Proceedings World Geothermal Congress 2026
Calgary, Canada, June 8 – 11, 2026

Increasing Information Content in Geothermal Production Data Case Study from Qinghe, China

Egill Júliusson[1,2*], Andrés Laverde[1], Tingting Zheng[1], Li Ming[1,3], Susan Sun[1,3]
Arctic Green Energy, Kopavogur, Iceland[1*], Reykjavik University, Reykjavik, Iceland[2], Sinopec Green Energy, Xiong'an, China[3]
egill@arcticgreen.com

Keywords: Reservoir modelling, calibration, optimization, flow rate scheduling, inter-well connectivity

INTRODUCTION

Geothermal reservoir models often suffer from lack of proper production data for characterization of key features such as the influence that production or injection in one well has on its neighboring production (or injection) wells. Some reasons for this may be the lack of monitoring (pressure or water-level, flowrate and temperature), improper installation of sensors, inability to account for wellbore, ESP or other equipment effects or noise, poor data curation practices, and a fundamental deficiency in the information content contained in the collected data. The last-named item will be the main topic of this paper.

1 BACKGROUND

Arctic Green Energy, along with it's subsidiary, Sinopec Green Energy, have been working on reservoir modelling of low-temperature (single phase liquid) sedimentary geothermal systems in Northern China. Production from these systems generally occurs throughout the winter (~4 months) for space heating, and after that the wells are closed and the system allowed to recover until the next season. This means that the production data are such that all the wells (both production and injection) are turned on at once in a more-or-less synchronized step function. As a result, it is very difficult to resolve inter-well signals and gauge reservoir heterogeneity.

The Arctic Green Innovation Center is developing resource monitoring, characterization and control systems focusing on low-enthalpy geothermal reservoirs. One of the objectives set out for the creation of these tools was to find practical ways to guide operators towards generating more informative production signals. The methodology would focus on pressure responses to changes in injection and production rates. and to validate this work a pressure signal test was designed and implemented in the Qinghe geothermal field in China.

2 METHODOLOGY

In this section we will discuss the methodology that was developed to assess information content and design practical flowrate scheduling strategies for low-temperature (liquid dominated) geothermal systems. We start by describing the governing equations used for the reservoir model, simplifying assumptions and possible additions. This is followed by a section on how this model can be represented as a state-space system, which are commonly used in control systems engineering, aerospace engineering and other engineering disciplines. With further simplifications and the model calibration process turns into a regression problem that can be related to the Fisher Information Matrix (FIM) and information content.

2.1 Reservoir Model Basics

The Qinghe reservoir is a relatively uniform sandstone basin found at approximately 1.3 km depth below Qinghe city. The goal of our study was to construct a model that would remain as simple as possible, yet capture some key elements to be able predict and control well operations:

- injection and production in active wells
- drawdown in and near active wells
- drawdown in the neighboring (either active or inactive) wells
- long-term drawdown in the system
- pressure boundary interactions.

The model focused on pressure response and the temperature of the fluid in the system was ignored to simplify the mathematics. As the temperature of fluid produced from the reservoir is only around 60-65°C and the

reinjection temperature is close to 30°C, the errors associated with this assumption should be quite limited[1]. Moreover, the sedimentary basin is relatively flat and therefore the gravity effects were ignored from block to block within the reservoir.

In line with the reasoning above a generalized block (or tank) model was constructed based on a set of mass balance equations:

$$\frac{\partial(\rho_f V\phi)_i}{\partial t} = \sum_{j\in J_i} \rho_f q_{ij} + \sum_{b\in B_i} \rho_f q_{ib} + \sum_{k\in K_i} \rho_f \lambda_{ik} q_k \tag{1}$$

Here ρ_f is the density of the fluid, V_i is the volume of grid block i, ϕ_i is the porosity of grid block i and q_k is the volumetric inflow in from external source k (e.g. a well). Grid block i is also assumed to be connected to other blocks as specified by the set J_i and q_{ij} denotes the volumetric inflow from the other blocks. Similarly, the set B_i denotes the boundaries that the block connects to. Finally, the set K_i contains the sources (and sinks) connected to the block and λ_{ik} denotes the fraction of the flow from these sources that goes into the block.

