https://doi.org/10.37965/jdmd.2026.1608

Fault Diagnosis of Electric Submersible Progressive Cavity Pump Systems Based on Multi-Physics Forward Modeling and Regularized Inversion

Fan Li[1] and Dechun Chen[1],*

[1] School of Petroleum Engineering, China University of Petroleum (East China), Qingdao 266580, China

*Corresponding author: chendc@upc.edu.cn (Dechun Chen)

Received 7 July , 2026; Revised 8 August 2026; Accepted 27 October, 2026; Published online 28 October 2026

Abstract: The electric submersible progressive cavity pump (ESPCP) is widely used in unconventional oil production, yet its downhole unit is subject to multiple progressive faults under coupled electromagnetic, fluid, and structural physics. Conventional monitoring methods often suffer from delayed detection and high false-alarm rates, while data-driven diagnosis requires large volumes of labeled fault samples that are rarely available. This paper reformulates fault diagnosis as a Tikhonov-regularized inverse problem, eliminating the need for historical fault data. A multi-physics forward model covering the cable, motor, coupling, pump, and tubing subsystems is constructed (fixed-point coupling of the electromechanical and hydraulic steady states, verified against the coupled transient dynamics), and 22 fault parameters are defined through explicit perturbation functions. Diagnosis is a two-stage pipeline: Stage A calibrates the healthy baseline of a 12-parameter reduced quasi-steady model from one normal-condition sweep; Stage B inverts each measured sweep with the L-curve-selected regularization and scores every fault signature direction by its restricted-fit evidence (the residual reduction of the best penalised 1-D fit along that direction, healthy projection subtracted), with a per-noise-level threshold calibrated on healthy draws. The monitoring vector comprises four surface channels (flow rate, supply current, wellhead pressure, electrical power) at 12 variable-speed-drive setpoints (m = 48). A 1,720-cell experiment (22 faults × 5 severities × 4 noise levels × 3 repetitions plus 400 healthy cells), aggregated from a single master results table, yields at 3% measurement noise a detection rate of 87.3% and a per-type classification accuracy of 41.2%, rising to 69.4% at the level of the measurably confusable fault groups, with a false-alarm rate of 17.5% at the 95th-percentile healthy threshold. The identifiability boundary of the four-channel quasi-steady measurement set is quantified (condition number 6.2e7; five weak parameter directions) and reported explicitly: the fault acting through the weakest parameter direction (bearing wear, through viscous friction B) is below the noise floor at moderate severity, as are the faults whose full-model effect is itself below the noise (backlash drag, tubing friction), and phase-attributed electrical faults collapse to their dq projection. A reduced-vs-full model discrepancy study over 135 operating-condition × fault-state combinations (27 conditions × 5 fault states) bounds the

https://doi.org/10.37965/jdmd.2026.1608

surrogate error. The method requires only normal-condition data for baseline calibration, demonstrating practical potential for field deployment.

Keywords: fault diagnosis; electric submersible progressive cavity pump; multi-physics modeling; regularized inversion; baseline calibration

1. Introduction

The electric submersible progressive cavity pump (ESPCP) integrates a downhole permanent magnet synchronous motor (PMSM) with a progressive cavity pump (PCP) through a coupling and protector assembly. Power is delivered from the surface via a three-phase armored cable extending through the wellbore [1,2]. Compared with conventional sucker-rod pumping systems, ESPCP systems offer the advantages of eliminating the rod string, reduced energy losses, suitability for deviated and horizontal wells, and strong adaptability to high-viscosity and sand-laden crude oils [3,4]. These characteristics have driven widespread adoption of ESPCP technology in unconventional oil reservoirs, particularly in China, Canada, and South America over the past decade [5].

Despite these advantages, ESPCP downhole units operate under harsh multi-physics conditions at depths typically exceeding 1,000 m. The electromagnetic field in the PMSM, the multiphase flow in the pump chambers and the tubing, and the structural contact field at the rotor-stator interface are strongly coupled; the wellbore temperature profile modulates the fluid and material properties and is carried as a prescribed input rather than solved. This coupling creates a complex fault landscape: stator wear caused by sand abrasion progressively enlarges the rotor-stator clearance, leading to increased internal leakage and declining volumetric efficiency; rubber swelling due to aromatic hydrocarbon absorption increases the interference fit and friction torque[21,22]; gas accumulation at the pump intake can cause gas locking, resulting in complete loss of production; wax deposition in the tubing reduces the effective flow area; and cable insulation degradation increases power losses and creates electrical safety hazards [6-9]. These faults typically initiate in subtle, progressive forms that are difficult to detect at early stages using conventional threshold-based monitoring.

Existing ESPCP fault diagnosis methods can be broadly classified into three categories. Empirical methods rely on operator experience and simple threshold comparisons [10], offering low cost but suffering from delayed detection, high false-alarm rates, and inability to handle compound faults. Data-driven methods have attracted significant recent attention: wavelet-package and Elman-network schemes [11], probabilistic neural networks [12], Elman-network time-series prediction [13], and random forests [14] have been reported for PCP and submersible screw-pump diagnosis; Liu et al. [15] developed a CNN-ResNet-RF fusion model with TimeGAN data augmentation achieving approximately 97% accuracy on 360 field records from Xinjiang Oilfield; and Ou et al. [16] employed contrastive learning with mixed-domain feature alignment for imbalanced SCADA datasets containing 27 fault cases. However, these data-driven approaches share three fundamental

https://doi.org/10.37965/jdmd.2026.1608

limitations: (i) they require large quantities of accurately labeled fault samples, which are extremely scarce and expensive to obtain in real oilfield operations; (ii) the diagnostic results lack physical interpretability, making it difficult for field engineers to trust and act upon the outputs; and (iii) models trained on data from specific wells exhibit limited transferability to wells with different configurations or operating conditions.

Mechanism-based simulation methods [17-19] offer the advantage of physical rigor but can only perform forward prediction—given fault parameters, they compute the resulting monitoring signals—and cannot inversely infer fault states from measured data. Furthermore, they do not account for the deviation between design parameters and actual operating parameters caused by manufacturing tolerances and installation variations.

To address these limitations, this paper reformulates ESPCP fault diagnosis as a Tikhonov-regularized inverse problem: given the measured monitoring vector d, find the fault-parameter vector θ such that the forward-model prediction m(θ) best matches the measurements while remaining close to a calibrated normal baseline. The approach eliminates the need for fault-labeled training data and provides physically interpretable diagnostic outputs. The principal contributions of this work are: (1) an implemented multi-physics forward model covering the cable, motor, coupling, pump and tubing subsystems, coupled through a fixed point of the electromechanical and hydraulic steady states and verified against the coupled transient dynamics, with 22 fault parameters defined by explicit perturbation functions (Table 2); (2) a 12-parameter reduced quasi-steady map for inversion (Table 4), with the L-curve-selected regularization and the singular-spectrum identifiability of the 48-component measurement set quantified explicitly (Section 3.3); (3) a two-stage diagnostic pipeline - Stage A calibrates the healthy baseline from normal-condition data, Stage B scores every fault signature direction by its restricted-fit evidence with per-noise-level thresholds calibrated on healthy draws (Section 3.4) - together with the measured fault-to-parameter mapping and the resulting confusion groups (Table 6); (4) a 1,720-case experiment (22 fault types × 5 severities × 4 noise levels × 3 repetitions plus 400 healthy cases, Section 4.1) reported from a single master results table, including the measured detection limits of the weakly identifiable faults (Section 4.3); and (5) a 135-condition study bounding the reduced-vs-full model discrepancy over the operating envelope (Section 4.9).

2. Multi-Physics Forward Modeling

2.1 System architecture and coupling structure

The ESPCP downhole unit (Fig. 1) comprises five functionally interconnected subsystems: (1) the three-phase armored cable (200 mm² Cu, 1300 m) delivering power from the surface variable-frequency drive (VFD); (2) the downhole permanent-magnet synchronous motor (PMSM, 16-pole, 8 pole pairs); (3) the universal-joint coupling with protector assembly; (4) the progressive-cavity pump (PCP) with four stages of 80 mm pitch; and (5) the production tubing (62 mm ID) carrying the oil-gas-water mixture to the wellhead. The four monitoring quantities used throughout this study are surface measurements: production flow rate, motor supply current, wellhead pressure, and electrical input power (Section 3.1, Table 3). The design parameters of the representative well are listed in Table 1.

https://doi.org/10.37965/jdmd.2026.1608

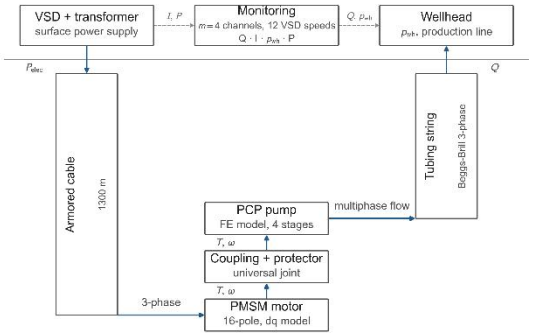


Fig. 1 ESPCP system architecture: five subsystems and four surface monitoring channels.

The forward model couples the electromagnetic, fluid and structural physics through the steady-state fixed point of the coupled electromechanical-hydraulic system (Fig. 2), solved at each VFD speed. The motor-side variables (supply current, electrical power) are computed from the shaft torque through the motor torque constant; the shaft torque is the sum of the pump hydraulic, friction, compression and contact torques transmitted through the coupling; the pump discharge pressure and surface flow rate are obtained from the tubing pressure traverse, which depends on the pump flow rate; and the pump flow rate itself is the difference between the theoretical displacement and the clearance leakage, which depends on the discharge pressure. The loop is closed by fixed-point iteration, which converges in 5-15 iterations at nine of the twelve sweep speeds; at the remaining three speeds the 60-iteration cap is reached with the residual already plateaued at 4-8·10^-9 (Fig. 3), i.e. the cap truncates a converged state. The steady-state limit is the stable fixed point of the coupled transient system (PMSM dq dynamics with PI speed control; Section 4.8 verifies this against a direct transient integration for all nine verification cases).

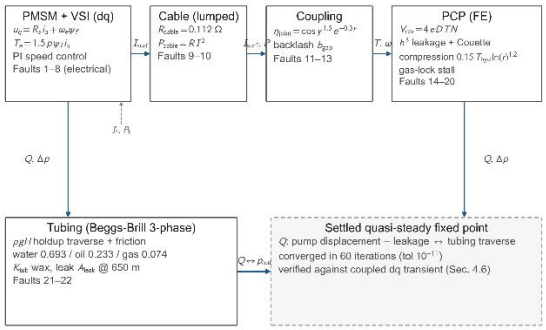


Fig. 2 Coupling structure of the implemented multi-physics forward model (steady-state fixed point of the coupled transient system).

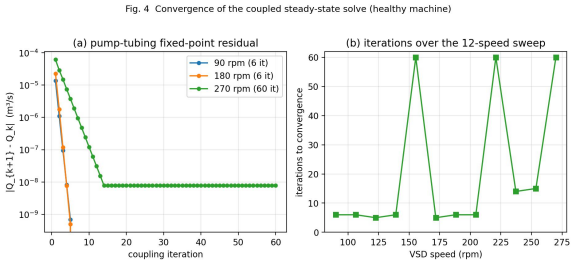


Fig. 3 Convergence of the coupled steady-state solve (healthy machine): (a) pump-tubing fixed-point residual at three speeds; (b) iterations to convergence over the 12-speed sweep.

The wellbore temperature enters the model only through the prescribed (measured) temperature profile used in the Beggs-Brill fluid property evaluation - a linear profile between the measured wellhead temperature 274.81 K and the bottom-hole temperature 320.05 K. A thermal solver (motor temperature) exists in the code family but is not part of the diagnostic forward model of this paper, because none of the four diagnostic channels is a temperature measurement.

https://doi.org/10.37965/jdmd.2026.1608

Table 1. Design parameters of the representative ESPCP well (code constants of the implemented forward model).

