Akademie věd České republiky
Ústav teorie informace a automatizace, v.v.i.

Academy of Sciences of the Czech Republic
Institute of Information Theory and Automation

RESEARCH REPORT

Kamil Dedecius

Centralized Estimation of Adhesion Loss in Wheel–Rail System Using Variational Bayes and Variational Message Passing

No. 2329 April 10, 2013

ÚTIA AV ČR, P.O.Box 18, 182 08 Prague, Czech Republic
Tel: +420 286892337, Fax: +420 266052570, Url: http://www.utia.cas.cz,
E-mail: dedecius@utia.cas.cz

List of physical variables

Notation	Meaning
i_α, i_β	Currents in $\alpha - \beta$ reference frame.
u_α, u_β	Input voltages in $\alpha - \beta$ reference frame.
ω_{me}	Electrical angular rotor speed.
ϑ_e	Electrical angular rotor position.
L_s, L_d	Stator inductances in $d - q$ reference frame.
$t, \Delta t$	Discrete time instant, time period.
R_s	Stator resistance.
p_p	Number of pole-pairs.
Ψ_{pm}	Flux linkage (permeability).
B	Friction.
J	Moment of inertia.
k_p	Park constant.
T_L	Torque load.
F_f	Friction force (static or kinetic)
F_n	Force in the normal direction
ν	Friction coefficient (static or kinetic)

List of abbreviations

Abbreviation	Meaning
EKF	Extended Kalman filter
VB	Variational Bayes
VMP	Variational message passing
PMSM	Permanent magnet synchronous motor
DTC	Direct torque control

1 Introduction

The sensing and control of railway traction devices like the permanent magnet synchronous motors (PMSM) represent very complex tasks, solved by numerous more or less elaborated methods. Traditionally, each wheel is controlled independently, using separate electric motorization with balanced transmitted torques. An identical torque on each motor is usually imposed by using a direct torque control (DTC). The difficulty of controlling such a system is its highly nonlinear character of the traction forces expressions. The loss of traction of one or more wheels, that is the loss of adhesion in the wheel–rail system, is likely to destabilize the vehicle Soylu [2011]. Furthermore, it can lead to threshold values of relevant electric variables in motors.

In this report, we focus on centralized estimation of wheel adhesion loss. The proposed methods, based on variational (mean-field) inference, exploit variables resulting from centralized sensor-less recursive estimation of PMSM drive state.

The report first discusses relevant modelling issues, connected with the traction/adhesion forces. Then, the extended Kalman filter (EKF), providing sensor-less estimation of drive state is presented. Its full treatment can be found, e.g., in Šmídl [2013]. The ensuing description of centralization of EKFs for several drives is followed by estimation of adhesion loss using two methods, allied in the variational framework: the variational Bayes and the variational message passing. The resulting algorithms are demonstrated on examples.

2 Characterization of adhesion loss

In a railway context, adhesion is the friction available to transmit tangential force between railway wheel and rail. It is always lower or (theoretically equal) to the maximum tangential force produced by a driving wheel before slipping,

$$F_f = \nu F_n$$

where F_f is the (static or kinetic) friction force, ν is the (static or kinetic) friction coefficient and F_N is the force in the normal direction (i.e. the weight) on a non-slipping wheel. Usually the static friction is greater than the kinetic friction, implying that the force needed to start sliding dominates the force needed to maintain it. While high adhesion is the essential factor for breaking, it may generate problems, e.g., in sharp curves, and in extreme cases lead even to derailment [Baek et al., 2008].

The adhesion coefficient is very sensitive to the environmental conditions, such as speeds, axle-loads, wheel and rail profiles, contamination of contact surfaces, weather etc. [Zhang et al., 2002]. Extensive research effort has been given the problem of adhesion loss caused by water, including one contaminated by various, mostly wear debris (even non-Newtonian), rust, oil and its mixtures containing chemically active compounds [Beagley and Pritchard, 1975, Beagley et al., 1975a,b]. If water firm exists in the wheel-rail contact surfaces, the adhesion coefficients decrease with an increase in speed and the traction forces decrease with an increase in exciting frequency. Furthermore, the adhesion coefficients decrease, independently of speed and contamination, with an increase in axle load [Zhang et al., 2002]. This might prove a significant factor in multiple unit trains, metro and tram units, where the axle load can be highly varying in time and space. Another frequent cause of traction problems are leaves on the rail lines, forming a low-adhesion layer when crushed by passing wheels. Also the toppling conditions in turns, when the centrifugal acceleration is sufficient to begin to lift the inner wheel, have significant impact on adhesion.

Tracking of a traction drive torque can be used for detection of a wheel slip. As described above, adhesion is influenced by a large number of phenomena at the rail lines. This imposes a limiting factor to distributed estimation of adhesion loss: the train/tram/metro wheels may face more or less different conditions at the same time and even if the torque control is employed, the true torque values are unlikely the same at individual wheels (disregarding the slipping wheels). One way around this problem would be exploiting a suitable probabilistic distribution of torque, e.g. the normal one, and view the torque value of a slipping wheel as an outlier. Similarly, we can base the distributed slip detection on classification on base of torque value. The latter approach is adopted in this report.

3 State-space estimation of PMSM drive

Let us now deal with estimation of PMSM drive state variables using the extended Kalman filter. We begin with a concise description of EKF, followed by its application to PMSM and transition to centralized estimation of several PMSM drives.

3.1 Basic of extended Kalman filter

Suppose the existence of a nonlinear dynamic system with a real single- or multivariate inner state variable x_t influencing the real single- or multivariate system output (measurement, observation) y_t, observed at discrete time instants $t = 1, 2, \ldots$ Furthermore suppose, that there exist differentiable functions f, h such that

$$\begin{aligned} x_t &\sim \mathcal{N}\left(f(x_{t-1}), Q\right) \\ y_t &\sim \mathcal{N}\left(h(x_t), R\right) \end{aligned} \tag{1}$$

where $\mathcal{N}(\mu, \Sigma)$ denotes normal distribution with mean value μ and covariance Σ, both of them having appropriate dimensions according to the argument. If f, h are linear functions, then the celebrated Kalman filter dominates solutions of the system (1). Under nonlinearity, the requirement of differentiability of f and h still provides ways to its analytic treatment via local linearization. Then, the extended Kalman filter (EKF) is a popular choice. We focus on the EKF. Its first-order linearization variant exploiting Taylor series with appropriate Jacobians is easily derived. We skip the derivation; it can be found, e.g., in Simon [2006].
The EKF works as follows: First, it is initialized with

$$\begin{aligned} \hat{x}_0^+ &= \mathbb{E}[x_0] \\ P_0^+ &= \mathbb{E}[(x_0 - \hat{x}_0^+)(x_0 - \hat{x}_0^+)']. \end{aligned} \tag{2}$$

That is, the initial estimates of state and estimate covariances are set. Then, the following relations provide *prediction* of prior ('-') estimates of P_t and $\hat{x}_t$, followed by update yielding their posterior ('+') estimates:

$$P_t^- = A_{t-1}P_{t-1}^+A_{t-1}' + Q_t \tag{3}$$
$$\hat{x}_t^- = f(\hat{x}_{t-1}^+) \tag{4}$$
$$K_t = P_t^- C_t'(C_t P_t^- C_t' + R_t)^{-1} \tag{5}$$
$$\hat{x}_t^+ = \hat{x}_t^- + K_t\left[y_t - h(\hat{x}_t^-)\right] \tag{6}$$
$$P_t^+ = (I - K_t C_t)P_t^- \tag{7}$$

where

$$A_{t-1} = \frac{\mathrm{d}}{\mathrm{d}x_t} f(x_t)\bigg|_{\hat{x}_{t-1}^+} \tag{8}$$
$$C_t = \frac{\mathrm{d}}{\mathrm{d}x_t} h(x_t)\bigg|_{\hat{x}_t^-} \tag{9}$$

are derivatives of the nonlinear model (1) and K_t denotes the Kalman gain. The matrix inversion in (5) usually calls for numerically stable evaluation, e.g. in the Cholesky factorization. The predictive distribution for y_t given previous observations $y_{1:t-1}$ is univariate normal,

$$y_t|y_{1:t-1} \sim \mathcal{N}\left(h(\hat{x}_{t-1}), R\right). \tag{10}$$

3.2 Application of EKF model to PMSM drive

For the purpose of state estimation, the discretized model of PMSM is transformed into a stationary $\alpha-\beta$ reference frame [Šmídl, 2013]. Further reduction of model complexity is reached by a common simplification, imposing the equality $L_s = L_d$. Since usually $|L_d/L_q| \leq 0.1$, the induced error is negligible in most applications.