2.2 State-Space Representation

Assuming the rock and fluid are slightly compressible, and doing some algebraic manipulation, the expression given by equation (1) can be transformed into a system of matrix equations with the following form:

$$\boldsymbol{S\dot{p} = Tp + Vp_b + \Lambda q} \tag{2}$$

In this equation the $\boldsymbol{S}$ is a diagonal matrix of storage coefficients, $\boldsymbol{T}$ is the block-to-block transmissibility matrix which is symmetric and diagonally dominant (i.e., invertible). The transmissibility values for the boundary connections are stored in the sparse matrix $\boldsymbol{V}$, where the boundary connections of the blocks determine which values are non-zero. The distribution of source flows between blocks is stored in the sparse matrix $\boldsymbol{\Lambda}$, which also has non-zero values corresponding to how the sources are connected to each block. The column vectors $\boldsymbol{\dot{p}}$, $\boldsymbol{p}$, $\boldsymbol{q}$, and $\boldsymbol{p_b}$ represent the time derivative of block pressure, the block pressure, volumetric inflow into each block, and boundary pressure values at each connection point, respectively.

The storage coefficient of each block is always non-zero and therefore the storage coefficient matrix is invertible. This means that equation (2) can be rewritten in the form:

$$\begin{aligned}\boldsymbol{\dot{p}}(t) &= \boldsymbol{S^{-1}Tp}(t) + \boldsymbol{S^{-1}\Lambda q}(t) + \boldsymbol{S^{-1}Vp_b} \\ &= \boldsymbol{Ap}(t) + \boldsymbol{Bq}(t) + \boldsymbol{Ep_b}\end{aligned} \tag{3}$$

The observation of pressure is in the wellbore (not inside the grid blocks themselves), often at a depth significantly different from the reservoir (usually water-level is being observed as opposed to sand-face pressure). Therefore, a second formulation is used to represent the observed pressure values. A general format, considering the possible inclusion of gravity and frictional pressure losses inside the well, would be of the form:

$$\boldsymbol{p_o}(t) = \boldsymbol{Cp}(t) + \boldsymbol{G} + \boldsymbol{Dq}(t) + \boldsymbol{Fq}^2(t) \tag{4}$$

where $\boldsymbol{C}$ denotes which pressure states (blocks) are being observed, $\boldsymbol{G}$ represents the gravitational weight of the water column between the observation depth and the reservoir depth, $\boldsymbol{D}$ represents the effects of the productivity (or injectivity) index (similar to skin effect), and $\boldsymbol{F}$ accounts for frictional pressure losses in the wellbore.

In these equations $\boldsymbol{A}$ is an N_T-by-N_T matrix, $\boldsymbol{B}$ is an N_T-by-N_S matrix, $\boldsymbol{E}$ is an N_T-by-N_B matrix, $\boldsymbol{C}$ is an N_O-by-N_T matrix of 1s and 0s and $\boldsymbol{D}$ and $\boldsymbol{F}$ would be N_O-by-N_S matrices of well connection parameters, and $\boldsymbol{G}$ is an N_O-by-1 vector. Here N_B is the number of boundary connections, N_S is the number of sources (wells). N_O is the number of observations and N_T is the number of blocks in the model.

These equations were set up in a computer code that starts by creating a grid around the wells, and transforms a given initial estimate of the rock and well properties into the set defined by equations (3) and (4). The observed

[1] The density changes are very small, but viscosity does change significantly over this temperature range. Fracture permeability may also change due to thermal contraction. The effects would be most apparent in the near wellbore regions of the injection wells where the increased viscosity of the colder water will decrease the transmissibility, but fracture dilation would increase it.

pressure values can then be simulated with a very efficient simulation code. Each run, for our test system took a fraction of a second to compute, with 1000 timesteps and more than 80 parameters.

With this general formulation the work of calibrating the model to production histories and optimizing flow rate schedules has become a tractable automation task. Further discussion on that will be given in future publications.

2.3 Information Content and the Fisher Information Matrix

As discussed earlier, the ESPs for many operators were run at maximum capacity throughout the heating season. This is not ideal because it wastes electricity for pumping when heat demand is low, and it gives limited information about the properties of the geothermal reservoir, e.g. inter-well connectivity.

The following strategy has been suggested for obtaining more informative data to identify reservoir properties. The methodology could be implemented in conjunction with the flowrate optimization, where a balance is struck between changing flow rates to minimize pumping cost, and collecting more informative reservoir data.