Item	Value
Well	depth 1300 m; tubing ID 62 mm; wellhead pressure 5 MPa; pump intake pressure 5 MPa
Fluid	oil density 850 kg/m3; viscosity 1.35 mPa.s; intake fractions water 0.693 / oil 0.233 / gas 0.074; oil gravity 0.8261; gas gravity 0.65; T_wh 274.81 K; T_bh 320.05 K
Cable	200 mm2 Cu, 1300 m, lumped R = 0.112 Ohm
Motor	PMSM, p = 8 pole pairs, Rs = 0.2 Ohm, Ld = Lq = 1 mH, psi_f = 0.1 Wb, J = 0.2 kg.m2, B = 0.05 N.m.s/rad, PI gains 0.1 / 0.5, VSI limit 200 V
Coupling	eta_c = 0.96, backlash half-width 5e-3 rad, fault resistance R_f = 1 Ohm
Pump	e = 2 mm, D = 20 mm, pitch = 80 mm, N = 4 stages, h0 = 40 um, L_seal = 0.32 m, mu_f = 0.1, P_MAX = 10 MPa, alpha_crit = 0.5, Q_ref = 1.2e-4 m3/s, dP_ref = 0.5 MPa, C_D = 0.6, tubing-leak depth 650 m
Monitoring	4 channels x 12 VFD speeds (90-270 rpm, uniform) = 48 components

2.2 Electromagnetic subsystem (PMSM + VSI)

The motor is modeled in the d-q rotating frame with surface-mounted magnets (L^d = L_q = 1 mH, Rs = 0.2 Ω, permanent-magnet flux linkage ψ_f = 0.1 Wb, p = 8 pole pairs). The torque-producing current follows from the electromagnetic torque

$$I_q = \frac{T_e}{1.5 p \psi_f}$$

$$V_{ph} = \sqrt{\left(R_s I_q\right)^2 + \left(\omega_e \psi_f\right)^2}$$

$$\omega_e = p\omega_m \quad (1)$$

and the phase voltage follows from the stator voltage equation; the VFD is a PI speed controller with output limit 0-200 V and gains Kp = 0.1, Ki = 0.5 (used in the transient verification, Section 4.8); the quasi-steady diagnostic model evaluates the settled limit, in which the motor speed equals the VFD setpoint. The motor power is P_motor = T_motor·ω_m + 1.5·Rs·I_q². Faults 1-8 perturb the electrical side: inter-turn shorting (Faults 1-3, one per phase) raises the mean phase resistance and lowers the flux linkage in the dq projection (Section 2.6); inter-phase insulation degradation (Faults 4-6) draws a circulating fault current I_f = 0.8·S·V_ph/(R_f + 0.1·Rs) between the affected pair with fault resistance R_f = 1 Ω, adding fault power P_f = 1.2·V_ph·I_f and increasing the supply current to I_surf = I_q + 1.2·I_f; PM demagnetization (Fault 7) reduces ψ_f; bearing wear / rotor eccentricity (Fault 8) raises the viscous friction coefficient B.

The three-phase cable is a lumped series resistance R_cable = 0.112 Ω (DC resistance of 200 mm² copper over 1300 m; skin effect negligible at the 12-36 Hz electrical frequencies involved). Its power loss R_cable·I_surf² is added to the surface power channel, and cable conductor degradation raises R_cable by up to +50% (Fault 9, resistivity factor) and up to +100% (Fault 10, conductor area loss).

2.3 Coupling subsystem

The universal-joint coupling is modeled by (i) a tilt-angle transmission efficiency η_joint = cos(γ)^1.5·exp(−0.3γ), γ in rad

https://doi.org/10.37965/jdmd.2026.1608

(Fault 12, misalignment up to 0.35 rad at S = 1), (ii) a backlash half-width b_gap (Fault 11) whose quasi-steady torque cost T_gap = 0.001·max(b_gap/B_gap0 − 1, 0)·ω_m is a small drag, and (iii) a transmission efficiency η_c = 0.96 (Fault 13). The motor shaft torque is

$$T_{motor} = \frac{T_{pump}}{\eta_c \eta_{joint}} + T_{gap} + B\omega_m \quad (2)$$

2.4 Pump subsystem (FE-based PCP model)

The PCP geometry follows the code-family FE pump module: eccentricity e = 2 mm, rotor diameter D = 20 mm, stator pitch 80 mm, N = 4 stages, giving the displacement

$$V_{rev} = 4eD\text{pitch}N$$
$$Q_{th} = V_{rev} n_{rps} \quad (3)$$

and the theoretical flow Q_th = V_rev·n_rps. The healthy rotor-stator seal clearance is h0 = 40 μm over the total sealing length L_seal = N·pitch = 0.32 m. Internal leakage follows the h³ clearance law plus the Couette drag of the eccentric rotor[20]:

$$Q_{leak} = HPF\frac{\pi D h^3 \Delta p}{12\mu_{hp} L_{seal}} + 0.5\pi D h \frac{\omega_m D}{2}$$
$$\mu_{hp} = \mu\left(1 + 0.002\frac{\Delta p}{10^6} + 10^{-5}\left(\frac{\Delta p}{10^6}\right)^2\right)$$
$$HPF = 1 + 0.015\left(\frac{\Delta p}{10^6}\right)^{0.8} + 0.0002\left(\frac{\Delta p}{10^6}\right)^{1.5} \quad (4)$$

where the clearance viscosity μ_hp accounts for piezoviscous thickening and the high-pressure correction HPF applies to the pressure-driven (Poiseuille) term only. The cubic dependence on h is the dominant fault sensitivity: stator wear (Fault 14) adds up to 80 μm to h at S = 1, which multiplies leakage by up to (120/40)³ = 27. Pump torque is the sum of four terms:

$$T_{hyd} = \frac{\Delta p V_{rev}}{2\pi}$$
$$t_{visc} = \frac{\mu\omega_m D}{2h}\pi D L_{seal}\frac{D}{2}$$
$$t_{cont} = (0.002\Delta p + 0.0001\Delta p^{0.7})\mu_f \frac{D}{2} k_{contact,23}$$
$$T_{fric} = HPF_f(t_{visc} + t_{cont})$$
$$HPF_f = 1 + 0.08\left(\frac{\Delta p}{10^6}\right)^{0.8} + 0.001\left(\frac{\Delta p}{10^6}\right)^{1.5}$$
$$T_{comp} = 0.15 T_{hyd}(\ln r)^{1.2}$$
$$r = \frac{p_{out}}{p_{in}} \quad (5)$$

with the stator contact friction coefficient μ_f = 0.1 and the contact multiplier k_contact = 1 healthy, raised by stator swelling (Fault 17, ×(1+1.5S)), rotor eccentricity deviation (Fault 15, ×(1+0.8S)) and wax blockage (Fault 19, ×(1+S)). Compressibility and gas: the compression factor 1/(1 + 0.1·ln(r)^1.5) reduces the effective displacement, and free gas at the intake (Fault 20, void fraction up to 0.45 at S = 1) reduces both the effective displacement (1 − 0.8α) and the maximum working pressure through the gas-lock stall factor

$$stall = \max\left(1 - \frac{\max(\Delta p - \Delta p_{cap}, 0)}{\max(\Delta p_{cap}, 1)}, 0\right)$$
$$\Delta p_{cap} = P_{MAX}(N/N_0)\max(1 - \alpha/\alpha_{crit}, 0)$$
$$P_{MAX} = 10\text{MPa}$$
$$\alpha_{crit} = 0.5 \quad (6)$$

The max guards saturate the stall factor at 1 below the cap and floor the denominator at 1 Pa when the gas-lock bound vanishes (the isothermal pressure-ratio bound 1 − 1/r_p of the code family, r_p = P_MAX/P_in = 2). The pump intake pressure follows p_in = P_IN − (K_inlet − 1)·Δp_ref·(Q/Q_ref)² with P_IN = 5 MPa, Q_ref = 1.2e-4 m³/s, Δp_ref = 0.5 MPa; Fault 18 raises K_inlet to 3 at S = 1 (inlet restriction). Stator pitch

https://doi.org/10.37965/jdmd.2026.1608

change (Fault 16, +25% at S = 1) scales the pitch directly.

2.5 Tubing subsystem (Beggs-Brill three-phase traverse)

The tubing pressure profile is integrated top-down with the Beggs-Brill three-phase closure (water cut 0.693, oil fraction 0.233, gas fraction 0.074; oil gravity 0.8261, gas gravity 0.65) over 20 segments with the linear temperature profile of Section 2.1:

$$\frac{dp}{dz} = \left(\frac{dp}{dz}\right)_{grav} + K_{tub}\left(\frac{dp}{dz}\right)_{fric} + \left(\frac{dp}{dz}\right)_{acc} \quad (7)$$

subject to the wellhead back-pressure p_wh = 5 MPa. The friction multiplier K_tub is the wax / scale parameter (Fault 21, 1 → 2). A tubing leak (Fault 22) is a mid-depth orifice at z = 650 m discharging into the annulus (water column): Q_leak,t = C_D·A_leak·√(2·Δp/ρ_w), C_D = 0.6, A_leak up to 1e-6 m² at S = 1; the leak flow is resolved by an inner fixed point (8 iterations) inside each traverse. The traverse returns the pump discharge pressure p_out and the surface flow Q_surf = Q − Q_leak,t. The full-model monitoring vector is the settled quasi-steady limit of this coupled system at 12 VFD speeds (90-270 rpm), stacked into one 48-component vector (Fig. 4; Section 3.1).

Fig. 3 Healthy full-model VSD sweep: the 48-component monitoring vector (4 channels × 12 speeds, 90

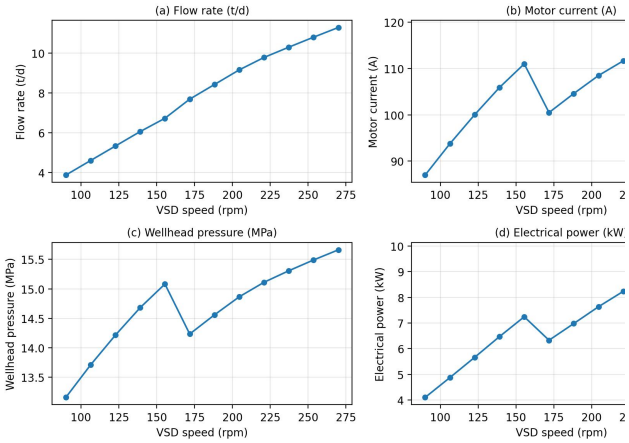


Fig. 4 Healthy full-model VSD sweep: the 48-component monitoring vector (4 channels × 12 speeds, 90-270 rpm).

2.6 Fault parameterization

Table 2 defines the 22 fault parameters $\theta_1..\theta_{22}$, each by its physical meaning, healthy value, fault range, and the explicit perturbation function applied to the full model at severity S ∈ [0,1] (severity = $|\theta^* - \theta 0|/(\theta max - \theta min)$). Three remarks follow from the perturbation functions. (i) Inter-turn shorting acts identically in all three phases in the dq projection - only the mean resistance increase and the flux reduction are observable, so F1-F3 share one signature (the faulted phase is not identifiable from the four channels); the same holds for the three inter-phase pairs (F4-F6). (ii) The cable faults F9/F10 both change the series resistance seen at the surface; they differ only in their severity scaling, so they form one family. (iii) F8, F11 and F21 enter the quasi-steady channels only through small secondary terms (viscous friction, backlash drag, tubing friction multiplier) whose quasi-steady channel effect at moderate severity is below the 3% measurement noise - the diagnostic limits for these three faults are quantified in Section 4.3.

Table 2. The 22 fault parameters: physical meaning, healthy value, fault range, and the perturbation applied to the full model at severity S.