Let us denote the measured values of currents $\bar{i}_{\alpha,t}$ and $\bar{i}_{\beta,t}$ and assume, that the associated measurements errors are $\varepsilon_{\alpha,t} \sim \mathcal{N}(0, r_\alpha)$ and $\varepsilon_{\beta,t} \sim \mathcal{N}(0, r_\beta)$, respectively. Furthermore, let $u_{\alpha,t}$ and $u_{\beta,t}$ be the reconstructed values of the input voltage. The resulting PMSM state-space model then evaluates the column state vector $x_{t+1} = [i_{\alpha,t+1}, i_{\beta,t+1}, \omega_{me,t+1}, \vartheta_{e,t+1}]'$ and the column vector of measurements $y_t = [\bar{i}_{\alpha,t}, \bar{i}_{\beta,t}]'$ as follows:

State equations of $f(x_t)$

$$\begin{aligned}
i_{\alpha,t+1} &= \theta_a i_{\alpha,t} + \theta_b \omega_{me,t}\sin\vartheta_{e,t} + \frac{\Delta t}{L_s}u_{\alpha,t} + \xi_\alpha \\
i_{\beta,t+1} &= \theta_a i_{\beta,t} - \theta_b \omega_{me,t}\cos\vartheta_{e,t} + \frac{\Delta t}{L_s}u_{\beta,t} + \xi_\beta \\
\omega_{me,t+1} &= \theta_d \omega_{me,t} + \theta_e\left(i_{\beta,t}\cos\vartheta_{e,t} - i_{\alpha,t}\sin\vartheta_{e,t}\right) - \frac{p_p}{J}T_L\Delta t + \xi_\omega \\
\vartheta_{e,t+1} &= \vartheta_{e,t} + \omega_{me,t}\Delta t + \xi_\vartheta
\end{aligned} \tag{11}$$

where

$$\theta_a = 1 - \frac{R_s}{L_s}\Delta t, \qquad \theta_b = \frac{\Psi_{pm}}{L_s}\Delta t \tag{12}$$
$$\theta_d = 1 - \frac{B}{J}\Delta t, \qquad \theta_e = \frac{k_p p_p^2 \Psi_{pm}}{J}\Delta t \tag{13}$$

Measurement equations of $h(x_t)$

$$\begin{aligned} \bar{i}_{\alpha,t} &= i_{\alpha,t} + \varepsilon_{\alpha,t} \\ \bar{i}_{\beta,t} &= i_{\beta,t} + \varepsilon_{\beta,t} \end{aligned} \tag{14}$$

In order to use the extended Kalman filter (Section 3.1), one has to determine the derivatives (9) of state equations (11). The resulting matrices A_{t-1} and C_t are as follows:

$$A_{t-1} = \frac{\mathrm{d}}{\mathrm{d}x_t} f(x_t)\bigg|_{\hat{x}^+_{t-1}} = \begin{pmatrix} \theta_a & 0 & \theta_b \sin\vartheta_{e,t} & \theta_b \omega_{me,t} \cos\vartheta_{e,t} \\ 0 & \theta_a & -\theta_b \cos\vartheta_{e,t} & \theta_b \omega_{me,t} \sin\vartheta_{e,t} \\ \theta_e \sin\vartheta_{e,t} & \theta_e \cos\vartheta_{e,t} & \theta_d & -\theta_e (i_{\beta,t} \sin\vartheta_{e,t} + i_{\alpha,t} \cos\vartheta_{e,t}) \\ 0 & 0 & \Delta t & 1 \end{pmatrix}$$

$$= \begin{pmatrix} \theta_a & 0 & \theta_b \sin\vartheta_{e,t} & \theta_b \omega_{me,t} \cos\vartheta_{e,t} \\ 0 & \theta_a & -\theta_b \cos\vartheta_{e,t} & \theta_b \omega_{me,t} \sin\vartheta_{e,t} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & \Delta t & 1 \end{pmatrix}$$

and

$$C_t = \frac{\mathrm{d}}{\mathrm{d}x_t} h(x_t)\bigg|_{\hat{x}^-_t} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}. \tag{15}$$

The terms $\theta_a, \theta_b, \theta_d$ and θ_e are identical with those in the state equations (11).

3.2.1 Initialization

The use of state-space models expects the setting of the initial covariance matrices. The state covariance matrix suitable for the concrete PMSM application is a symmetric square matrix with diagonal

$$\operatorname{diag} Q_0 = [0.05, 0.05, 0.001, 0.0001, 0.1]'$$

and zeros elsewhere. The increased dimension of covariance matrices is due to incorporated estimation of torque, made simply by addition of a random walk on related variable, cf. (11).

The white noise covariance is set with 0.01 in places of diagonal elements and zeros elsewhere, as it is assumed low and uncorrelated. The initial estimates covariance is a diagonal matrix with low integers on diagonal.

3.3 Centralized estimation of several PMSM drives

The previously described approach to state estimation of a single PMSM drive can be easily extended to centralized independent estimation of several drives at once. If we consider these drives independent, that is, they have independent states and measurements, we can proceed with a basic algebraic modifications. For N drives, the state and measurement variables are augmented as follows

$$x_t = \begin{bmatrix} x_{1,t} \\ \vdots \\ x_{N,t} \end{bmatrix} \qquad y_t = \begin{bmatrix} y_{1,t} \\ \vdots \\ y_{N,t} \end{bmatrix} \tag{16}$$

where $x_{i,t}$ and $y_{i,t}$ are state and measurement vectors related to the ith drive. The independence of evaluations is achieved by block-diagonal matrices with non-overlapping rows/columns,

$$A_t = \text{diag}(A_{1,t}, \dots, A_{N,t}) \qquad \text{and} \qquad C_t = \text{diag}(C_{1,t}, \dots, C_{N,t}) \tag{17}$$

for matrices A and C. Identical treatment is made for all other terms involved, i.e. P, K, R, Q. Remind that this approach neglects the simple fact that part of the state variables (namely the speed $\omega_{i,t}$ and the torque load $T_{L,i,t}$) are common for all drives $i = 1, \cdots, N$ under ideal conditions. However, the reasons discussed in Section 2 advocate sticking with independent state evaluation, accompanied by subsequent adhesion loss estimation. We will focus on it in the ensuing sections of this report.

4 Estimation of adhesion loss

This section is devoted to estimation of adhesion loss of a wheel attached to a PMSM drive. We propose two methods based on variational (mean-field) inference – the variational Bayesian inference [Jaakkola and Jordan, 2000] and the variational message passing [Winn and Bishop, 2005]. Both of them exploit the results of the centralized EKF.

The following subsections briefly introduce the principle of each method, followed by its application to the issue considered.

4.1 Variational Bayes

The variational Bayesian (VB) inference, rooted in the field of calculus of variations, serves for analytic approximation of the posterior pdf of parameters and potentially other latent variables [Jaakkola and Jordan, 2000]. Let us denote $Z = (Z_1, \dots, Z_n)$ as the set comprising both parameters and latent variables. The goal is to find analytically tractable approximation $q(Z)$ of $f(Z|X)$. It is possible to write

$$\log f(X) = \mathcal{L}(q) + \mathcal{D}(q||f), \tag{18}$$

where

$$\mathcal{D}(q||f) = -\int q(Z) \log \frac{f(Z|X)}{q(Z)} dZ$$
$$\mathcal{L}(q) = \int q(Z) \log \frac{f(Z|X)}{q(Z)} dZ$$

are the Kullback-Leibler divergence [Kullback and Leibler, 1951] of q and f and the variational lower bound, respectively. Note, that this order of arguments makes the Kullback-Leibler divergence the case of zero-forcing case of α-divergence family. As a result, the (suboptimal) resulting distribution has most of its mass concentrated at a mode of the unknown true distribution, while other modes are neglected.

Unlike in the celebrated EM algorithm [Dempster et al., 1977], the elements of Z are factorized into M independent factors $Z_i, i = 1, \dots, M$, such that

$$q(Z) = \prod_{i=1}^{M} q_i(Z_i). \tag{19}$$

This, put back into (18) yields

$$\begin{aligned}\mathcal{L}(q) &= \int \prod_i q_i(Z_i) \left[\log f(X,Z) - \sum_i \log q_i(Z_i) \right] dZ \\ &= \int q_j(Z_j) \mathbb{E}_{i\neq j}[\log f(X,Z)] dZ \\ &- \mathbb{E}_j[\log q_j(Z_j)] + \text{const.}\end{aligned}$$

where

$$\mathbb{E}_{i\neq j}[\log f(X,Z)] = \int \log f(X,Z) \prod_{i\neq j} q_i(Z_i) dZ_i.$$

This directly yields the VB-optimal factors

$$\log q_j^*(Z_j) = \mathbb{E}_{i\neq j}[\log f(X,Z)] + \text{const.}$$

The additive constant changes to multiplicative in exponentiation, providing the solution

$$q_j^*(Z_j) \propto \exp\left\{ \mathbb{E}_{i\neq j}\left[\log f(X,Z)\right]\right\}. \tag{20}$$

The resulting algorithm is very similar to the expectation-maximization, but unlike it, VB computes the posterior distributions of all parameters. The expectations are taken with respect to variables not in the current factor, which, in turn, are recomputed in the same way. The algorithm is guaranteed to converge and, under convexity of the lower bound, to the global maximum [Boyd and Vandenberghe, 2004].