2.3.1 Estimating Information Content – Fisher Information Matrix (FIM)

A well-known approach for estimating information content is through the Fisher Information Matrix. Consider a linear time-invariant state-space system of the form:

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{A}x(t) + \boldsymbol{B}\boldsymbol{u}(t) + \boldsymbol{E}\boldsymbol{d}(t), \tag{5}$$

$$\boldsymbol{y}(t) = \boldsymbol{C}\boldsymbol{x}(t) + \boldsymbol{D}\boldsymbol{u}(t) + \boldsymbol{v}(t) \tag{6}$$

Where $\boldsymbol{x}(t)$ is the state vector, $\boldsymbol{u}(t)$ is the input vector (to be designed), $\boldsymbol{y}(t)$ is the measured output, $\boldsymbol{d}(t)$ is disturbance, and $\boldsymbol{v}(t)$ is measurement noise. The parameter set includes matrices $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$, $\boldsymbol{D}$, and $\boldsymbol{E}$ or a subset of these parameters. The analogy between this standard state-space notation and the notation used in equations (3) and (4) is clear, with a few exceptions. With this formulation, it is assumed that observations are made at the sand-face ($\boldsymbol{G}$ and $\boldsymbol{F}$ could have been evaluated separately from well test and used to convert water-level readings to readings at reservoir depth, or pressure readings are collected at reservoir depth). The boundary pressure fits into the form here as a disturbance and a noise feature is added for the measured output signal.

Given input signals $u(t)$, the Fisher Information Matrix (FIM) is defined as:

$$\boldsymbol{\mathcal{I}}(\boldsymbol{\theta}) = E\left\{ \left(\frac{\partial}{\partial \boldsymbol{\theta}} \ln p(\boldsymbol{y}|\boldsymbol{\theta},\boldsymbol{u})\right)\left(\frac{\partial}{\partial \boldsymbol{\theta}} \ln p(\boldsymbol{y}|\boldsymbol{\theta},\boldsymbol{u})\right)^{\top} \right\} \tag{7}$$

where:

- $p(\boldsymbol{y}|\boldsymbol{\theta},\boldsymbol{u})$ is the likelihood of the observed outputs y given parameters $\boldsymbol{\theta}$ and inputs $\boldsymbol{u}$.
- $\boldsymbol{\theta}$ represents the unknown system parameters (elements of $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$, $\boldsymbol{D}$, and $\boldsymbol{E}$).

For Gaussian noise and disturbances, this typically simplifies into:

$$\boldsymbol{\mathcal{I}}(\boldsymbol{\theta}) = \sum_{t=1}^{N} \left(\frac{\partial \hat{\boldsymbol{y}}(t|\boldsymbol{\theta},\boldsymbol{u})}{\partial \boldsymbol{\theta}}\right)^{\top} \boldsymbol{R}^{-1} \left(\frac{\partial \hat{\boldsymbol{y}}(t|\boldsymbol{\theta},\boldsymbol{u})}{\partial \boldsymbol{\theta}}\right) \tag{8}$$

where $\hat{\boldsymbol{y}}(t|\boldsymbol{\theta},\boldsymbol{u})$ is the model-predicted output and $\boldsymbol{R}$ is the covariance matrix of the measurement noise.

The larger the determinant (or eigenvalues) of the Fisher Information Matrix, the more information the experiment provides about the parameters.

2.3.2 Evaluating Input Signals

A practical way to evaluate the information content of the input signals starts by computing the sensitivity functions:

$$\boldsymbol{S}_{\theta}(t) = \frac{\partial \hat{\boldsymbol{y}}(t|\boldsymbol{\theta},\boldsymbol{u})}{\partial \boldsymbol{\theta}} \tag{9}$$

These sensitivities can be numerically estimated by finite differences or analytically. The Fisher Information Matrix (FIM) is computed as:

$$\boldsymbol{\mathcal{I}}(\boldsymbol{\theta}) = \sum_t \boldsymbol{S}_\theta^\top(t)\boldsymbol{R}^{-1}\boldsymbol{S}_\theta(t) \tag{10}$$

And the information metric can be either the log of the determinant $\log[\det(\boldsymbol{\mathcal{I}}(\boldsymbol{\theta}))]$ or trace of $\mathrm{tr}(\boldsymbol{\mathcal{I}}^{-1}(\boldsymbol{\theta}))$. Larger determinants indicate larger confidence bounds for the estimates (D-optimality), while a low trace indicates a small average variance of all estimated parameters (A-optimality). Yet another metric would be the minimum eigenvalue of $\boldsymbol{\mathcal{I}}(\boldsymbol{\theta})$, which represents the most uncertain parameter combination (E-optimality).