No.	Fault (physical meaning)	Normal value	Fault range	Perturbation at severity S
F1	Phase-A winding resistance (inter-turn A)	0.2 Ohm	[0.2, 0.3] Ohm	Rs_avg = 0.2(1+0.5S/3) Ohm; psi_f = 0.1(1-0.2S) Wb
F2	Phase-B winding resistance (inter-turn B)	0.2 Ohm	[0.2, 0.3] Ohm	Rs_avg = 0.2(1+0.5S/3) Ohm; psi_f = 0.1(1-0.2S) Wb

https://doi.org/10.37965/jdmd.2026.1608

F3	Phase-C winding resistance (inter-turn C)	0.2 Ohm	[0.2, 0.3] Ohm	Rs_avg = 0.2(1+0.5S/3) Ohm; psi_f = 0.1(1-0.2S) Wb
F4	Inter-phase insulation (A-B)	1e+06 Ohm	[100, 1e+06] Ohm	ins_S = S (I_f = 0.8S.V_ph/(R_f+0.1 Rs), P_f = 1.2V_ph.I_f)
F5	Inter-phase insulation (A-C)	1e+06 Ohm	[100, 1e+06] Ohm	ins_S = S (I_f = 0.8S.V_ph/(R_f+0.1 Rs), P_f = 1.2V_ph.I_f)
F6	Inter-phase insulation (B-C)	1e+06 Ohm	[100, 1e+06] Ohm	ins_S = S (I_f = 0.8S.V_ph/(R_f+0.1 Rs), P_f = 1.2V_ph.I_f)
F7	PM demagnetisation	0.1 Wb	[0.08, 0.1] Wb	psi_f = 0.1(1-0.2S) Wb
F8	Rotor eccentricity (bearing wear)	0 -	[0, 0.5] -	B = 0.05(1+0.5S) N.m.s/rad
F9	Cable resistivity factor	1 -	[1, 1.5] -	R_cable = 0.112(1+0.5S) Ohm
F10	Cable conductor area loss	0 -	[0, 0.5] -	R_cable = 0.112/(1-0.5S) Ohm
F11	Coupling backlash	0.005 rad	[0.005, 0.05] rad	b_gap = 5e-3(1+9S) rad
F12	Coupling misalignment angle	0 rad	[0, 0.35] rad	gamma = 0.35S rad
F13	Coupling transmission efficiency	0.96 -	[0.7, 0.96] -	eta_c = 0.96 - 0.26S
F14	Stator wear depth	0 mm	[0, 0.08] mm	h = 40um + 0.08mm.S
F15	Pump rotor eccentricity deviation	0 -	[0, 1] -	h x(1+0.6S); k_contact x(1+0.8S)
F16	Stator pitch change	0 mm	[0, 20] mm	pitch = 80(1+0.25S) mm
F17	Stator swelling	0 -	[0, 1] -	h x(1-0.6S); k_contact x(1+1.5S)
F18	Pump inlet flow coefficient	1 -	[1, 3] -	K_inlet = 1 + 2S
F19	Pump wax blockage	0 -	[0, 1] -	h x(1-0.5S); k_contact x(1+1.0S)
F20	Gas volume fraction at intake	0 -	[0, 0.45] -	alpha_g = 0.45S
F21	Tubing flow coefficient	1 -	[1, 2] -	K_tub = 1 + S
F22	Tubing leak area	0 m2	[0, 1e-06] m2	A_leak = 1e-6.S m2

2.7 Retained, approximated and omitted physical effects

Retained (explicit): dq motor with torque constant and copper loss; lumped cable resistance; coupling tilt efficiency + backlash drag; pump displacement geometry; h^3 clearance leakage with piezoviscosity and high-pressure correction; Couette drag; pressure-dependent contact friction; compression torque; gas void fraction with stall collapse; Beggs-Brill three-phase tubing traverse with slip closure; tubing friction multiplier; mid-depth tubing leak; inlet restriction; the pump-tubing fixed point (the electromagnetic-mechanical-hydraulic loop).

Approximated: quasi-steady (settled) evaluation instead of the full transient at every diagnostic step (verified equivalent, Section 4.8); prescribed linear temperature profile instead of a solved thermal field; lumped instead of distributed cable; analytical clearance leakage instead of a CFD solution of the seal gap; contact friction via the pressure-dependent empirical law of the code family instead of a hyperelastic contact solve; gas-lock via the stall factor instead of a two-phase chamber simulation.

https://doi.org/10.37965/jdmd.2026.1608

Omitted (not in the diagnostic forward model): thermal transients and motor temperature dynamics; vibration / acoustic channels; phase-level current imbalance waveforms (only the quasi-steady aggregate I_f is retained); magnetic saturation and iron losses (the code family's Steinmetz term is not needed for the settled current channel).

3. Regularized Inversion Formulation and Experimental Protocol

3.1 Monitoring vector and the reduced quasi-steady model

The diagnostic vector d ∈ R^48 is formed by evaluating, at 12 VFD speeds (90, 106, ..., 270 rpm, uniform 12-point sweep) and in the settled quasi-steady state, the four surface channels of Table 3: flow rate, supply current, wellhead pressure and electrical power. Torque is NOT measured independently: it is inferred from the supply current through the motor torque constant ($I_q = T_{motor}/(1.5 \cdot p \cdot \psi_f)$). The sweep is a single monitoring cycle; the 48 components are stacked into one data vector, and the inversion uses all of them jointly (no per-speed inversions). In the simulation study, measurement noise is multiplicative Gaussian per channel (1/3/5/10% standard deviation). The transient settling between sweep points is not part of the diagnostic data - each sweep point is the settled limit (Section 4.8 verifies this attractivity by direct transient integration for all nine verification cases).

Table 3. Measurement inventory of the diagnostic vector (all channels surface-mounted); m = 4 channels x 12 VSD speeds = 48.

Q_surf	**t/d**	**flow meter, wellhead**	**tubing traverse Q - Q_leak,t (multi-phase slip closure)**	**settled value at each of the 12 VSD speeds 90-270 rpm; stacked into one 48-vector; no filtering**
I_surf	A	current transformer, VFD output panel	I_q + 1.2 I_f, where I_q = T_motor/(1.5 p psi_f)	settled value per speed; torque is NOT measured - inferred from current via the motor torque constant
p_out	MPa	pressure transducer, wellhead	tubing traverse discharge pressure (Beggs-Brill)	settled value per speed
P_surf	kW	power meter, VFD panel (or V.I with PF)	P_motor + P_f + P_cable (copper + fault + cable losses)	settled value per speed; computed quantity, not a separate sensor

The inversion cannot afford the full model's Beggs-Brill traverse inside every objective evaluation - each inversion evaluates the forward map thousands of times (Stage A of this study: 4,017 function evaluations), and each forward evaluation re-solves the pump-tubing fixed point; the reduced model therefore replaces the tubing traverse with a calibrated surrogate that preserves the traverse's key qualitative feature - p_out rising with flow rate through gas-holdup relaxation - which a pure quadratic friction term cannot represent:

$$p_{out}(Q) = p_{wh} + \rho g H(1 - K_{HOLD}\exp(-Q/Q_{HOLD})) - C_{tub}\left(\frac{Q}{Q_{ref}}\right)^2 \quad (8)$$

https://doi.org/10.37965/jdmd.2026.1608

with K_HOLD = 0.5697 and Q_HOLD = 0.4855·Q_ref fitted once to the healthy full-model traverse (RMSE 0.31 MPa) and then frozen - they are not inversion parameters. The 12 inversion parameters (Table 4) are θ = [p_wh, Rs, ψ_f, B, D, e, pitch, N, η_v, η_m, C_leak, C_tub] with bounds and baselines in Table 4. The reduced map is the 30-iteration fixed point

$$Q = Q_{th}\eta_{v,eff} - C_{leak}\left(\frac{e}{E_0}\right)^3 \frac{PITCH_0}{pitch}\Delta p\frac{\mu_{base}}{\mu}$$
$$\eta_{v,eff} = \max\left(\eta_{v'}\left(1 - 0.1\Delta p/P_{max,eff}\right)\ 0.3\right)$$
$$P_{max,eff} = P_{MAX}\frac{N}{N_0}$$
(9)

and the torque chain mirrors the full model's structure (same four torque terms):

$$T_{hyd} = \frac{\Delta p V_{rev}}{2\pi\eta_m}$$
$$T_{fric} = \left(0.5 + 0.3 n_{rps}\,\mu/\mu_{base}\right)\left(\frac{D}{D_0}\right)^3\left(\frac{e}{E_0}\right)^2$$
$$T_{comp} = 0.15 T_{hyd}(\ln r)^{1.2}$$
$$T_{cont} = (0.002\Delta p + 0.0001\Delta p^{0.7})\mu_f\frac{D}{2}$$
$$T_e = T_{hyd} + T_{comp} + T_{cont} + T_{fric} + B\omega_m$$
$$I_q = \frac{T_e}{1.5 p \psi_f}$$
$$P_{elec} = T_e\omega_m + 1.5 R_s I_q^2$$
(10)

Table 4. The 12 reduced-model parameters: name, bounds, baseline, Stage-A estimate, and unit.

Parameter	Name	Bounds	Baseline	Stage-A estimate	Unit
x1	**wellhead pressure (MPa)**	**[1, 10]**	**5**	**5.012**	**MPa**
x2	stator resistance Rs (Ohm)	[0.05, 0.5]	0.2	0.2747	Ohm
x3	magnet flux psi_f (Wb)	[0.05, 0.3]	0.1	0.097	Wb
x4	viscous friction B (N.m.s/rad)	[0.01, 0.2]	0.05	0.05192	N.m.s/rad
x5	rotor diameter D (mm)	[15, 25]	20	20.11	mm
x6	eccentricity e (mm)	[1, 5]	2	1.998	mm
x7	pitch length T (mm)	[60, 100]	80	80	mm
x8	number of stages N	[2, 8]	4	3.873	-
x9	volumetric eff. eta_v	[0.6, 0.95]	0.85	0.839	-
x10	mechanical eff. eta_m	[0.5, 0.98]	0.9	0.8808	-
x11	leakage coeff. C (1e-12)	[0.1, 10]	1	0.7759	1e-12
x12	tubing flow coeff. C_tub (1e5 Pa)	[0.2, 20]	2	0.2	1e5 Pa
	N is optimized as a continuous effective stage count (Section 3.1); Rs is an effective stator resistance absorbing the cable series drop (Section				

https://doi.org/10.37965/jdmd.2026.1608

	3.1).				

Two of the 12 parameters deserve comment. (i) The stage count N is optimized as a CONTINUOUS effective coefficient in [2, 8], not as an integer: it scales the displacement volume V_rev = 4eDTN and the stall pressure P_max_eff = P_MAX·(N/N0). At the healthy design it recovers ≈ 3.87 (Stage A, Table 4), i.e. close to the true integer stage count N0 = 4; between integers it interpolates the displacement/stall capacity continuously, and the diagnostic rule interprets deviations of N from its calibrated baseline as displacement-capacity changes rather than rounding it to an integer. (ii) The stator resistance Rs is an EFFECTIVE parameter: the reduced model has no cable resistance, so the calibrated Rs absorbs the cable series drop (Stage A recovers Rs = 0.275 Ω against the design 0.2 Ω, consistent with R_cable = 0.112 Ω), and cable faults F9/F10 are therefore diagnosed through their projection onto Rs (Table 6).

3.2 Tikhonov formulation, scaling and regularization selection

The inverse problem minimizes, over the scaled parameters x = (θ − LO)/(HI − LO) ∈ [0,1]^12,

$$J(x) = \frac{1}{2}\left\| W\left(d - m\left(\theta(x)\right)\right)\right\|^2 + \frac{\lambda}{2}\|x - x_0\|^2 \quad (11)$$

with channel weights W = diag(1/|d_nom|) (equal relative weight per channel). The solution is computed with multi-start L-BFGS-B (K = 3 starts: the baseline x0, a 5% perturbation of x0, and a uniform draw; ftol 1e-14, gtol 1e-10, maxiter 300). A differential-evolution global search was benchmarked against this solver on 12 representative cases (Section 4.8, Table 8): DE with L-BFGS-B polish reaches the same optimum but costs 17.8× more function evaluations (DE-lite at reduced population: 4.0×), so the multi-start L-BFGS-B is the standard solver of the pipeline and DE is not used adaptively.