It is necessary to stress that the variational Bayesian method provides analytic approximations of the posterior distribution of parameters and latent variables. The sacrifice is their factorized treatment (18), neglecting the dependency properties carried by the true joint posterior pdfs. An alternative *expectation propagation* algorithm [Minka, 2001] overcomes this issue by exploiting reversed order of pdfs in the Kullback-Leibler divergence in (18). The price is elevated level of computational difficulties.

4.2 Application of VB to estimation of adhesion loss

The variational estimation of adhesion loss is based on an *external observer model*, in which a central unit has access to angular speeds $\omega_{i,t} \equiv \omega_{me,i,t}$, torque loads $T_{i,t} \equiv T_{L,i,t}$ or both, for several PMSM drive indexed by integers $i = 1, \dots, N$. Obviously it does not matter whether these variables are directly measured or estimated by a number of extended Kalman filters presented in Section 3.2).

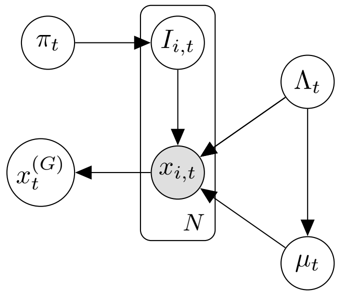

Figure 1: Graphical representation of the situation: The observation of the angular speed $\omega_{i,t}$ or torque $T_{i,t}$ or both, at time t, are represented by a variable $x_{i,t}$. The particular observations come either from a PMSM device operating under normal conditions with latent dichotomous variable $I_{i,t} = 0$, or under the loss of adhesions with $I_{i,t} = 1$. The probability of $I_{i,t} = 1$ is driven by a coefficient π_t. The Bayesian setting of the problem exploits the normal distribution for observations $x_{i,t}$ and assigns a prior distribution for its parameters, namely mean value μ_t and precision Λ_t. Finally, a global value $x_t^{(G)}$ can be produced.

To distinguish between normal operating conditions and adhesion loss, we introduce a dichotomous latent indicator $I_{i,t} \in \{0, 1\}$, where 0 means no adhesion loss and 1 stands for loss. In addition, we define $I = (I_1, \dots, I_N)$ the vector of indicators for PMSM drive $1, \dots, N$. Also, we drop time index t for easier reading.

To summarize, if π is a probability of adhesion loss, then

$$I_i = \begin{cases} 0 & \text{if } \textit{no loss of adhesion}, \\ 1 & \text{if } \textit{adhesion is lost}. \end{cases} \tag{21}$$

$$p(I|\pi) = \prod_{i=1}^{N} \pi^{I_i} (1 - \pi)^{(1-I_i)} \tag{22}$$

$$p(X|I, \mu, \Lambda) = \prod_{i=1}^{N} p(x_i|\mu_0, \Lambda_0)^{1-I_i} p(x_i|\mu_1, \Lambda_1)^{I_i} \tag{23}$$

where μ_j and $\Lambda_j, j \in \{0, 1\}$ are mean and precision of corresponding normal distributions under regular conditions (0) and adhesion loss (1), respectively.

In concordance with the Bayesian realm, we introduce prior distributions over the parameters π and μ, Λ. A particularly plausible is the following choice:

- The weight π follows the ordinary beta distribution with nonnegative real parameters α_0, α_1,

  $$\pi \sim \mathcal{B}(\alpha_0, \alpha_1), \tag{24}$$

  with a pdf

  $$p(\pi) = \frac{1}{B(\alpha_0, \alpha_1)} \pi^{\alpha_1 - 1} (1 - \pi)^{\alpha_0 - 1} \tag{25}$$

  $$\propto \pi^{\alpha_1 - 1} (1 - \pi)^{\alpha_0 - 1}, \tag{26}$$

  where $B(\alpha_0, \alpha_1)$ is a beta function.

- The vector of means $\mu = (\mu_0, \mu_1)$ and precisions $\Lambda = (\Lambda_0, \Lambda_1)$ follow, under independences among 0th and 1st elements, the form of a product of normal-Wishart (or normal-gamma) distributions, such that

  $$\mu|\Lambda \sim \mathcal{N}(\mu|m^-, (\beta^-\Lambda)^{-1})$$
  $$\Lambda \sim \mathcal{W}(\Lambda|W^-, \nu^-),$$

  yielding a compound pdf

  $$p(\mu, \Lambda) = p(\mu_0|m^-, (\beta^-\Lambda_0)^{-1})p(\Lambda_0|W^-, \nu^-) \tag{27}$$
  $$\times\, p(\mu_1|m^-, (\beta^-\Lambda_1)^{-1})p(\Lambda_1|W^-, \nu^-) \tag{28}$$

The variational distribution exploits the joint distribution of all of the variables,

$$p(X, I, \pi, \mu, \Lambda) = p(X|I, \mu, \Lambda)p(I|\pi)p(\pi)p(\mu|\Lambda)p(\Lambda) \tag{29}$$

The joint variational distribution of parameters and latent variables with pdf $q(I, \pi, \mu, \Lambda)$ factorizes between the latent indicator part and parameters as

$$q(I, \pi, \mu, \Lambda) = q(I)q(\pi, \mu, \Lambda). \tag{30}$$

$$\begin{aligned}\log q^*(I) &\propto \mathbb{E}_{\pi,\mu,\Lambda}\left[\log p(X, I, \pi, \mu, \Lambda)\right] \\ &\propto \mathbb{E}_{\pi}\left[\log p(I|\pi)\right] + \mathbb{E}_{\mu,\Lambda}\left[\log p(X|I, \mu, \Lambda)\right] \\ &\propto \sum_{i=1}^{N}(1 - I_i)\log \rho_{0,i} + I_i \log \rho_{1,i},\end{aligned} \tag{31}$$

where

$$\begin{aligned}\log \rho_{0,i} = \mathbb{E}\left[\log(1 - \pi)\right] + \frac{1}{2}\mathbb{E}\left[\log\det \Lambda_0\right] - \frac{D}{2}\log\tau \\ - \frac{1}{2}\mathbb{E}_{\mu_0,\Lambda_0}\left[(x_i - \mu_0)^T\Lambda_0(x_i - \mu_0)\right]\end{aligned} \tag{32}$$

and where $D = \dim x_i$ for all i and $\tau = 2\pi = 6.283\ldots$
Analogously, we define $\log \rho_{1,i}$,

$$\begin{aligned}\log \rho_{1,i} = \mathbb{E}\left[\log \pi\right] + \frac{1}{2}\mathbb{E}\left[\log\det \Lambda_1\right] - \frac{D}{2}\log\tau \\ - \frac{1}{2}\mathbb{E}_{\mu_1,\Lambda_1}\left[(x_i - \mu_1)^T\Lambda_1(x_i - \mu_1)\right].\end{aligned} \tag{33}$$

Taking the exponential of (31) we obtain

$$q^*(I) \propto \prod_{i=1}^{n}\left[\rho_{0,i}\right]^{1-I_i}\left[\rho_1\right]^{I_i} \tag{34}$$

which takes the same functional form as the prior $p(I|\pi)$ in (22). The posterior expected value of the adhesion loss indicator I_i given the data X simply takes the form

$$\mathbb{E}\left[I_i\right] = \frac{\rho_{1,i}}{\rho_{0,i} + \rho_{1,i}} \tag{35}$$

which corresponds to the posterior probability that the adhesion is lost at the ith PMSM drive.