3 QINGHE PRESSURE SIGNAL TEST

A pressure signal test was designed and carried out in the Qinghe sandstone reservoir to evaluate hydraulic connectivity and pressure interference through a coordinated multi-well interference experiment. Two production wells (FHQY1 and FHQY4) and two reinjection wells (FHQY2 and FHQY3) formed the active control group, while seven surrounding wells monitored pressure propagation. Preparatory work in August 2024 included instrument installation and calibration, with both offline pressure and temperature loggers and online sensors deployed at ~150 m depth to capture high-resolution water-level fluctuations. Flow changes were applied systematically between August 27 and September 27, 2024, creating eight distinct production-reinjection stages with rates ranging from 0 to 60 m^3/h. This controlled sequence generated measurable pressure responses across the field, enabling the characterization of well-to-well connectivity and near-surface pressure recovery trends.

The data revealed strong pressure variations in the active wells and weaker but detectable signals in observation wells, confirming lateral communication within the reservoir. Corrections were applied to account for frictional and hydrostatic effects, isolating the true reservoir response from wellbore influences. Subsequent well test analysis provided transmissivity, storativity, and skin factors, highlighting moderate formation permeability and significant near-wellbore pressure losses—particularly in reinjection wells where skin-related drops reached 3–4 bar at 60 m^3/h. The test demonstrated that wellbore friction and skin effects can mask actual reservoir behavior, underscoring the importance of downhole sensors or corrections and continuous flow monitoring. The test thus provided a robust dataset for calibrating numerical and network-based models, offering the first high-resolution view of pressure dynamics and connectivity in the Qinghe geothermal reservoir.

3.1 Evaluating Signal Information Content

The information content method described above was validated as part of a pressure signal test. After calibrating the model, it was clear that the transient state from one fixed flow rate configuration to the next was short lived (several hours). Therefore, the knowing transient part of the reservoir model would have limited value, at least for short term optimization of flow rate schedules. Thus, equation (3) would simplify to the steady-state version:

$$\mathbf{0} = \boldsymbol{T}\boldsymbol{p}_{\boldsymbol{ss}} + \boldsymbol{\Lambda}\boldsymbol{q}_{\boldsymbol{ss}} + \boldsymbol{V}\boldsymbol{p}_{\boldsymbol{b}} \tag{11}$$

Combining this with equation (4), after the observations have been converted to reservoir depth (using $\boldsymbol{F}$ and $\boldsymbol{G}$), yields the steady-state observation equation:

$$\boldsymbol{p}_{o,ss} = (\boldsymbol{D} - \boldsymbol{C}\boldsymbol{T}^{-1}\boldsymbol{\Lambda})\boldsymbol{q}_{\boldsymbol{ss}} - \boldsymbol{C}\boldsymbol{T}^{-1}\boldsymbol{V}\boldsymbol{p}_{\boldsymbol{b}} \tag{12}$$

It is assumed that the influence from the boundary is constant over time and therefore $-\boldsymbol{C}\boldsymbol{T}^{-1}\boldsymbol{V}\boldsymbol{p}_{\boldsymbol{b}}$ is equal to the undisturbed observation pressure, $\boldsymbol{p}_{o,\mathbf{0}}$. It can therefore be subtracted from the pressure observation vector, leading to an expression that evaluates the relative pressure change, $\boldsymbol{x}_o$, due to a given flow rate configuration:

$$\boldsymbol{x}_{o,ss} = (\boldsymbol{D} - \boldsymbol{C}\boldsymbol{T}^{-1}\boldsymbol{\Lambda})\boldsymbol{q}_{\boldsymbol{ss}} = \boldsymbol{\Psi}\boldsymbol{q}_{\boldsymbol{ss}} \tag{13}$$

