Schreiner, A. Pagès-Bernaus, and A. Tomasgard, "Optimizing a CO_2 value chain for the Norwegian continental shelf," *Energy Policy*, vol. 38, no. 11, pp. 6604–6614, Nov. 2010.

[8] A. Selot, L. K. Kuok, M. Robinson, T. L. Mason, and P. I. Barton, "A short-term operational planning model for natural gas production systems," *AIChE J.*, vol. 54, no. 2, pp. 495–515, Feb. 2008.

[9] F. Rømo, A. Tomasgard, L. Hellemo, M. Fodstad, B. H. Eidesen, and B. Pedersen, "Optimizing the Norwegian natural gas production and transport," *Interfaces*, vol. 39, no. 1, pp. 46–56, Feb. 2009.

[10] A. H. Ibrahim, P. Kumam, and W. Kumam, "A family of derivative-free conjugate gradient methods for constrained nonlinear equations and image restoration," *IEEE Access*, vol. 8, pp. 162714–162729, 2020.

[11] L. Li, A. F. Saldivar, Y. Bai, Y. Chen, Q. Liu, and Y. Li, "Benchmarks for evaluating optimization algorithms and benchmarking MATLAB derivative-free optimizers for Practitioners' rapid access," *IEEE Access*, vol. 7, pp. 79657–79670, 2019.

[12] W. Li, M. Assaad, and S. Zheng, "Distributed derivative-free learning method for stochastic optimization over a network with sparse activity," *IEEE Trans. Autom. Control*, vol. 67, no. 5, pp. 2221–2236, May 2022.

[13] R. Huang, Y. Chen, T. Yin, X. Li, A. Li, J. Tan, W. Yu, Y. Liu, and Q. Huang, "Accelerated derivative-free deep reinforcement learning for large-scale grid emergency voltage control," *IEEE Trans. Power Syst.*, vol. 37, no. 1, pp. 14–25, Jan. 2022.

[14] D. E. Ciaurri, T. Mukerji, and L. J. Durlofsky, "Derivative-free optimization for oil field operations," in *Computational Optimization and Applications in Engineering and Industry*. Cham, Switzerland: Springer, 2011, pp. 19–55.

[15] M. Khoshkbarchi, M. Rahmanian, J. Cordazzo, and L. Nghiem, "Application of mesh adaptive derivative-free optimization technique for gas-lift optimization in an integrated reservoirs, wells, and facilities modeling environment," in *Proc. SPE Canada Heavy Oil Conf.*, Sep. 2020.

[16] T. L. Silva, M. C. Bellout, C. Giuliani, E. Camponogara, and A. Pavlov, "Derivative-free trust region optimization for robust well control under geological uncertainty," *Comput. Geosci.*, vol. 26, no. 2, pp. 329–349, Apr. 2022.

[17] E. Camponogara, A. S. Machado, T. L. Silva, C. M. Giuliani, B. F. Vieira, and F. A. Teixeira, "Derivative-free optimization of offshore production platforms sharing a subsea gas network," *IFAC-PapersOnLine*, vol. 51, no. 8, pp. 185–190, 2018.

[18] L. O. Seman, L. K. Miyatake, E. Camponogara, C. M. Giuliani, and B. F. Vieira, "Derivative-free parameter tuning for a well multiphase flow simulator," *J. Petroleum Sci. Eng.*, vol. 192, Sep. 2020, Art. no. 107288.

[19] C. M. Giuliani, E. Camponogara, and A. R. Conn, "A derivative-free exact penalty algorithm: Basic ideas, convergence theory and computational studies," *Comput. Appl. Math.*, vol. 41, no. 1, pp. 1–36, Feb. 2022.

[20] S. F. Stefenon, L. O. Seman, C. S. F. Neto, A. Nied, D. M. Seganfredo, F. G. D. Luz, P. H. Sabino, J. T. González, and V. R. Q. Leithardt, "Electric field evaluation using the finite element method and proxy models for the design of stator slots in a permanent magnet synchronous motor," *Electronics*, vol. 9, no. 11, p. 1975, 2020.