Regularization parameters are chosen on the L-curve of the linearized system at the calibration point (Hansen[24]corner), verified by nonlinear solves (Fig. 5): λ*(1% noise) = 0.146, λ*(3%) = 0.533, λ*(5%) = 1.017, λ*(10%) = 2.984. These per-noise corners are reported for the selection and identifiability analysis only; the experiment itself fixes one value per stage. Stage A (healthy calibration) uses λ_cal = 0.18 (≈ the 1% corner), and Stage B (diagnosis) uses λ_diag = 0.53 (≈ the 3% corner) for ALL four noise levels, so that the signature set and the Stage-C evidence score of every cell are directly comparable; the noise-dependent scaling of the null score is instead absorbed by the per-noise detection threshold τ (Section 4.2, Table 5).

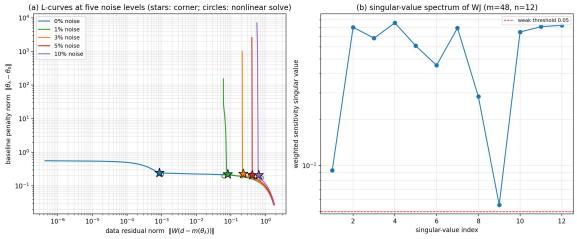


Fig. 5 Regularization-parameter selection: (a) measured L-curves at five noise levels (stars: Hansen corners; circles: nonlinear multi-start solves at the corner λ); (b) singular-value spectrum of the weighted sensitivity matrix W·J_x (m = 48, n = 12).

3.3 Identifiability of the 12 reduced parameters

The weighted sensitivity matrix A = W·J_x (J_x = ∂m/∂x at the fault-state operating point of Table 4's design factors) has the singular-value spectrum (Fig. 5(b))

[25.53, 7.22, 2.87, 2.04, 1.44, 0.409, 0.109, 1.56e-2, 2.02e-3, 1.20e-3, 2.28e-5, 4.14e-7],

condition numbers $\kappa(A) = 6.2e7$ and $\kappa(J^T W^2 J) = 3.8e15$, and five singular values below the 0.05 weak threshold. The weak subspace (participation in the five smallest right singular vectors, weakest first) is: viscous friction B (0.979), pitch (0.922), rotor diameter D (0.867), volumetric efficiency η_v (0.725), stage count N (0.487), then η_m (0.351), C_{leak} (0.338), e (0.321); the remaining four directions (C_{tub} 0.009, ψ_f 0.0002, Rs 0.0002, p_{wh} 0.0001) are effectively identified. Two consequences. First, B, pitch, D, η_v are only weakly constrained by the four quasi-steady channels, which is precisely why F8 (bearing wear, acting through B) is at the detection limit (Section 4.3) - regularization stabilizes the estimates but cannot create information; F11 and F21 are at the detection limit for a different reason - their full-model effects are below the noise (Section 4.5). Second, the strong directions are the electrical parameters (Rs, ψ_f) and the pressure-observable parameters (p_{wh}, C_{tub}), plus the leakage/clearance composite ($C_{leak}\cdot(e/E0)^3$, D through the friction and leakage geometry): the 22 faults project onto these directions in the signature space, which is the mapping documented in Section 4.7 (Table 6).

3.4 Two-stage diagnosis pipeline

Stage A (baseline calibration). From one noise-free healthy sweep, invert with λ_{cal} from the design baseline: the result of this study is θ_{cal} = [5.012, 0.275, 0.097, 0.052, 20.11, 1.998, 80.0, 3.873, 0.839, 0.881, 0.776, 0.2] (units as in Table 4), with residual norm 0.251 and maximum per-component relative error 10.4% at the 155.5-rpm point of the power channel; the pressure channel's maximum is 3.73% at the 171.8-rpm point, where the Beggs-Brill traverse switches flow pattern - the irreducible surrogate error of the smooth reduced map. The calibration residual also defines the nominal model-form error floor for diagnosis (Section 4.4 shows a Stage-B fit with the healthy baseline overlaid, Fig. 7).

Stage B (signature construction). For each fault i = 1..22, the full model is evaluated at severity S = 0.5 with 3% noise (3 repetitions); each repetition is inverted with λ_{diag}; the fault signature v_i is the unit normalization of the mean of the three raw deviations. The pairwise cosine similarity of the signatures quantifies a-priori confusability (Fig. 12(a)): phases of the same fault family (F1-F3, F4-F6) share one perturbation by construction (measured pairwise cosines 0.90-0.96 and 0.90-0.999 respectively); coupling transmission loss F13 lands on the inter-turn direction (F13-to-F1/F2/F3 cosines 0.956-0.983); cable resistivity/area (F9/F10) form one pair (0.93); stator wear and tubing leak (F14/F22) share the leakage-eccentricity direction (0.97); stator swelling and wax blockage (F17/F19) overlap in the clearance-contact direction (0.97).

Stage B (diagnosis). For a measured sweep d: invert with λ_{diag} from x_{cal}, and score every signature direction by its restricted-fit evidence

$$score_i(d) = \frac{1}{2}\left\|W\left(d - m(x_{cal})\right)\right\|^2 - \min_s\left[\frac{1}{2}\left\|W\left(d - m(x_{cal} + s v_i)\right)\right\|^2 + \frac{\lambda_{diag}}{2}s^2\right] - base_i \quad (12)$$

where $base_i$ is the same expression evaluated on the noise-free healthy sweep (the model-form-error projection of direction i; measured 7.2e-8..9.4e-5 with

median 1.7e-6, negligible against the noise level). The cell is declared fault i* = argmax_i score_i when the score exceeds τ(pct) - the 95th percentile of 60 healthy calibration draws at that noise level - and "normal" otherwise; the false-alarm rate is measured on 40 held-out healthy draws per level (Table 5). The winning fit's s* is the severity estimate. Why the deviation-cosine rule is not used: with λ_diag = 0.53 the recovered deviation is shrinkage- and noise-dominated (median $\|\hat{x} - x_{cal}\| = 0.028$ for healthy vs 0.066 for faulted cells at S = 0.5, 3% noise), so healthy and faulted unit deviations occupy the same subspace and no threshold separates them; the evidence score evaluates each fault hypothesis in the data space instead (Section 4.2 reports the measured null distributions).

3.5 Experimental grid

22 fault types × 5 severities {0.1, 0.3, 0.5, 0.7, 1.0} × 4 noise levels {1, 3, 5, 10%} × 3 repetitions = 1,320 fault cells, plus 400 healthy cells (100 per noise level: 60 threshold-calibration draws and 40 held-out false-alarm draws), 1,720 inversions in total. Every accuracy and detection aggregate reported in Section 4 is regenerated from the single master results table (supplementary material, master_table.csv): overall accuracy, per-noise and per-severity accuracy, per-fault and per-group accuracies, the confusion matrix, the detection thresholds and the false-alarm rates. Fig. 6, Fig. 10, Fig. 11 and Fig. 12(b) are computed exclusively from this table; the signature cosine matrix (Fig. 12(a)), the fault mapping (Table 6), the solver benchmark (Table 8), the transient verification (Table 9) and the discrepancy study (Table 10) are the separate protocol runs of Sections 4.7-4.9, each regenerated by one script from its own recorded results.

4. Results and Discussion

4.1 Experimental setup

The experimental grid of Section 3.5 is executed with the full model of Section 2 as the data generator (a virtual test rig; no field SCADA records are used): 22 fault types × 5 severities {0.1, 0.3, 0.5, 0.7, 1.0} × 4 noise levels {1, 3, 5, 10%} × 3 repetitions = 1,320 fault cells plus 400 healthy cells (100 per noise level: 60 threshold-calibration draws and 40 held-out false-alarm draws), 1,720 Stage-B inversions in total (multiplicative Gaussian noise, seeded RNG, seed 2026). Every cell records: the predicted class (or "normal" when below the threshold τ), the evidence score, the correctness flag, the severity estimate s* of the winning signature direction, the solver function-evaluation count and success flag, and the residual norm. The complete table is the single source of every accuracy, detection, false-alarm and confusion aggregate in this section (supplementary material, master_table.csv); Fig. 6, Fig. 10, Fig. 11, Fig. 12(b) and Tables 5 and 7 are computed from it by one script, so no two of these numbers can disagree with each other. Tables 6, 8, 9 and 10 and Fig. 12(a) report the separate protocol runs of Sections 4.7-4.9, each regenerated by one script from its own recorded results.

4.2 Stage A calibration, Stage B signatures, and the evidence rule

Stage A. Inverting the noise-free healthy sweep with λ_cal = 0.18 from the design baseline gives θ_cal = [5.012, 0.275, 0.097, 0.052, 20.11, 1.998, 80.0, 3.873, 0.839, 0.881, 0.776, 0.2] (4,017 function evaluations). The weighted residual norm is 0.251; the largest single-component error (10.4%) is the 155.5-rpm point of the power channel; the pressure channel's largest error is 3.73% at 171.8 rpm, where the Beggs-Brill traverse switches flow pattern - a local feature the smooth holdup surrogate cannot reproduce. The effective parameters are

https://doi.org/10.37965/jdmd.2026.1608

physically interpretable: Rs = 0.275 Ω against the design 0.2 Ω - the reduced model's 1.5·Rs·Iq² term absorbs the full model's cable loss R_cable·I_surf², which the fit achieves with the increment 0.075 ≈ R_cable/1.5 (the reduced model has no cable term); N = 3.87 against the integer design N0 = 4; η_m = 0.881 against the healthy design 0.90. The residual 0.251 is the model-form error floor carried into every Stage-B inversion.

Stage B. Inverting each fault at S = 0.5 (3% noise, 3 reps) and unit-normalizing the mean deviation gives the signature set; the pairwise cosine matrix is Fig. 12(a). Signatures above 0.90 cosine: F1/F2/F3 with F13 (0.95-0.98), F4-F6 among themselves (0.90-0.999), F9/F10 (0.93), F14/F22 (0.97), F17/F19 (0.97). These are the a-priori confusable pairs.

Decision rule. Classification by raw deviation cosine was found inadequate: with λ = 0.53 the recovered deviation is dominated by shrinkage and noise, so healthy and faulted unit deviations occupy the same subspace (measured: median fault deviation norm 0.066 vs 0.028 healthy at S = 0.5/3% noise; healthy max-cosines reach 0.99 with median 0.76). The final rule therefore scores each signature direction by its data-space evidence (Eq. (12)): the cell is declared fault i* = argmax_i score_i when the score exceeds τ(pct) = the 95th percentile of the 60 healthy calibration draws at that noise level, and "normal" otherwise; the false-alarm rate is measured on the 40 held-out draws (Table 5). s* of the winning fit is the severity estimate. The threshold is calibrated per noise level because the null score scales with the noise power (measured null medians 1e-4 / 8e-4 / 2e-3 / 9e-3 at 1/3/5/10%).

Table 5. Detection threshold tau and false-alarm rate (FAR) per noise level. tau = 95th percentile of the 60 healthy calibration draws at that level; FAR on the 40 held-out healthy draws.