This simple linear system simulates the steady-state pressure around the wells as a function of fixed production and injection rates from each well. The $\boldsymbol{\Psi}$ matrix maps a unit change in flow rate in each control well to a pressure change at an observation well. Knowing this relationship is useful in terms of application to field management for short-term (hourly to daily) scheduling of injection and production rates. With this simplification of the system, the sensitivity matrix, with respect to the individual elements of $\boldsymbol{\Psi}$, denoted by θ, can be computed from the Kronecker product:

$$\boldsymbol{S}_{\theta} = \boldsymbol{q}_{ss}^{T} \otimes \boldsymbol{I}_{\boldsymbol{N_o}} \tag{14}$$

where $\boldsymbol{I}_{\boldsymbol{N_o}}$ is the identity matrix of dimension N_o. Thus, the FIM (equation (10)) only depends on the input flowrates. Furthermore, the number of linearly independent input configurations must equal the number of active wells for $\boldsymbol{\mathcal{J}}(\boldsymbol{\theta})$ to have full rank (i.e. to fully characterize $\boldsymbol{\Psi}$). Importantly, this means that a test design cannot have a combination of perfectly balanced injection and production rates for all cases. The test design used for the Qinghe pressure signal test, along with the computed optimality measures, is presented in Table 1.

Table 1: Flowrate [m^3/h] configurations for the Qinghe pressure signal test along with the development of information metrics.

Config# / Well ID	1	2	3	4	5	6	7	8	9	10	11	12	13
FHQY#1	-60	-59	0	-24	-39	-43	0	0	0	0	0	0	0
FHQY#2	0	0	0	24	39	43	0	0	0	0	0	42	0
FHQY#3	0	59	0	0	0	0	0	40	25	8	0	0	0
FHQY#4	0	0	0	0	0	0	0	-40	-25	-8	0	-42	0
D-opt	-inf	-inf	-inf	-inf	-inf	-inf	-inf	320.0	323.3	323.6	323.6	333.6	333.6
A-opt*10^3	2.8	8.4	8.4	28.6	16.0	13.7	13.7	25.6	23.9	23.8	23.8	19.4	19.4
E-opt	0	0	0	0	0	0	0	558.7	589.4	591.6	591.6	658.1	658.1
rank($\mathcal{J}$)	10	20	20	30	30	30	30	40	40	40	40	40	40

The experimental design included fall-off periods (Config# 3, 7, 11, and 13) and which do not contribute to the identification of $\boldsymbol{\Psi}$, but were recorded to improve the identification of $\boldsymbol{A} = \boldsymbol{S}^{-1}\boldsymbol{T}$. Step-rate tests with the linearly dependent injection/production relationships (Config# 4-5-6 and 8-9-10) were also included for individual well testing but do not contribute much to the inter-well connectivity estimates. The test included 10 observation wells (including the 4 active wells) and therefore the FIM has full rank (40) when four linearly independent configurations have been tested (e.g. Config# 1, 2, 4, and 8). The optimality criteria have limited meaning before the full rank condition has been reached, but after that the criteria follow the expected paths (D and E optimality increase and A optimality decreases) when more information is added.

Overall, the results show that all the parameters for the connectivity matrix $\boldsymbol{\Psi}$ are identifiable (full-rank FIM) and each of the optimality measures indicate a well-designed test. The conditioning number for the FIM was also computed as 22 after the 13 configurations, indicating that the parameters have been uniformly excited and are relatively insensitive to noise in the data. In this case the noise covariance matrix $\boldsymbol{R}$ was ignored but should not change the overall test design given that the same type of sensor was used for each well.

4 CONCLUSION

The work presented shows how a generalized mass balance model can be used to model pressure variations in liquid dominated sedimentary geothermal systems. The model is converted to state-space form and further simplified to a linear well connectivity model. This model is useful for explaining how information content can be increased and evaluated with smart pressure signaling between wells.

This paper gives a simple introduction to information content and methods for optimal experimental design. Assessing information content for other parameters, adding thermal effects, test configurations, automated data curation and optimal flowrate scheduling are other methods being developed.

Acknowledgements

The authors gratefully acknowledge assistance and funding from Sinopec Green Energy.

REFERENCES

Juliusson, E., Laverde, A and Zheng, T.: Internal Project Report: Qinghe Sandstone Reservoir – Pressure Signal Test, *Arctic Green Energy Innovation Center,* Kopavogur, Iceland (2025).