[21] S. F. Stefenon, L. O. Seman, B. A. Pavan, R. G. Ovejero, and V. R. Q. Leithardt, "Optimal design of electrical power distribution grid spacers using finite element method," *IET Gener., Transmiss. Distrib.*, vol. 16, no. 9, pp. 1865–1876, May 2022.

[22] A. Hoffmann and M. Stanko, "Short-term model-based production optimization of a surface production network with electric submersible pumps using piecewise-linear functions," *J. Petroleum Sci. Eng.*, vol. 158, pp. 570–584, Sep. 2017.

[23] R. K. Ahuja, T. L. Magnanti, and J. B. Orlin, *Network Flows: Theory, Algorithms, and Applications*. Upper Saddle River, NJ, USA: Prentice-Hall, 1993.

[24] A. Codas, S. Campos, E. Camponogara, V. Gunnerud, and S. Sunjerga, "Integrated production optimization of oil fields with pressure and routing constraints: The Urucu field," *Comput. Chem. Eng.*, vol. 46, pp. 178–189, Nov. 2012.

[25] B. Grimstad, B. Foss, R. Heddle, and M. Woodman, "Global optimization of multiphase flow networks using spline surrogate models," *Comput. Chem. Eng.*, vol. 84, pp. 237–254, Jan. 2016.

[26] T. L. Silva, E. Camponogara, A. F. Teixeira, and S. Sunjerga, "Modeling of flow splitting for production optimization in offshore gas-lifted oil fields: Simulation validation and applications," *J. Petroleum Sci. Eng.*, vol. 128, pp. 86–97, Apr. 2015.

[27] E. Camponogara and A. M. D. Conto, "Lift-gas allocation under precedence constraints: MILP formulation and computational analysis," *IEEE Trans. Autom. Sci. Eng.*, vol. 6, no. 3, pp. 544–551, Jul. 2009.

[28] C. Borraz-Sánchez and D. Haugland, "Optimization methods for pipeline transportation of natural gas with variable specific gravity and compressibility," *TOP*, vol. 21, no. 3, pp. 524–541, Oct. 2013.

[29] *PIPESIM Steady-State Multiphase Flow Simulator*, Schlumberger, Houston, TX, USA.

[30] *MATLAB, Version R2018b*, The MathWorks Inc., Natick, MA, USA, 2018.

[31] O. Nelles, *Nonlinear System Identification: From Classical Approaches to Neural Networks and Fuzzy Models*. Cham, Switzerland: Springer, 2001.

[32] A. Afram and F. Janabi-Sharifi, "Black-box modeling of residential HVAC system and comparison of gray-box and black-box modeling methods," *Energy Buildings*, vol. 94, pp. 121–149, May 2015.

[33] M. Raissi, P. Perdikaris, and G. E. Karniadakis, "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations," *J. Comput. Phys.*, vol. 378, pp. 686–707, Feb. 2019.

[34] J. P. Vielma, S. Ahmed, and G. Nemhauser, "Mixed-integer models for nonseparable piecewise-linear optimization: Unifying framework and extensions," *Oper. Res.*, vol. 58, no. 2, pp. 303–315, Apr. 2010.

[35] T. L. Silva and E. Camponogara, "A computational analysis of multi-dimensional piecewise-linear models with applications to oil production optimization," *Eur. J. Oper. Res.*, vol. 232, no. 3, pp. 630–642, Feb. 2014.

[36] R. L. C. Beltrão, C. L. Sombra, A. C. V. M. Lage, J. R. F. Netto, and C. D. C. Henriques, "Challenges and new technologies for the development of the pre-salt cluster, Santos Basin, Brazil," in *Proc. Offshore Technol. Conf. (OTC)*, Houston, TX, USA, May 2009.

[37] M. A. Abramson, C. Audet, J. E. Dennis, and S. L. Digabel, "OrthoMADS: A deterministic MADS instance with orthogonal directions," *SIAM J. Optim.*, vol. 20, no. 2, pp. 948–966, Jan. 2009.