Now, let us consider the second factor $q(\pi, \mu, \Lambda)$ in the variational posterior distribution. According to the principle of variational inference, we obtain

$$\begin{aligned}\log q^*(\pi, \mu, \Lambda) \propto \log p(\pi) + \log p(\mu_0, \Lambda_0) + \log p(\mu_1, \Lambda_1) + \mathbb{E}_I\left[\log p(I|\pi)\right] \\ + \sum_{i=1}^{n} \mathbb{E}\left[I_i\right]\left[\log p(x_i|\mu_0, \Lambda_0) + \log p(x_i|\mu_1, \Lambda_1)\right],\end{aligned} \tag{36}$$

where the last two term in brackets are pdfs of corresponding normal distributions. Careful examination of (36) revels two groups of terms, one involving only π and the other only μ and Λ. Thus it is possible to further factorize $q(\pi, \mu, \Lambda)$ into separate factors,

$$q(\pi, \mu, \Lambda) = q(\pi) q(\mu_0, \Lambda_0) q(\mu_1, \Lambda_1).$$

The terms involving π in (36) produce

$$\begin{aligned}\log q^*(\pi) \propto (\alpha_0 - 1)\log(1-\pi) + (\alpha_1 - 1)\log \pi \\ + \sum_{i=1}^{N} \left\{(1 - \mathbb{E}[I_i])\log(1-\pi) + \mathbb{E}[I_i]\log \pi\right\}.\end{aligned} \tag{37}$$

Taking the exponential reveals in $q^*(\pi)$ the pdf of a beta distribution with parameters

$$\begin{aligned}\alpha_0^+ &= \alpha_0 + N_0 \\ \alpha_1^+ &= \alpha_1 + N_1\end{aligned}$$

where

$$\begin{aligned}N_0 &= \sum_{i=1}^{N}(1 - \mathbb{E}[I_i]) \\ N_1 &= \sum_{i=1}^{N}\mathbb{E}[I_i] = N - N_0.\end{aligned}$$

represent the effective numbers of drives without (0) or with (1) adhesion loss. Similarly, reading off the terms involving (μ_0, Λ_0) and (μ_1, Λ_1), respectively, reveals two normal-Wishart distributions given by

$$q^*(\mu_k, \Lambda_k) = p(\mu_k | m_k^+, (\beta_k^+ \Lambda_k)^{-1}) p(\Lambda_k | W_k^+, \nu_k^+), \qquad k = 0, 1, \tag{38}$$

where

$$\beta_k^+ = \beta_k + N_k \tag{39}$$

$$m_k^+ = \frac{\beta_k m_k + N_k \bar{x}_k}{\beta_k^+} \tag{40}$$

$$(W_k^{-1})^+ = W_k^{-1} + N_k S_k + \frac{\beta_k N_k (\bar{x}_k - m_k)(\bar{x}_k - m_k)^T}{\beta_k + N_k} \tag{41}$$

$$\nu_k^+ = \nu_k + N_k \tag{42}$$

with weighted first order raw and second order central moments being, for $k = \{0, 1\}$

$$\bar{x}_k = \frac{1}{N_k} \sum_{i=1}^{N} \mathbb{E}[I_i] x_i \tag{43}$$

$$S_k = \frac{1}{N_k} \sum_{i=1}^{N} \mathbb{E}[I_i](x_i - \bar{x}_k)(x_i - \bar{x}_k)^T. \tag{44}$$

Returning to (32) and (33) with parameters obtained so far yields

$$\mathbb{E}[\log \pi] = \psi(\alpha_1^+) - \psi(\alpha_0^+ + \alpha_1^+) \tag{45}$$

$$\mathbb{E}[\log \det \Lambda_k] = \sum_{d=1}^{D} \psi\left(\frac{\nu_k^+ + 1 - d}{2}\right) + D \log 2 + \log \det W_k^+ \tag{46}$$

$$\mathbb{E}_{\mu_k, \Lambda_k}\left[(x_i - \mu_k)^T \Lambda_k (x_i - \mu_k)\right] = \frac{D}{\beta_k^+} + \nu_k^+ (x_i - m_k^+)^T W_k^+ (x_i - m_k^+) \tag{47}$$

where ψ denotes the digamma function.

4.3 Variational Message Passing

The Variational Message Passing (VMP) algorithm, rooted in the field of variational inference methods, was originally proposed in a paper Winn and Bishop [2005]. In many aspects it resembles the variational Bayes discussed above. Unlike it, VMP allows to infer parameters and latent variables of a large class of Bayesian networks without the need to derive application-specific update equations. Indeed, VMP alleviates the cumbersome derivations typical for the ordinary Variational Bayes, only at low price. It is the restriction of the method, imposing the conditional distributions at graph nodes to be exponential family distributions, summarized either by their natural parameter vector or by a vector of moments. This rule does not apply to the marginal distributions of observations, which can come from a broader class of distributions. These properties allow application of the framework on a number of popular algorithms, comprising hidden Markov models, principal component analysis, factor analysis, Kalman filters, probabilistic mixtures and hierarchical models.

A conditional distribution is a member of the exponential family if its pdf has the form

$$f(X|Y) = \exp\left[\phi(Y)' u(X) + f(X) + g(Y)\right], \tag{48}$$

where $\phi(Y)$ is the natural parameter, $u(X)$ is the natural sufficient statistics, $f(Y)$ is an arbitrary function and $g(Y)$ is the logarithm of the normalization constant (called log-partition function) ensuring unity of the area under $f(X|Y)$. One of numerous advantages of the exponential family distributions is the simplicity of evaluation of expectations. Under the knowledge of $\phi(Y)$, the pdf (48) can be rewritten as

$$f(X|\phi) = \exp\left[\phi' u(X) + f(Y + \tilde{g}(\phi))\right].$$

Since

$$\int_X \frac{d}{d\phi} f(X|\phi) dX = \int_X f(X|\phi) \left[u(X) + \frac{d\tilde{g}(\phi)}{d\phi}\right] dX = 0,$$

the expectation of the natural sufficient statistics is simply the differential of the log-partition

$$\mathbb{E}_{X|\phi}\left[u(X)\right] = -\frac{d\tilde{g}(\phi)}{d\phi}. \tag{49}$$

Suppose that we have a Bayesian network, represented by an ordered graph of a classical type $(N, E, \prec)$ where N denotes the set of nodes, E the set of edges and $\prec$ represents ordering. Then we distinguish among the following node types:

Parent nodes pa_i of a node $X_i \in N$ is the set of adjacent nodes that are joined to it and precede it in the ordering. That is, pa_i are parent nodes of X_i if $\forall X_j \in pa_i : (X_j, X_i) \in E$ and $X_j \prec X_i$.

Child nodes ch_i of a node $X_i \in N$ is the set of adjacent nodes that are joined to it and follow it in the ordering. That is, ch_i are child nodes of X_i if $\forall X_k \in ch_i : (X_i, X_k) \in E$ and $X_i \prec X_k$.

Co-parent nodes $cp_k^{(i)}$ of a node X_i with respect to the node $X_k \in ch_i$ is the set of non-adjacent nodes of X_i, that are parents of ch_i. That is, $cp_k^{(i)}$ are co-parent nodes of X_i if $\forall X \in cp_k^{(i)} : X \in pa_k, X_i \in pa_k, X \neq X_i$.

The structure of parents, child nodes and co-parents for a chosen node Z is schematically depicted in Figure 2.

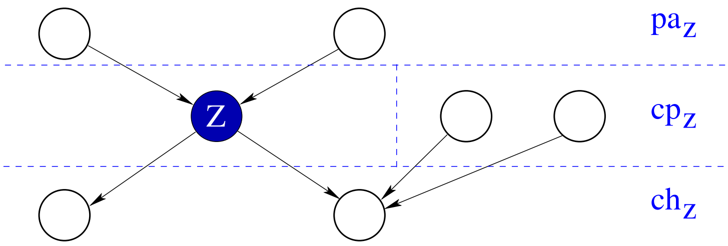

Figure 2: Structure of nodes around a node Z.

We assume that the joint distribution $f(\mathcal{X}) \equiv f(X, Z)$ can be factorized in the Bayesian network as

$$f(\mathcal{X}) = \prod_i f(\mathcal{X}_i|pa_i). \tag{50}$$

Identically (in respect to the notation and principle) to the variational Bayes, we have

$$\log f(X) = \mathcal{L}(q) + \mathcal{D}(q||f), \tag{51}$$

where

$$\mathcal{D}(q||f) = -\int q(Z)\log\frac{f(Z|X)}{q(Z)}dZ$$
$$\mathcal{L}(q) = \int q(Z)\log\frac{f(Z|X)}{q(Z)}dZ.$$

The computational simplicity is again provided by the factorization principle (19),

$$q(Z) = \prod_i q_i(Z_i).$$

After identical steps, this yields the same result as the Variational Bayes,

$$q_j^*(Z_j) \propto \exp\{\mathbb{E}_{i\neq j}[\log f(X, Z)]\}. \tag{52}$$

Until here, the derivations of the VMP are identical to VB. This is due to the adopted variational framework, dictating the same exclusive (zero-forcing) Kullback-Leibler divergence, accompanied by simplification in the form of factorized posterior distributions (19) and (50). From this point, the VB framework would derive the analytical relations among particular distributions in accordance with (20). The VMP framework, treated from now on, bypasses this by formulating a message passing procedure, avoiding these rather cumbersome derivations in favor of exchange of moments and parameters among the distributions.

Let us substitute (50) into (52)

$$q_j^*(Z_j) \propto \exp\left\{\mathbb{E}_{i\neq j} \sum_i f(X_i, Z_i|pa_i)\right\}. \tag{53}$$

Any terms independent of Z_j will be constant under the expectation and can be subsumed into the constant term. This leaves only $f(Z_j|pa_j)$ together with children ch_j, as they have $Z_j \in pa_j$. Hence

$$q_j^*(Z_j) \propto \exp E_{i\neq j}\left[f(Z_j|pa_j)\right] + \sum_{k\in ch_j} \mathbb{E}_{i\neq j}\left[f(X_k, Z_k|pa_k)\right]. \tag{54}$$

Verbally, the expectations required to evaluate $q_j^*(Z_j)$ involve only the variables lying in the Markov blanket of Z_j, i.e. its parents, children and co-parents.