Noise level (%)	tau (detection threshold)	False-alarm rate
1	0.0004537	0.025
3	0.002052	0.175
5	0.006613	0.100
10	0.02889	0.050

4.3 Detection limits

Fig. 6 maps the detection rate for every fault × severity at 3% noise. Three faults are at the detection limit by construction (Section 2.6): F8 (bearing wear, acts through the viscous friction coefficient B, whose participation in the weak subspace is 0.979 - Section 3.3) is detected in 26.7% of cells, and F11 (backlash drag) and F21 (tubing friction multiplier) in 20.0%/6.7%; their per-severity detection rates (0.33/0.33/0.00/0.33/0.33, 0.33/0.67/0.00/0.00/0.00 and 0.33/0.00/0.00/0.00/0.00 at S = 0.1/0.3/0.5/0.7/1.0) show no systematic rise with severity - F8 stays at the 3% noise floor through the weak subspace, while F11 and F21 stay at the floor because their full-model effects are below the noise (Section 4.5). All electrical faults (F1-F7, F9/F10, F12/F13), the clearance faults (F14, F15, F17, F19, F22), F16 and F20 are detected at rates 0.33-1.00 from S = 0.1 upward; F18 (inlet restriction) joins them from S = 0.3 upward (0.00/1.00/1.00/1.00/1.00 over the severity grid).

https://doi.org/10.37965/jdmd.2026.1608

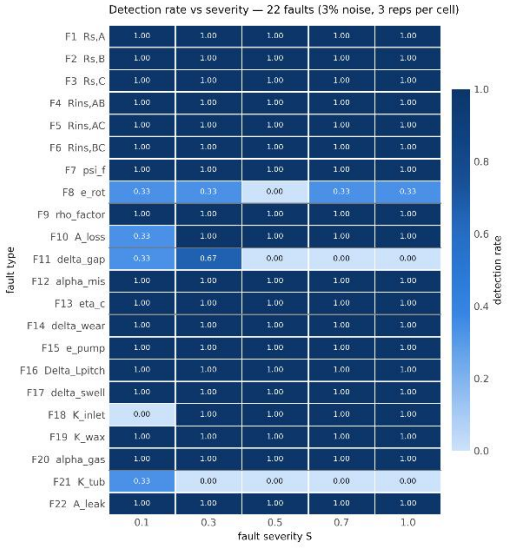


Fig. 6 Detection rate vs severity for all 22 faults at 3% noise (3 reps per cell, 15 cells per fault).

4.4 Inversion fitting examples

Fig. 7 shows one Stage-B inversion at the representative cell (stator wear F14 at S = 0.5, 3% noise): measured, prediction, and healthy baseline for all 48 components. The per-channel root-mean-square errors of the prediction against the noisy measurement are 0.25 t/d (flow), 4.30 A (current), 0.39 MPa (pressure) and 0.45 kW (power) - 3.2%, 4.1%, 2.7% and 6.3% of the healthy nominal values respectively, i.e. the fit tracks the measurement at the noise level on every channel; the weighted residual norm of the fit is 0.292 against the Stage-A floor of 0.251. The evidence scores of the 22 directions at this cell are led by F14 with score 1.746 and severity estimate s* = 0.30 against the true 0.5 - the winning direction is the correct one, and the severity is under-estimated rather than mis-attributed (the regularized 1-D fit along the signature direction trades magnitude for robustness; direction, not magnitude, is the diagnostic output).

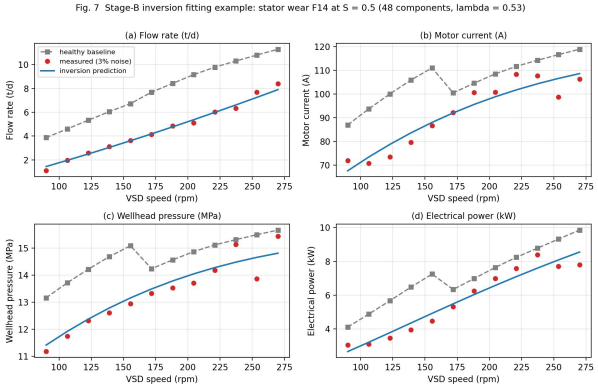
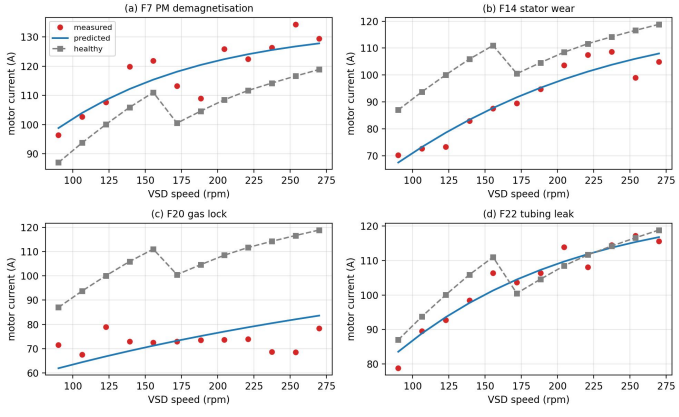


Fig. 7 Stage-B inversion fitting example: stator wear F14 at S = 0.5 (48 components, λ = 0.53, 3% noise): measured vs inversion prediction with the healthy baseline.

Fig. 8 repeats the exercise on the motor-current channel for four faults spanning the signature space (F7 demagnetisation, F14 wear, F20 gas lock, F22 tubing leak, each S = 0.5, 3% noise): the current-channel root-mean-square errors are 5.30 A, 3.96 A, 7.36 A and 2.72 A (5.0%, 3.7%, 6.9% and 2.6% of the nominal 106 A). Gas lock shows the largest misfit because the reduced map has no gas parameter - the stall collapse must be approximated through the pressure and volumetric parameters - but the fit still captures the signature direction (a current decrease), which is what the evidence rule tests.



https://doi.org/10.37965/jdmd.2026.1608

Fig. 8 Motor-current channel: measured vs inversion prediction for four faults at S = 0.5 (3% noise).

4.5 Fault response signatures

Fig. 9 compares the four-channel sweeps of five representative fault states (F7, F14, F16, F20, F21 at S = 0.5) against the healthy sweep; Table 6 (Section 4.7) reports the maximum per-channel change for all 22 faults. The five patterns of Fig. 9 illustrate the discrimination basis of the inversion (mean changes over the 12-speed sweep): stator wear F14 enlarges the clearance so that the flow channel drops by 44.9% - and, because the enlarged clearance also relaxes the viscous and contact friction while the reduced differential pressure lowers the hydraulic torque, the current and power channels drop by 14.9% and 23.2% in the same direction; pitch change F16 rescales the displacement volume and thereby scales flow, current and power upward together (+11.4%, +13.1%, +23.2%); gas lock F20 collapses the flow channel (mean −69.9%) while the reduced load lowers the current by 30.7% - the opposite current direction to every mechanical-load fault, which is its discriminating hallmark; demagnetisation F7 raises the current by 11.1% (reduced torque constant) with no flow response; tubing friction F21 changes all four channels by less than 0.03% at S = 0.5 (max over channels, on the power channel) - the traverse friction is negligible at these flow rates - which is precisely why it sits at the detection limit of Section 4.3. These channel-level directions, not any single scalar, are what the 12-parameter inversion and the evidence scores operate on.

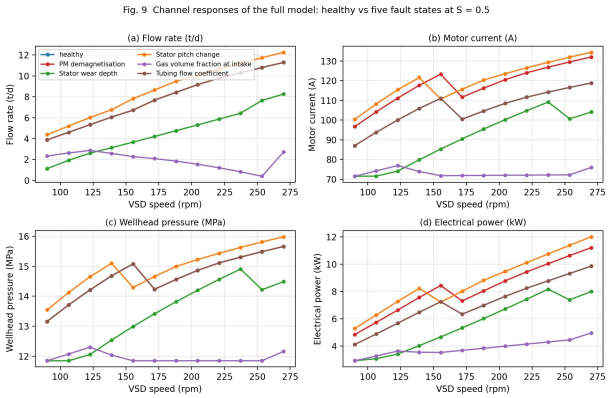


Fig. 9 Channel responses of the full model: healthy vs five fault states at S = 0.5.

4.6 Headline accuracy and its subsets

At 3% noise, over all 22 faults and 5 severities (22 × 5 × 3 = 330 fault cells), the per-type classification accuracy is 41.2% and the detection rate is 87.3%. By noise level: acc 45.2%/41.2%/33.0%/26.1% and detection 86.4%/87.3%/78.8%/70.0% at 1/3/5/10% (Fig. 10). By severity at 3% noise: acc 27.3%, 45.5%, 45.5%, 45.5%, 42.4% for S = 0.1/0.3/0.5/0.7/1.0. Group-level accuracy at 3% noise is 69.4% (per group: G1 0.767, G2 0.956, G3 0.733, G4 0.000, G5 0.733, G6 0.067, G7 0.000, G8 0.867, G9 1.000, G10 1.000, G11 0.867, G12 0.667, G13 0.933, G14 0.000); false-alarm rates per noise level 2.5%/17.5%/10.0%/5.0% (n = 40 held-out healthy draws per level). These false-alarm rates are point estimates on 40 held-out draws, not the nominal 5% exceedance one reads off the 95th-percentile threshold: the threshold is the empirical 95th percentile of only 60 calibration draws – itself a noisy estimate of the population quantile – and the false alarms are counted out-of-sample on a fresh 40 draws. At 3% noise 7 of 40 held-out healthy draws exceed the threshold (17.5%), whose 95% Wilson confidence interval is [8.7%, 31.9%]; the corresponding intervals at 1/5/10% noise are [0.4%, 12.9%] / [4.0%, 23.1%] / [1.4%, 16.5%]. Every interval contains the nominal

https://doi.org/10.37965/jdmd.2026.1608

5% rate, so the 17.5% value reflects the binomial sampling variance of the 40-draw set rather than a mis-calibrated threshold. All aggregates here are computed from the single master results table with the denominators stated explicitly. Per-fault accuracy and detection at 3% noise are Table 7; the full per-fault detection vs severity map is Fig. 6, and per-type correct-rate vs severity curves for representative faults are Fig. 12(b).

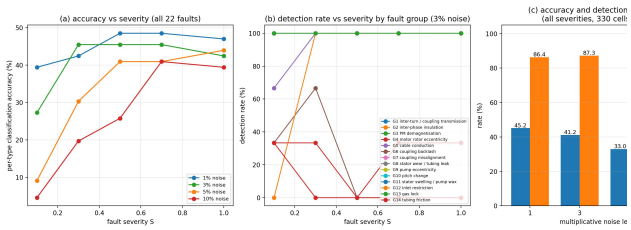


Fig. 10 (a) Per-type accuracy vs severity by noise level (each point: all 22 faults × 3 reps = 66 cells at that severity and noise level); (b) detection rate vs severity by fault group at 3% noise (each point: the group's fault cells, 3 reps each); (c) accuracy and detection vs noise level (all severities, 330 cells per level).

4.7 Confusion groups, signature structure, and the fault mapping

The 23-class confusion matrix at 3% noise (Fig. 11) realizes the signature structure: the off-diagonal mass concentrates in the measured signature groups {1,2,3,13}, {4,5,6}, {9,10}, {14,22}, {17,19}. Within a group, the faults are not distinguishable from the four quasi-steady channels at any severity - a fundamental identifiability statement, not a solver artifact (Section 3.3): the faulted PHASE of an inter-turn or inter-phase fault, and the pair attribution, are invisible to the dq projection. The evidence rule distributes the cells of a group among its members according to which member's signature best absorbs the residual, so the per-type score within a group is low by construction while the group is nearly always identified. We therefore report group-level accuracy alongside per-type accuracy (Table 7): at group level the accuracy is 69.4% at 3% noise. The clearest measured case where detection and attribution diverge is F12 (coupling misalignment): every one of its 15 cells is detected, yet none is typed as F12 - 10 of the 15 cells land in the inter-turn group G1 {F1,F2,F3,F13} (F13 alone takes 4), and the remaining 5 scatter on F7 (2), F9 (1), F15 (1) and F19 (1) (Table 7, group G7 0.000).