[38] C. Audet, A. L. Custódio, and J. E. Dennis, "Erratum: Mesh adaptive direct search algorithms for constrained optimization," *SIAM J. Optim.*, vol. 18, no. 4, pp. 1501–1503, Jan. 2008.

[39] A. R. Mello, J. de Matos, M. R. Stemmer, A. de Souza Britto Jr., and A. L. Koerich, "A novel orthogonal direction mesh adaptive direct search approach for SVM hyperparameter tuning," 2019, *arXiv:1904.11649*.

[40] S. Le Digabel, "Algorithm 909: NOMAD: Nonlinear optimization with the MADS algorithm," *ACM Trans. Math. Softw.*, vol. 37, no. 4, pp. 1–15, Feb. 2011.

[41] A. J. Booker, J. E. Dennis, P. D. Frank, D. B. Serafini, V. Torczon, and M. W. Trosset, "A rigorous framework for optimization of expensive functions by surrogates," *Struct. Optim.*, vol. 17, no. 1, pp. 1–13, Feb. 1999.

[42] C. Audet and D. Orban, "Finding optimal algorithmic parameters using derivative-free optimization," *SIAM J. Optim.*, vol. 17, no. 3, pp. 642–664, Jan. 2006.

CARLOS LUGUESI received the degree in technical-professionalizing (electronics technician) and the degree in industrial electrical engineering emphasis electronic from the Federal Technological University of Paraná, in 2000 and 2005, respectively. He is currently an Automation Engineer with Holic Automação Industrial. He has experience in electrical engineering, with emphasis on industrial electronics, and electronic systems and controls.

EDUARDO CAMPONOGARA (Member, IEEE) received the Ph.D. degree in electrical and computer engineering from Carnegie Mellon University, Pittsburgh, PA, USA, in 2000. He was a Postdoctoral Fellow at the Institute for Complex Engineered Systems, USA. He joined as a Faculty Member with the Department of Automation and Systems Engineering, Federal University of Santa Catarina, Brazil, in 2002. His research interests include systems optimization, distributed decision making, and traffic control engineering.

LAIO ORIEL SEMAN received the Ph.D. degree in electrical engineering from the Federal University of Santa Catarina, in 2017. He is currently a University Professor with the University of Vale do Itajaí, Brazil. His research interests include strategies for static and dynamic optimization, along with applications in traffic control, cyber-physical systems, and oil and gas production systems.

JOSÉ TORREBLANCA GONZÁLEZ is currently a Lecturer of electronic technology with the Department of Applied Physics, University of Salamanca. He worked as a Technical Engineer in the design, assembly, and repair of all types of refrigeration systems, from the assembly of industrial air conditioners to the assembly of freezing and conservation centers at private enterprise for two years. He has extensive knowledge in the operation of power electronics circuits and works on several research projects. One is dedicated to renewable energies, especially photovoltaic installations, another project is involved in the development and creation of test benches to study the behavior of photovoltaic modules in real sunlight. On the other hand, he is dedicated to intelligent devices and the study of sensors that can be used to help people, more specifically patients with different types of illnesses such as diabetes. Currently, he also dedicates another part of his time to the study of energy efficiency and implementation of systems with lower consumption and higher efficiency.

VALDERI REIS QUIETINHO LEITHARDT (Senior Member, IEEE) received the Ph.D. degree in computer science from INF-UFRGS, Brazil, in 2015. He is currently a Professor with the Polytechnic Institute of Portalegre and a Researcher integrated with the VALORIZA—Research Centre for Endogenous Resource Valorization. He is also a Collaborating Researcher with the Expert Systems and Applications Laboratory (ESALab), University of Salamanca, Spain. His research interests include distributed systems with a focus on data privacy, communication, and programming protocols, involving scenarios and applications for the Internet of Things, smart cities, big data, cloud computing, and blockchain.

• • •