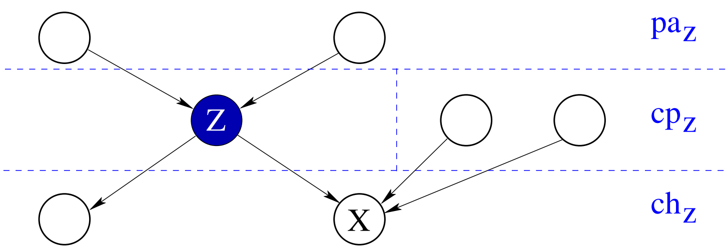

Figure 3: Structure of nodes around a node Z with an explicitly depicted child X.

Let us now describe in detail the VMP principle on Fig. 3. A node Z, representing for instance a parameter of a model $f(X|Z,\dots)$, is assigned an exponential family pdf

$$\log f(Z|pa_Z) = \phi_Z(pa_Z)' u_Z(Z) + f_Z(Z) + g_Z(pa_Z). \tag{55}$$

Similarly, a node X, obeys

$$\begin{aligned} \log f(X|Z, cp_Z) &= \phi_X(Z, cp_Z)' u_x(X) + f_X(X) + g_X(Z, cp_Z) \\ &= \phi_{XZ}(X, cp_Z)' u_Z(Z) + \lambda(X, cp_Z). \end{aligned} \tag{56}$$

The second representation of $\log f(X|Z, cp_Z)$ in (56) is obtained by a simple change of variables in the exponent, made in order to provide conjugacy. The principle dictates that for any random variable, say A, the term $\log f(A|pa_A)$ must be a multilinear in each of function of $u_A(A)$ and its parents. That is, it is linear in each of pa_A and $u_A(A)$, yet it may be affine in their combination.

The variational update for a node Z then reads

$$\begin{aligned}\log q_Z^*(Z) \propto \mathbb{E}_{\setminus Z}\left[\phi_Z(pa_Z)' u_Z(Z) + f_Z(Z) + g_Z(pa_Z)\right] \\ + \sum_{k \in ch_Z} \mathbb{E}_{\setminus Z}\left[\phi_{XZ}(X_k, cp_k)' u_Z(Z) + \lambda(X_k, cp_k)\right].\end{aligned}$$

By principle, it yields a function of the same type as the prior $f(Z|pa_Z)$, now with an updated natural parameter

$$\phi_Z^+ = \mathbb{E}\left[\phi_Z(pa_Z)\right] + \sum_{k \in ch_Z} \mathbb{E}\left[\phi_{XZ}(X_k, cp_k\right].$$

Since the expectations are evaluated on base of the expectations of the related sufficient statistics, we exploit (49). In other words, we can define

$$\begin{aligned}\tilde{\phi}_Z(\{\mathbb{E}\left[u_i\right]\}_{i \in pa_Z}) = \mathbb{E}[\phi_Z(pa_Z)] \\ \tilde{\phi}_{XZ}(\mathbb{E}\left[u_k\right], \{\mathbb{E}\left[u_j\right]\}_{j \in cp_k}) = \mathbb{E}[\phi_{XZ}(X_k, cp_k)].\end{aligned}$$

The VMP update is therefore made by exchanging and incorporating messages among adjacent nodes. These messages are of two types:

i) $Z \to X$: $m_{Z \to X} = \mathbb{E}[u_Z]$,

ii) $X \to Z$: $m_{X \to Z} = \tilde{\phi}_{XZ}(\mathbb{E}[u_x], \{m_{i \to X}\}_{i \in cp_Z})$.

In terms of messages, the update (57) reads

$$\phi_Z^+ = \tilde{\phi}_Z(\{m_{i \to Z}\}_{i \in pa_Z}) + \sum_{j \in ch_Z} m_{j \to Z}.$$

4.4 Application of VMP to estimation of adhesion loss

The application of the variational message passing algorithm to the adhesion loss estimation is entirely straightforward. Let us proceed with the same notation as in the variational Bayesian version. The situation is depicted in Fig. 4.4. Further on, the time indices are dropped. The variable I_i stands again for adhesion loss indicator as in (21) with a pdf (22) driven by a beta-distributed π, and the data obey a normal distribution with a pdf

$$p(X|I, \mu, \Lambda) = \prod_{i=1}^{N} p(x_i|\mu_0, \Lambda_0)^{1-I_i} p(x_i|\mu_1, \Lambda_1)^{I_i}. \tag{57}$$

Unlike in the VB, we consider $\mu_{i,t}$ and $\Lambda_{i,t}$ for $i \in \{0, 1\}$ independent, which leads to a higher degree of computational simplification for only a low price of inaccuracy. Their distributions are normal (for means $\mu_{i,t}$) and gamma (for precisions $\Lambda_{i,t}$) with pdfs $f(\mu|m, \beta)$ and $f(\Lambda|a, b)$, respectively.

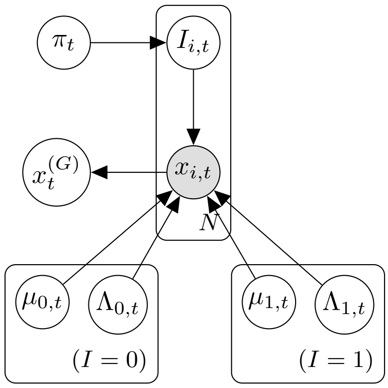

Figure 4: Variational message passing for estimation of adhesion loss. If we focus on the plate $(I = 0)$, then $\mu_{0,t}$ has child $x_{i,t}$, co-parents $\Lambda_{0,t}$, members of plate $(I = 1)$ and the indicator variable $I_{i,t}$. The system gives rise to the global variable $x_t^{(G)}$.

The logarithms of pdfs, corresponding to the model (56) and priors (55) are

$$\log f(x_i|\mu, \Lambda^{-1}, I) = (1 - \mathbb{E}[I_i]) \left\{ \begin{bmatrix} \Lambda_0 x_i \\ -\Lambda_0/2 \end{bmatrix}' \begin{bmatrix} \mu_0 \\ \mu_0^2 \end{bmatrix} + \frac{1}{2}(\log \Lambda_0 - \Lambda_0 x_i^2 - \log 2\pi) \right\}$$

$$+ \mathbb{E}[I_i] \left\{ \begin{bmatrix} \Lambda_1 x_i \\ -\Lambda_1/2 \end{bmatrix}' \begin{bmatrix} \mu_1 \\ \mu_1^2 \end{bmatrix} + \frac{1}{2}(\log \Lambda_1 - \Lambda_1 x_i^2 - \log 2\pi) \right\}$$

$$= (1 - \mathbb{E}[I_i]) \left\{ \begin{bmatrix} 1/2 \\ -(x_i - \mu_0)^2/2 \end{bmatrix}' \begin{bmatrix} \log \Lambda_0 \\ \Lambda_0 \end{bmatrix} - \log 2\pi \right\}$$

$$+ \mathbb{E}[I_i] \left\{ \begin{bmatrix} 1/2 \\ -(x_i - \mu_1)^2/2 \end{bmatrix}' \begin{bmatrix} \log \Lambda_1 \\ \Lambda_1 \end{bmatrix} - \log 2\pi \right\}$$

$$\log f(\mu|m_j, \beta_j) = \begin{bmatrix} \beta_j m_j \\ -\beta_j/2 \end{bmatrix}' \begin{bmatrix} \mu_j \\ \mu_j^2 \end{bmatrix} + \frac{1}{2}(\log \beta_j - \beta_j m_j^2 - \log 2\pi)$$

$$\log f(\lambda|a_j, b_j) = \begin{bmatrix} a_j - 1 \\ -b_j \end{bmatrix}' \begin{bmatrix} \log \lambda_j \\ \lambda_j \end{bmatrix} + a \log b - \log \Gamma(a), \qquad j \in \{0, 1\}.$$

The message passing algorithm is achieved by recursive run of the following steps, evaluated in arbitrary order:

(i) For $j \in \{0, 1\}$, messages $m_{\Lambda_j \to x_i}$ are sent from nodes Λ_j (co-parents of the respective μ_j) to observed nodes x_i. These messages are

$$m_{\Lambda_j \to x_i} = E[u_{\Lambda_j}(\Lambda_j)] = \begin{bmatrix} \mathbb{E}[\log \Lambda_j] \\ \mathbb{E}[\Lambda_j] \end{bmatrix}.$$

(ii) The nodes x_i use obtained messages $m_{\Lambda_j \to x_j}$ to form messages

$$m_{x_i \to \mu_j} = \begin{bmatrix} \mathbb{E}[\Lambda_j] x_i \\ -\mathbb{E}[\Lambda_j]/2 \end{bmatrix}.$$

These messages are scaled by $(1 - \mathbb{E}[I_i])$ and $\mathbb{E}[I_i]$ for $I = 0$ and $I = 1$, respectively. These values are obtained from the relevant plates $(I = 0)$ and $(I = 1)$.