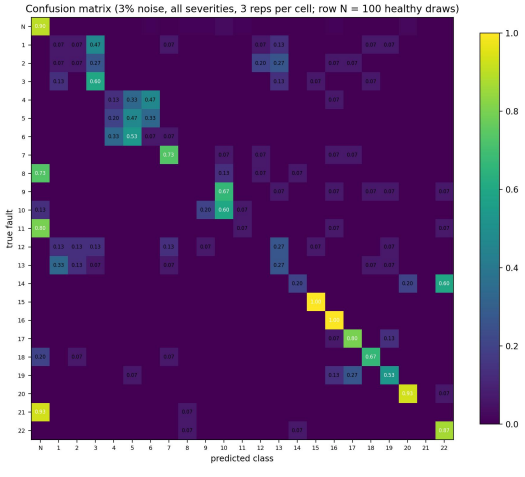


Fig. 11 Confusion matrix (3% noise, all severities, 3 reps per cell; row N = 100 healthy draws).

Table 6 closes the loop between the two model levels: for every fault it lists the full-model perturbation (Table 2), the reduced-model representation actually recovered by inverting the noise-free full-model sweep at S = 0.5 from the calibrated baseline, the affected channels (maximum relative change per channel), and the decision rule, with group labels as finalized above. The 12 reduced parameters are the quantities the inversion estimates directly; the 22 fault labels are inferred from the signature directions, and several fault families map to

https://doi.org/10.37965/jdmd.2026.1608

the same reduced direction - the table states these overlaps explicitly rather than implying 22 independent identifiability claims.

Table 6. Mapping from the 22 full-model faults to the 12 reduced-model parameters, measured by inverting the noise-free full-model sweep at S = 0.5 (lambda = 0.53) from the calibrated baseline; affected channels give the maximum relative change of the full-model sweep vs healthy. Group labels are finalized from the experiment confusion matrix (Table 7). Decision rule (per row): the restricted-fit evidence score of Eq. (12) against the signature direction, thresholded per noise level by the 95th percentile of the healthy calibration draws (Section 4.2); faults sharing one group are reported at group level.

	Phase-A winding resistance (inter-turn A)	**Rs_avg = 0.2(1+0.5S/3) Ohm; psi_f = 0.1(1-0.2S) Wb**	**stator resistance Rs (Ohm) +0.0359, magnet flux psi_f (Wb) -0.034, mechanical eff. eta_m -0.015**	**flow 0.00 current 0.11 pressure 0.00 power 0.24**	**restricte evidence of Eq. (1 along signatur threshol tau(pct); group la shared w confusal faults**
	Phase-B winding resistance (inter-turn B)	Rs_avg = 0.2(1+0.5S/3) Ohm; psi_f = 0.1(1-0.2S) Wb	stator resistance Rs (Ohm) +0.0359, magnet flux psi_f (Wb) -0.034, mechanical eff. eta_m -0.015	flow 0.00 current 0.11 pressure 0.00 power 0.24	restricted evidence of Eq. (1 along sig v_2, thre tau(pct); label sha with con faults
	Phase-C winding resistance (inter-turn C)	Rs_avg = 0.2(1+0.5S/3) Ohm; psi_f = 0.1(1-0.2S) Wb	stator resistance Rs (Ohm) +0.0359, magnet flux psi_f (Wb) -0.034, mechanical eff. eta_m -0.015	flow 0.00 current 0.11 pressure 0.00 power 0.24	restricted evidence of Eq. (1 along sig v_3, thre tau(pct); label sha with con faults
	Inter-phase insulation (A-B)	ins_S = S (I_f = 0.8S.V_ph/(R_f+0.1Rs), P_f =	stator resistance Rs (Ohm) -0.0489, magnet flux psi_f (Wb) -	flow 0.00 current 0.13 pressure 0.00 power 0.10	restricted evidence of Eq. (1 along sig v_4, thre
		1.2V_ph.I_f)	0.0343, mechanical eff. eta_m -0.0123		tau(pct); group label shared with confusable faults
hase ion (A-	ins_S = S (I_f = 0.8S.V_ph/(R_f+0.1Rs), P_f = 1.2V_ph.I_f)	stator resistance Rs (Ohm) -0.0489, magnet flux psi_f (Wb) -0.0343, mechanical eff. eta_m -0.0123	flow 0.00 current 0.13 pressure 0.00 power 0.10	restricted-fit evidence score of Eq. (12) along signature v_5, threshold tau(pct); group label shared with confusable faults	group: inter-p insulat
hase ion (B-	ins_S = S (I_f = 0.8S.V_ph/(R_f+0.1Rs), P_f = 1.2V_ph.I_f)	stator resistance Rs (Ohm) -0.0489, magnet flux psi_f (Wb) -0.0343, mechanical eff. eta_m -0.0123	flow 0.00 current 0.13 pressure 0.00 power 0.10	restricted-fit evidence score of Eq. (12) along signature v_6, threshold tau(pct); group label shared with confusable faults	group: inter-p insulat
netisatio	psi_f = 0.1(1-0.2S) Wb	magnet flux psi_f (Wb) -0.0352, mechanical eff. eta_m -0.0118, volumetric eff. eta_v -0.00892	flow 0.00 current 0.11 pressure 0.00 power 0.18	restricted-fit evidence score of Eq. (12) along signature v_7, threshold tau(pct); group label shared with confusable faults	group: demag n
icity g wear)	B = 0.05(1+0.5S) N.m.s/rad	mechanical eff. eta_m -0.00717, leakage coeff. C (1e-12) -0.00716, stator resistance Rs (Ohm) +0.00565	flow 0.00 current 0.00 pressure 0.00 power 0.00	restricted-fit evidence score of Eq. (12) along signature v_8, threshold tau(pct); group label shared with confusable faults	group: motor eccent
ity	R_cable = 0.112(1+0.5S) Ohm	stator resistance Rs (Ohm) +0.041, mechanical eff. eta_m -0.0107, volumetric eff. eta_v -0.00808	flow 0.00 current 0.00 pressure 0.00 power 0.05	restricted-fit evidence score of Eq. (12) along signature v_9, threshold tau(pct); group label shared with confusable faults	group: cable condu
tor area	R_cable = 0.112/(1-0.5S) Ohm	stator resistance Rs (Ohm) +0.0528, mechanical eff. eta_m -0.0121, volumetric eff. eta_v -0.00907	flow 0.00 current 0.00 pressure 0.00 power 0.07	restricted-fit evidence score of Eq. (12) along signature v_10, threshold tau(pct); group label shared with confusable faults	group: cable condu

https://doi.org/10.37965/jdmd.2026.1608

	Coupling backlash	b_gap = 5e-3(1+9S) rad	leakage coeff. C (1e-12) -0.00782, mechanical eff. eta_m -0.00679, stator resistance Rs (Ohm) +0.00558	flow 0.00 current 0.00 pressure 0.00 power 0.00	restricted evidence of Eq. (1 along sig v_11, thr tau(pct); label sha with con faults		resistance Rs (Ohm) +0.0323		label shared with confusable faults	
						nlet efficient	K_inlet = 1 + 2S	mechanical eff. eta_m -0.0311, volumetric eff. eta_v -0.0261, number of stages N -0.0236	flow 0.09 current 0.17 pressure 0.07 power 0.30	restricted-fit evidence score of Eq. (12) along signature v_18, threshold tau(pct); group label shared with confusable faults	group: inlet re
	Coupling misalignment angle	gamma = 0.35S rad	magnet flux psi_f (Wb) -0.0221, stator resistance Rs (Ohm) +0.0197, mechanical eff. eta_m -0.0156	flow 0.00 current 0.08 pressure 0.00 power 0.14	restricted evidence of Eq. (1 along sig v_12, thr tau(pct); label sha with con faults					
						wax ge	h x(1-0.5S); k_contact x(1+1.0S)	leakage coeff. C (1e-12) -0.0354, magnet flux psi_f (Wb) -0.0319, stator resistance Rs (Ohm) +0.0249	flow 0.07 current 0.15 pressure 0.02 power 0.28	restricted-fit evidence score of Eq. (12) along signature v_19, threshold tau(pct); group label shared with confusable faults	group: stator / pump
	Coupling transmission efficiency	eta_c = 0.96 -0.26S	magnet flux psi_f (Wb) -0.0419, stator resistance Rs (Ohm) +0.0307, mechanical eff. eta_m -0.023	flow 0.00 current 0.16 pressure 0.00 power 0.29	restricted evidence of Eq. (1 along sig v_13, thr tau(pct); label sha with con faults					
						lume n at	alpha_g = 0.45S	number of stages N -0.312, stator resistance Rs (Ohm) +0.193, magnet flux psi_f (Wb) -0.188	flow 0.96 current 0.38 pressure 0.24 power 0.52	restricted-fit evidence score of Eq. (12) along signature v_20, threshold tau(pct); group label shared with confusable faults	group: lock
	Stator wear depth	h = 40um + 0.08mm.S	leakage coeff. C (1e-12) +0.219, eccentricity e (mm) +0.183, number of stages N -0.141	flow 0.71 current 0.26 pressure 0.15 power 0.40	restricted evidence of Eq. (1 along sig v_14, thr tau(pct); label sha with con faults					
						flow ient	K_tub = 1 + S	leakage coeff. C (1e-12) -0.00823, mechanical eff. eta_m -0.00658, stator resistance Rs (Ohm) +0.00556	flow 0.00 current 0.00 pressure 0.00 power 0.00	restricted-fit evidence score of Eq. (12) along signature v_21, threshold tau(pct); group label shared with confusable faults	group: tubing
	Pump rotor eccentricity deviation	h x(1+0.6S); k_contact x(1+0.8S)	leakage coeff. C (1e-12) +0.0978, eccentricity e (mm) +0.064, number of stages N -0.0398	flow 0.14 current 0.21 pressure 0.06 power 0.38	restricted evidence of Eq. (1 along sig v_15, thr tau(pct); label sha with con faults					
						leak	A_leak = 1e-6.S m2	leakage coeff. C (1e-12) +0.144, eccentricity e (mm) +0.0878, number of stages N -0.0818	flow 0.35 current 0.08 pressure 0.04 power 0.14	restricted-fit evidence score of Eq. (12) along signature v_22, threshold tau(pct); group label shared with confusable faults	group: stator tubing
	Stator pitch change	pitch = 80(1+0.25S) mm	leakage coeff. C (1e-12) -0.0453, eccentricity e (mm) +0.0278, mechanical eff. eta_m -0.0194	flow 0.16 current 0.15 pressure 0.05 power 0.29	restricted evidence of Eq. (1 along sig v_16, thr tau(pct); label sha with con faults					
	Stator swelling	h x(1-0.6S); k_contact x(1+1.5S)	magnet flux psi_f (Wb) -0.0466, leakage coeff. C (1e-12) -0.0333, stator	flow 0.08 current 0.22 pressure 0.02 power 0.42	restricted evidence of Eq. (1 along sig v_17, thr tau(pct);					

Fig. 12 summarizes the two measured structures behind the accuracy: (a) the signature cosine matrix, whose block pattern is the a-priori counterpart of the realized confusion groups, and (b) the per-type correct-classification rate vs severity at 3% noise for representative faults, including the

https://doi.org/10.37965/jdmd.2026.1608

weak-subspace faults F8 and F21 whose curves remain at the floor.

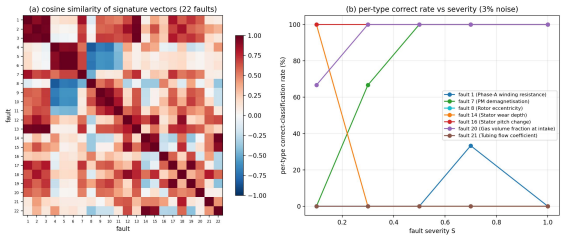


Fig. 12 (a) Cosine similarity of the 22 Stage-B signature vectors (22 × 22 matrix); (b) per-type correct-classification rate vs severity at 3% noise (3 cells per point, representative faults).

Table 7. Per-fault detection rate and per-type accuracy at 3% noise (15 cells per fault: 5 severities x 3 reps), and group-level accuracy. Group = measurably confusable set; group-level accuracy counts a prediction inside the true group as correct.