(iii) The nodes $\mu_j, j \in \{0, 1\}$ update their natural parameters

$$\phi^{+}_{\mu_j} = \begin{bmatrix} \beta_j m_j \\ -\beta_j/2 \end{bmatrix} + \sum_{n=1}^{N} m_{x_i \to \mu_j}.$$

(iv) The nodes $\mu_j, j \in \{0, 1\}$ evaluate relevant expectations in order to form messages

$$m_{\mu_j \to x_i} = \begin{bmatrix} \mathbb{E}[\mu_j] \\ \mathbb{E}[\mu_j^2] \end{bmatrix}.$$

(v) The nodes x_i form messages for Λ_j such that

$$m_{x_i \to \Lambda_j} = \begin{bmatrix} 1/2 \\ -(x_i - \mathbb{E}[\mu_j])^2 \end{bmatrix},$$

and scale it as the messages for μ_j.

(vi) The nodes $\Lambda_j, j \in \{0, 1\}$ update their natural parameters according to

$$\phi^{+}_{\Lambda_j} = \begin{bmatrix} a_j - 1 \\ -b_j \end{bmatrix} + \sum_{n=1}^{N} m_{x_i \to \Lambda_j}.$$

5 Experiments

The adhesion loss estimation in three-PMSM drives setting is demonstrated using four different methods, in each case on base of the torque load. The torque load (here M_z) itself is estimated together with other state variables – currents $i_{s\alpha}, i_{s\beta}$, electric angle ϑ_e, angular speed ω_e with EKF (see Section 3.2). The 20 000 data points are artificial, generated by a PMSM simulator written in Matlab.

The torque load is preset as follows:

- $t < 8000 - M_z = 10$;
- $t \in [8000, 16000) - M_z = 20$;
- $t \geq 16000 - M_z = 30$;

The transitions are not sharp, but are smoothed by a twice-differentiable function. The centered normal cumulative distribution function with scale 30 is used for this purpose.

The first PMSMS drive's torque load is corrupted as follows:

- $t \in [500, 1500]$: the torque load drops to zero.
- $t \in [10000, 12000]$: the torque load drops to zero.
- $t \in [17000, 18000]$: a negative part of the sinusoid with amplitude 10 is added to the torque load.

The estimation of state and adhesion loss run in Python, both evaluated at each $t = 1, \cdots, 20000$.

The four methods chosen for demonstration are:

Variational Bayes – five iterations of VB are run after the state is obtained with EKF. The initial means are preset at 0 and $\max(M_{z,i,t})$ where $i \in \{1, 2, 3\}$ is an index of PMSMS drive and $t = 1, \cdots, 20000$. The initial $\nu = 3$ and $W = 100$ for both cases (regular/adhesion loss), $\pi = 0.5$ and $\alpha = \beta = 1$.

Variational message passing – again five iterations of the algorithm are run after the state is obtained with EKF. The initial means are set identically to VB, the precisions of normals are 0.01. The gamma distributions are preset noninformative with hyperparameters (0.1, 0.1) as proposed in Gelman et al. [2003].

k-means – for comparison, the data are clustered with two centroids represented by means. The scipy method `kmean2` is directly used. Its initialization is at randomly chosen points, only two iterations are used.

Median filtering – the simplest approach. After the state is obtained from EKF, the median of M_z over all actually estimated torque loads is selected and the other are compared to it.

Each method produces a 'global' estimate $M_z^{(G)}$. In the variational methods and k-means it corresponds to the mean, in median filtering it is directly the median. The results of EKF estimation (true value in red, estimated blue) together with differences $M_z - M_z^{(G)}$ are depicted in Figures 5 for VB, 6 for VMP, 7 for k-means and finally 8 for median filtering.

Further insight into results provides Table 1, summarizing statistics of differences between M_z estimates of each PMSM and the global estimate $M_z^{(G)}$. These statistics are minimum, all three quartiles, maximum, mean, standard deviation and inter-quartile range. The differences are also depicted by histograms in Figure 9.

Further improvements of variational methods can be made by reusing of flattened posterior distribution as the prior in the next time step (it is assumed that the dynamics is reasonably slower in reality). Also, instead of a fixed number of iterations, the variational lower bound could be used as a stopping criterion. Since its evaluation is computationally expensive, we have abandoned this idea.

Method, PMSM no	**Min**	$\mathbf{q_1}$	$\widetilde{\mathbf{x}}$	$\mathbf{q_3}$	**Max**	$\bar{\mathbf{x}}$	**s**	**IQR**
VB.PMSM1	-21.36	-0.04	-0.01	0.01	9.74	-2.61	6.21	0.05
VB.PMSM2	-3.97	-0.01	0.00	0.01	3.18	0.10	0.57	0.02
VB.PMSM3	-2.77	-0.01	0.00	0.01	3.19	0.12	0.55	0.02
VMP.PMSM1	-21.36	-0.04	-0.01	0.01	9.64	-2.67	6.25	0.05
VMP.PMSM2	-1.66	-0.01	0.00	0.01	1.33	0.04	0.27	0.02
VMP.PMSM3	-1.02	-0.01	0.00	0.01	1.33	0.06	0.24	0.02
k-means.PMSM1	-21.36	-0.06	0.00	0.00	9.64	-2.72	6.29	0.06
k-means.PMSM2	-0.94	0.00	0.00	0.00	0.23	-0.01	0.09	0.00
k-means.PMSM3	-0.40	0.00	0.00	0.00	0.94	0.01	0.08	0.01
median.PMSM1	-21.36	-0.06	0.00	0.00	8.76	-2.72	6.27	0.06
median.PMSM2	-1.87	0.00	0.00	0.00	0.19	-0.02	0.17	0.00
median.PMSM3	-0.08	0.00	0.00	0.00	0.11	0.00	0.02	0.00

Table 1: Statistics of $M_z - M_z^{(G)}$.

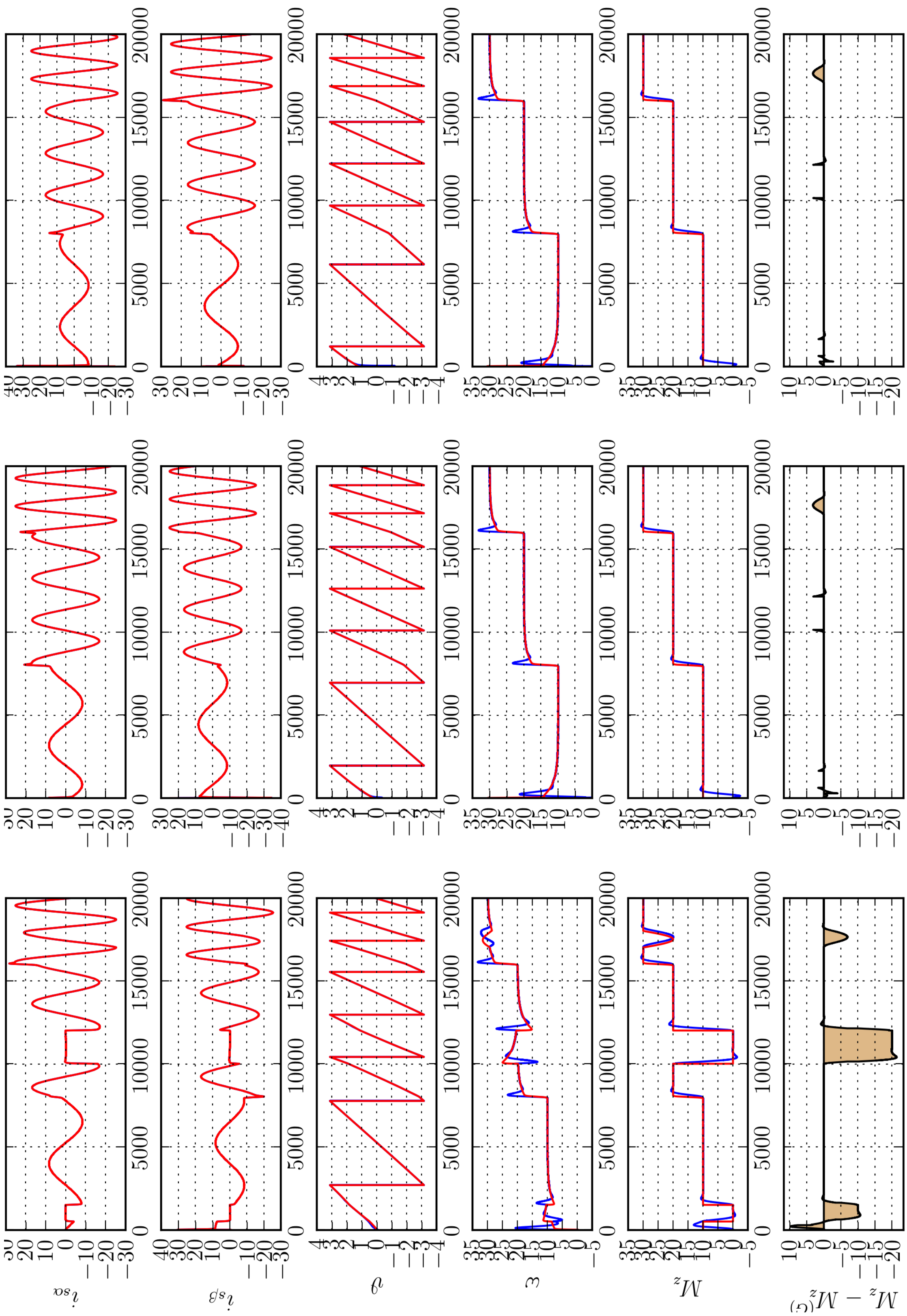

Figure 5: Estimation of PMSM with adhesion loss detection with variational Bayes.

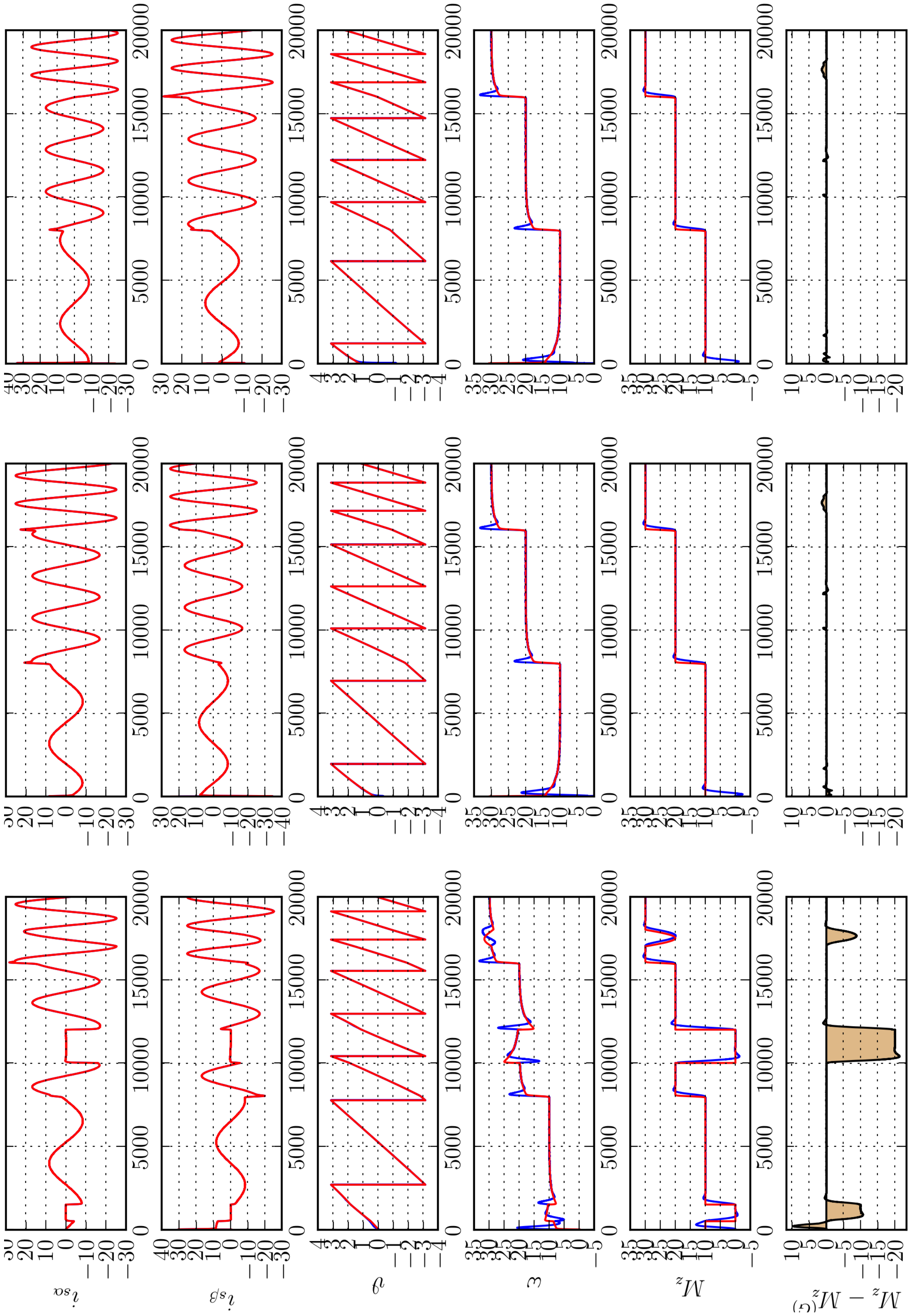

Figure 6: Estimation of PMSM with adhesion loss detection with variational message passing.

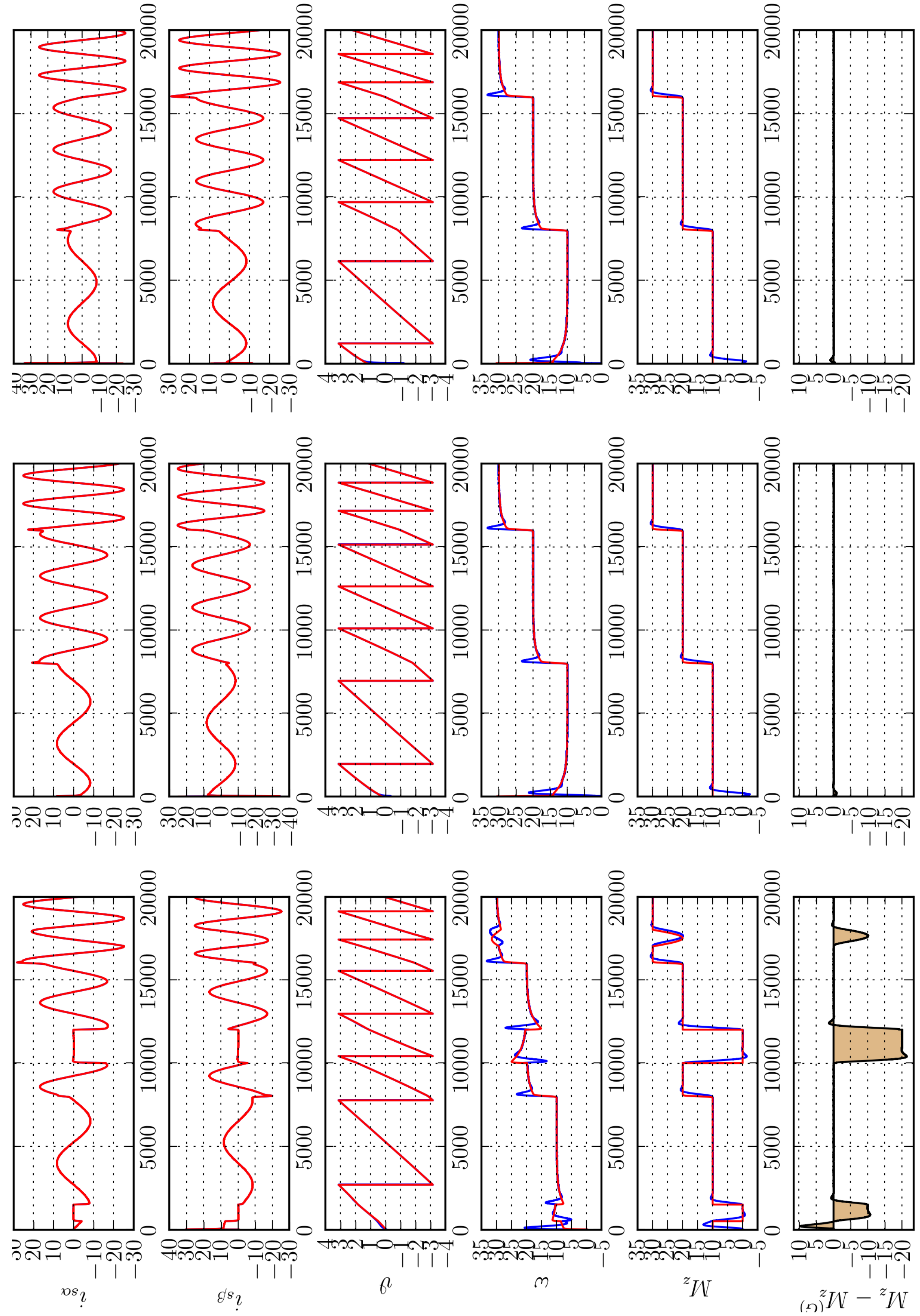

Figure 7: Estimation of PMSM with adhesion loss detection with k-means clustering.