Fault	**Fault (physical meaning)**	**Detection rate**	**Per-type accuracy**	**Group (measurably confusable set)**
F1	Phase-A winding resistance (inter-turn A)	1.000	0.067	G1 inter-turn / coupling transmission
F2	Phase-B winding resistance (inter-turn B)	1.000	0.067	G1 inter-turn / coupling transmission
F3	Phase-C winding resistance (inter-turn C)	1.000	0.600	G1 inter-turn / coupling transmission
F4	Inter-phase insulation (A-B)	1.000	0.133	G2 inter-phase insulation
F5	Inter-phase insulation (A-C)	1.000	0.467	G2 inter-phase insulation
F6	Inter-phase insulation (B-C)	1.000	0.067	G2 inter-phase insulation
F7	PM demagnetisation	1.000	0.733	G3 PM demagnetisation
F8	Rotor eccentricity (bearing wear)	0.267	0.000	G4 motor rotor eccentricity
F9	Cable resistivity factor	1.000	0.000	G5 cable conduction
F10	Cable conductor area loss	0.867	0.600	G5 cable conduction
F11	Coupling backlash	0.200	0.067	G6 coupling backlash
F12	Coupling misalignment angle	1.000	0.000	G7 coupling misalignment
F13	Coupling transmission efficiency	1.000	0.267	G1 inter-turn / coupling transmission
F14	Stator wear depth	1.000	0.200	G8 stator wear / tubing leak
F15	Pump rotor eccentricity deviation	1.000	1.000	G9 pump eccentricity
F16	Stator pitch change	1.000	1.000	G10 pitch change
F17	Stator swelling	1.000	0.800	G11 stator swelling / pump wax
F18	Pump inlet flow coefficient	0.800	0.667	G12 inlet restriction
F19	Pump wax blockage	1.000	0.533	G11 stator swelling / pump wax
F20	Gas volume fraction at intake	1.000	0.933	G13 gas lock
F21	Tubing flow coefficient	0.067	0.000	G14 tubing friction
F22	Tubing leak area	1.000	0.867	G8 stator wear / tubing leak

https://doi.org/10.37965/jdmd.2026.1608

overal l	(330 cells)	0.873	0.412	group acc 0.694
	G1 inter-turn / coupling transmission {1,2,3,13}			0.767
	G2 inter-phase insulation {4,5,6}			0.956
	G3 PM demagnetisati on {7}			0.733
	G4 motor rotor eccentricity {8}			0.000
	G5 cable conduction {9,10}			0.733
	G6 coupling backlash {11}			0.067
	G7 coupling misalignment {12}			0.000
	G8 stator wear / tubing leak {14,22}			0.867
	G9 pump eccentricity {15}			1.000
	G10 pitch change {16}			1.000
	G11 stator swelling / pump wax {17,19}			0.867
	G12 inlet restriction {18}			0.667
	G13 gas lock {20}			0.933
	G14 tubing friction {21}			0.000

4.8 Solver benchmark and transient verification

Solver benchmark (Table 8): on 12 representative cells at 3% noise, multi-start L-BFGS-B (K = 3 starts) reaches objective 0.0972 in 4,036 function evaluations and 1.36 s per cell; DE with popsize 144 / 40 generations + polish reaches the same optimum (12/12 cells, mean difference < 1e-12) with 71,896 evaluations and 24.1 s; DE-lite (popsize 60 / 20 gen) with 16,293 evaluations and 5.15 s. The multi-start L-BFGS-B is the standard solver; DE is benchmarked for reference only and is not used by the diagnostic pipeline.

Table 8. Solver benchmark on 12 representative cells (3% noise). A = multi-start L-BFGS-B (K = 3); B = differential evolution (popsize 144, 40 generations) + polish; C = DE-lite (popsize 60, 20 generations) + polish. All three reach the same optimum on every cell (mean |df| < 1e-12). DE is benchmarked for reference only; the pipeline uses solver A.

Solver	Objective (min)	Function evaluations	Time (s per cell)
A L-BFGS-B multi-start	**0.09718**	**4036**	**1.36**
B DE 144 x 40 + polish	0.09718	71896	24.06
C DE 60 x 20 + polish	0.09718	16293	5.15

Transient verification (Table 9): integrating the coupled dq ODE with the pump load recomputed from the instantaneous speed at every step (PI-controlled VSI; the PI integral is carried as an ODE state with anti-windup), initialized at a perturbed state around the settled point (0.8 × the setpoint speed, 0.5 × the settled q-current and integrator charge), for 9 cases (healthy at 90/180/270 rpm; f14 at 90/270 rpm; f7/f20/f21 at 270 rpm; f17 at 180 rpm). All nine trajectories settle back to the quasi-steady limit: the speed stays within 1% of the setpoint from at most 28.5 s to the

https://doi.org/10.37965/jdmd.2026.1608

end of the 60 s window, the settled speed error is at most 0.14 rpm and the settled q-current error at most 0.0001 of the steady-state value. Every fixed point is locally exponentially stable: the linearized eigenvalues of the ODE at each fixed point are all strictly negative (slowest between -0.07 and -0.38 s^{-1}, evaluated with both one-sided Jacobians). Two features of the load curve are crossed during the approaches and shape the settling transients: (i) the Beggs-Brill traverse switches flow pattern at a flow of 8.3-8.9 m^3/d, collapsing the discharge pressure by 1.1-1.2 MPa, releasing the stall factor, and dropping the load torque by 16-21 N·m - the trajectories starting from 144 rpm cross the switch at 168.8 rpm (healthy) and 160.0 rpm (f17 swelling), and the f14 trajectory at 270 rpm (starting from 216 rpm) crosses it at 250.8 rpm, producing a transient speed excursion (Table 9) before the speed re-settles; (ii) the f20 gas-lock quasi-steady point at 270 rpm sits inside the stall-collapse region (dP = 7.16 MPa against the 5.5 MPa gas-lock cap, stall factor 0.70) on a mildly negative load slope (-0.21 N·m/rpm); its approach crosses the gas-lock cap-release jump at 258.6 rpm (a +6.7 N·m torque step as the discharge pressure rises off the cap, 6.85 to 7.32 MPa, stall factor 0.76 to 0.67), across which the trajectory bounces 14 times over about 2 s before escaping upward, yet the equilibrium remains stable (slowest eigenvalue -0.22 s^{-1}) because the electromechanical coupling supplies positive damping, and the trajectory settles there with the settled q-current agreeing with the quasi-steady map to within 1e-8 relative (integrated with fixed-step RK4, h = 5 ms, because the load iteration's flip-flopping across the cap branches defeats adaptive step control). The settled-sweep assumption therefore holds for all nine cases.

Table 9. Settled-vs-transient verification (Section 4.8): the coupled dq ODE with the pump load recomputed at the instantaneous speed (PI-controlled VSI; PI integral carried as an ODE state with anti-windup) is integrated from a perturbed state around the settled point (0.8 x setpoint speed, 0.5 x settled q-current and integrator charge) for 9 cases. 9 of the 9 trajectories settle back to the quasi-steady limit used by the experiment (settled speed error at most 0.14 rpm, settled q-current error at most 0.0001 of the steady-state value, settling time at most 28.5 s). t_settle = first time after which the speed stays within 1% of the setpoint to the end of the 60 s window; w_exc = maximum speed deviation from the setpoint during the approach (the excursion across the traverse flow-pattern switch); tail span = oscillation range over the last 25% of the window. Each returned trajectory is checked against one-grid-step re-solves with a fine explicit integrator (DOP853, rtol 1e-8, max step 1e-4; the F20 gas-lock case, whose load iteration flip-flops across the gas-lock cap branches and defeats adaptive step control, is integrated with fixed-step RK4 h = 5 ms and re-solved at h = 5e-5); worst mismatch < 1e-3.

Setpoint (rpm)	Fault state	iq_ss (A)	iq_settled (A)	rel. error iq	speed error (rpm)	t_settle (s)
90	healthy	86.997	86.995	0	0.00	19.0
180	healthy	102.568	102.569	0	0.00	16.3
270	healthy	118.849	118.837	0.0001	-0.09	26.9
270	f14 wear	104.069	104.062	0.0001	-0.02	16.7
270	f7 demag	132.054	132.035	0.0001	-0.14	28.5

https://doi.org/10.37965/jdmd.2026.1608

90	f14 wear	71.429	71.429	0	0.00	11.9
270	f20 gas	75.914	75.919	0.0001	0.00	19.5
270	f21 tubing	118.863	118.849	0.0001	-0.09	26.9
180	f17 swell	123.427	123.424	0	-0.01	16.6

4.9 Reduced-vs-full model discrepancy

Table 10 / Fig. 13: over 135 operating-condition × fault-state combinations (viscosity × 0.5/1.0/2.0 design, water cut 0.20/0.50/0.693, gas fraction 0.02/0.074/0.15, five fault states), the Stage-A-calibrated reduced model (fixed θ) has per-channel mean errors of [20.3, 16.7, 9.3, 26.8]% of the healthy nominal value (flow/current/pressure/power) and maximum [99.5, 43.4, 25.6, 62.2]%; refitting the 12 parameters at each condition (the best the reduced model can do) leaves mean [3.8, 6.7, 3.9, 10.5]% and maximum [16.0, 15.1, 8.0, 17.9]%. The 99.5% flow maximum is the μ × 0.5 / f14 corner - all nine half-viscosity wear cells collapse toward stall; the other 11 cells above 30% are also F14 cells (0.30-0.72 at μ × 1 and μ × 2, largest 0.72 at fw 0.693 / fg 0.15), and the remaining 115 of the 135 cells stay below 30% of nominal. Errors here are normalized by the healthy nominal value per channel (the inversion's own W-weighting) because instantaneous relative error diverges where a channel value approaches zero (mean instantaneous flow error 460% fixed / 63% refit - reported for completeness in the supplementary table). Interpretation: the reduced model's form error is bounded by ~18% of nominal per channel over the operating envelope, i.e. the surrogate is adequate for the diagnosis channels but the calibrated baseline should be re-estimated (Stage A re-run) when the operating point moves far from design - the two-stage pipeline's answer to operating-condition changes. The effects retained / approximated / omitted in the reduced map are listed in Section 2.7.

Table 10. Full-vs-reduced model discrepancy over 135 operating-condition x fault-state combinations (mu x 0.5/1/2 design, water cut 0.20/0.50/0.693, gas fraction 0.02/0.074/0.15, five fault states). Errors are per channel, max over the 12-speed sweep, normalized by the healthy nominal value of that channel (the inversion weighting); instantaneous relative errors are in the supplementary table.

fixed (Stage-A theta)	**mean**	**0.203**	**0.167**	**0.093**	**0.268**
fixed (Stage-A theta)	max	0.995	0.434	0.256	0.622
refit (12 params)	mean	0.038	0.067	0.039	0.105
refit (12 params)	max	0.160	0.151	0.080	0.179

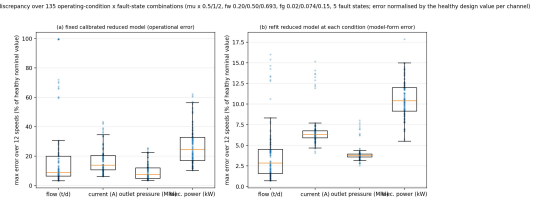


Fig. 13 Full-vs-reduced model discrepancy over 135 operating-condition × fault-state combinations: fixed calibrated reduced model vs refit, per channel.

4.10 Discussion of the detection boundary

The measurement set of four surface quasi-steady channels defines an information boundary: (a) phase-attributed electrical faults collapse to their dq projection (groups F1-F3, F4-F6); (b) series-resistance faults collapse to one direction (F9/F10 and, through the effective Rs, a component of every motor-side fault); (c) clearance faults

project onto the leakage composite $C_{leak}\cdot(e/E0)^3$ (F14/F15/F22 share its direction; F17/F19 mix it with the contact torque); (d) torque-transmission faults project onto the electrical channels through the torque constant (F12/F13, the latter collinear with the inter-turn group); (e) faults acting only through the weak subspace (B: F8; tubing friction: F21; backlash drag: F11) remain at the chance detection floor across severities (Section 4.3). Extending the channel set (vibration, phase-current waveforms, downhole pressure) would shrink the groups and lift the detection limits; within this paper's channel set, the reported group structure is the exact identifiability statement of the four-channel quasi-steady measurement.

5. Conclusions

This paper has presented a fault diagnosis methodology for ESPCP systems based on multi-physics forward modeling and Tikhonov-regularized inversion. The principal findings are as follows.

(1) An implemented multi-physics forward model of the ESPCP well - cable, PMSM with PI-controlled VSI, coupling, PCP, and Beggs-Brill tubing - was built, and its settled quasi-steady fixed point was verified against the coupled transient dynamics for all nine verification cases, including the severe-wear and gas-lock operating points near the traverse flow-pattern switch and the stall boundary (Section 4.8). Twenty-two fault parameters are defined by explicit perturbation functions (Table 2).

(2) Fault diagnosis was reformulated as Tikhonov-regularized inversion of a 12-parameter reduced quasi-steady map with L-curve-selected regularization (λ^* = 0.146/0.533/1.017/2.984 at 1/3/5/10% noise). The singular spectrum of the weighted Jacobian ($\kappa(A)$ = 6.2e7, five weak directions) quantifies what the four-channel quasi-steady set can and cannot identify, and the measured confusion groups state the resulting boundary.

(3) The two-stage pipeline - Stage A baseline calibration from one healthy sweep, Stage B restricted-fit evidence scoring with per-noise-level thresholds - was validated on 1,720 cells aggregated from a single master results table: 41.2% per-type accuracy and 87.3% detection at 3% noise, 69.4% group-level accuracy, and a false-alarm rate of 17.5%. The detection-limit faults - bearing wear F8 (acting through the weakest parameter direction) and backlash F11 / tubing friction F21 (whose full-model effects are themselves below the noise) - are reported at their measured detection limits rather than hidden in the aggregate.

(4) The fault-to-parameter mapping (Table 6) and the confusion groups {1,2,3,13}, {4,5,6}, {9,10}, {14,22}, {17,19} make explicit which fault distinctions the four surface channels cannot resolve: phase attribution of electrical faults, series-resistance families, and faults acting only through the weak parameter subspace.

(5) The 135-condition discrepancy study bounds the reduced-model surrogate error at ~18% of nominal per channel in the worst case after refitting (Table 10), motivating Stage-A re-calibration when the operating point moves.

(6) Future work: extending the channel set (vibration, phase-current waveforms, downhole pressure) to shrink the confusion groups and lift the detection limits, and field validation across wells with diverse operating conditions.

Nomenclature

θ	parameter vector: the 22 full-model fault parameters (Section 2.6), or the 12 reduced-model inversion parameters in the inversion context (Section 3.2)
x	scaled parameter vector, x = (θ − LO)/(HI − LO) ∈ [0,1]^12
x_cal	Stage-A calibrated healthy baseline (scaled)
v_i	signature direction of fault i (unit-normalized deviation)
score_i	restricted-fit evidence score of direction i (Eq. (12))
τ(pct)	detection threshold: 95th percentile of the healthy calibration draws
d, m	monitoring-data vector (48 components) / forward-model prediction
W	diagonal channel weighting, W = diag(1/|d_nom|)
J(x)	regularized objective function
λ	Tikhonov regularization parameter
A = W·J_x	weighted sensitivity (Jacobian) matrix
κ	condition number of A
S	fault-severity index (0-1)
s*	severity estimate of the winning signature direction
T_e, I_q	electromagnetic torque, q-axis current
Q_leak, h, Δp, μ	leakage rate, clearance height, pressure difference, viscosity
p_wh, Rs, ψ_f, B, D, e, pitch, N, η_v, η_m, C_leak, C_tub	the 12 reduced-model parameters (Table 4)
f_w	water cut (fraction)

Notation convention: all vector quantities (θ, d, m, u, and the velocity field v) are set in boldface throughout; scalars are set in italic; matrices are set in bold upright.

Conflict of Interest

The authors declare no conflicts of interest.

References

[1] WANG SJ, SUN H. Research of Real-time Monitoring Technology for Progressing Cavity Pumping Operation[C]//International Conference on Frontiers of Manufacturing and Design Science (ICFMD2010). FRONTIERS OF MANUFACTURING AND DESIGN SCIENCE, PTS 1-4, 2011, 44-47: 4094-4099. DOI: 10.4028/www.scientific.net/AMM.44-47.4094.

[2] MÜLLER J, LEONOW S, MÖNNIGMANN M. Optimal stator adjustment with minimal sensor requirements for variable stator progressing cavity pumps[C]//22nd World Congress of the International Federation of Automatic Control (IFAC). IFAC PAPERSONLINE, 2023, 56(2): 4614-4619. DOI: 10.1016/j.ifacol.2023.10.969.

[3] ULMANU V, BADICIOIU M, CALTARU MM. Fatigue Behaviour of Elastomers Used for Progressive Cavity Pumps Manufacture[J]. MATERIALE PLASTICE, 2015, 52(4): 494-499.

[4] MRINAL KR, SAMAD A. Performance prediction of kinetic and screw pumps delivering slurry[J]. PROCEEDINGS OF THE INSTITUTION OF MECHANICAL ENGINEERS PART A-JOURNAL OF POWER AND ENERGY, 2018, 232(7): 898-911. DOI: 10.1177/0957650918760161.

[5] DENNEY D. Fourteen years of progressing cavity pumps[J]. JOURNAL OF PETROLEUM TECHNOLOGY, 2003, 55(5): 63-65.

[6] TIMASHEV EO, LATYPOV BM, URAZAKOV KR. TRIBOTECHNICAL CHARACTERISTICS OF THE SCREW PAIR OF THE PROGRESSING CAVITY PUMP[J]. BULLETIN OF THE TOMSK POLYTECHNIC UNIVERSITY-GEO ASSETS ENGINEERING, 2021, 332(6): 19-27. DOI: 10.18799/24131830/2021/06/3232.

[7] CHEN J, LIU H, WANG FS, et al. Numerical prediction on volumetric efficiency of progressive cavity pump with fluid-solid interaction model[J]. JOURNAL OF PETROLEUM SCIENCE AND ENGINEERING, 2013, 109: 12-17. DOI: 10.1016/j.petrol.2013.08.019.

[8] ZU HY, CUI ZY, XU MH, et al. Experimental study on fatigue life of oil pump rubber stator under oil immersion and temperature effects[J]. OCEAN ENGINEERING, 2026, 358: 125698. DOI: 10.1016/j.oceaneng.2026.125698.

[9] KIM DY, KIM WB, CHOI J, et al. Modeling consolidation of wax deposition for progressive cavity pump using computational fluid dynamics[J]. ENGINEERING SCIENCE AND TECHNOLOGY-AN INTERNATIONAL JOURNAL-JESTECH, 2023, 41: 101384. DOI: 10.1016/j.jestch.2023.101384.

[10] MÜLLER J, LEONOW S, SCHULZ J, et al. Online wear detection for joints in progressing cavity pumps[C]//16th IFAC Symposium on Advanced Control of Chemical Processes (ADCHEM). IFAC PAPERSONLINE, 2021, 54(3): 188-193. DOI: 10.1016/j.ifacol.2021.08.240.

[11] REN WJ, LU Y, DONG HL, et al. Fault Diagnosis of Progressing Cavity Pump Well Based on Wavelet Package and Elman Neural Network[C]//8th World Congress on Intelligent Control and Automation (WCICA). 2010 8TH WORLD CONGRESS ON INTELLIGENT CONTROL AND AUTOMATION (WCICA), 2010: 2741-2745. DOI: 10.1109/WCICA.2010.5554848.

[12] DONG KX, LI QE, ZHANG ZH, et al. Submersible Screw Pump Fault Diagnosis Method Based on a Probabilistic Neural Network[J]. JOURNAL OF APPLIED SCIENCE AND ENGINEERING, 2022, 25(6): 915-923. DOI: 10.6180/jase.202212_25(6).0002.

[13] LV XR, LUO X, WANG SJ, et al. Short-term Prediction on the Time Series of PCP Speed Based on Elman Neural Network[C]//International Conference on Advanced Materials Design and Mechanics (ICAMDM 2012). ADVANCED MATERIALS DESIGN AND MECHANICS, 2012, 569: 749-753. DOI: 10.4028/www.scientific.net/AMR.569.749.

[14] JIANG MZ, CHENG TC, DONG KX, et al. Fault diagnosis method of submersible screw pump based on random forest[J]. PLOS ONE, 2020, 15(11): e0242458. DOI: 10.1371/journal.pone.0242458.

[15] LIU XF, SHAN JP, LIU CH, et al. An Operating Condition Diagnosis Method for Electric Submersible Screw Pumps Based on CNN-ResNet-RF[J]. PROCESSES, 2025, 13(7): 2043. DOI: 10.3390/pr13072043.

[16] OU YJ, CHEN WZ, GAO XY, et al. A novel fault diagnosis method for electric submersible progressing cavity pumps: Addressing imbalanced mixed-domain datasets with complex subdomains[J]. NEUROCOMPUTING, 2025, 655: 131420. DOI: 10.1016/j.neucom.2025.131420.

[17] PALADINO EE, LIMA JA, PESSOA PAS, et al. A computational model for the flow within rigid stator progressing cavity pumps[J]. JOURNAL OF PETROLEUM SCIENCE AND ENGINEERING, 2011, 78(1): 178-192. DOI: 10.1016/j.petrol.2011.05.008.

[18] OTTE J, LEONOW S, MÖNNIGMANN M. Integrating Fluid Rheology into a Physics-Based Torque Model of Progressive Cavity Pumps for Water Treatment Applications[C]//3rd IFAC Workshop on Control Methods for Water Resource Systems (CMWRS). IFAC PAPERSONLINE, 2025, 59(33): 19-24. DOI: 10.1016/j.ifacol.2025.12.416.

[19] BECERRA D, ZAMBRANO A, ASUAJE M, et al. Experimental and CFD modeling of a progressive cavity pump (PCP) using overset unstructured mesh part 2: Two-phase flow.[J]. GEOENERGY SCIENCE AND ENGINEERING, 2024, 233: 212601. DOI: 10.1016/j.geoen.2023.212601.

[20] GAMBOA J, OLIVET A, ESPIN S. New approach for modeling progressing-cavity-pump performance[J]. JOURNAL OF PETROLEUM TECHNOLOGY, 2004, 56(5): 56-57.

[21] STARKE L, HOPPE AF, SARTORI A, et al. Interference recommendation for the pump sizing process in progressive cavity pumps using graph neural networks[J]. SCIENTIFIC REPORTS, 2023, 13(1): 16884. DOI: 10.1038/s41598-023-43972-4.

[22] GU DW, YANG JL, LONG Z, et al. Optimal Design of Diameter Reduction for Progressive Cavity Pumps based on Interference Values[J]. CHEMICAL ENGINEERING & TECHNOLOGY, 2025, 48(9): e70105. DOI: 10.1002/ceat.70105.

[23] PAN L, TAN JZ. Numerical investigation of contact stress between rotor and stator in a two-lead progressing cavity pump[J]. JOURNAL OF PETROLEUM SCIENCE AND ENGINEERING, 2015, 134: 176-185. DOI: 10.1016/j.petrol.2015.07.026.

[24] HANSEN PC. Analysis of discrete ill-posed problems by means of the L-curve[J]. SIAM REVIEW, 1992, 34(4): 561-580. DOI: 10.1137/1034115.

Conflict of Interest

The authors declare no conflicts of interest.