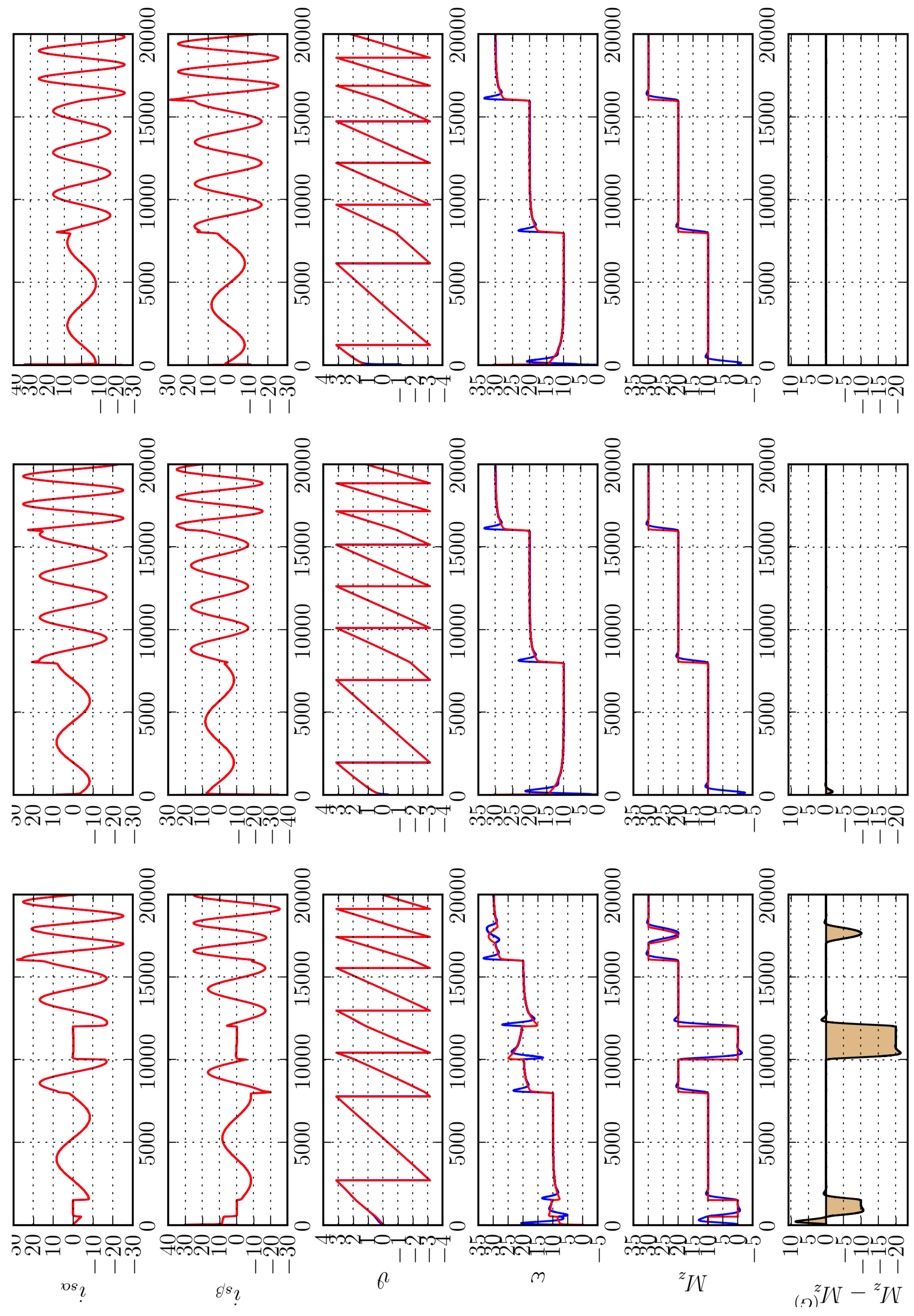

Figure 8: Estimation of PMSM with adhesion loss detection with median filter.

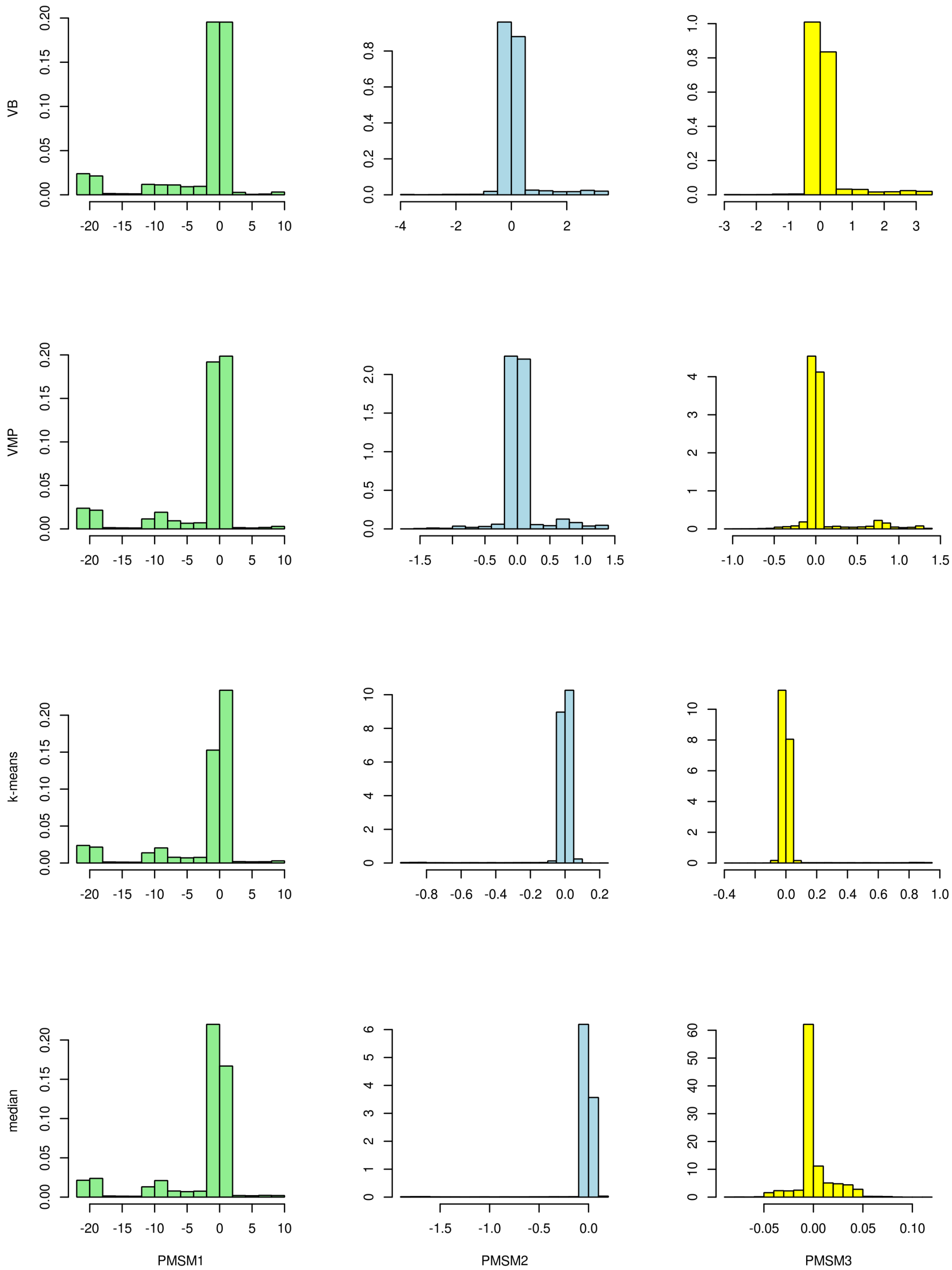

Figure 9: Histograms of $M_z - M_z^{(G)}$

6 Conclusion

The report provides insight into the adhesion loss estimation based on sensorless concurrent estimation of several PMSM drives using extended Kalman filter. Two variational methods – the variational Bayes method and the variational message passing – were derived for this purpose. They are compared to two ad-hoc simplistic approaches: the k-means clustering and median filtering.

The presented work is based on centralized data processing. It may be interesting to try a fully distributed algorithms, in which the variable of interest is estimated at each network node with the help of information obtained from other nodes. An example of a fully distributed variational Bayesian algorithm for estimation of normal mixtures in sensor networks is discussed in Safarinejadian et al. [2010]. While fully decentralized, it still assumes a potential risk in the form of a Hamiltonian cycle.

Acknowledgement

This work was supported by the Czech Science Foundation, project No. P102/11/0437.

References

Koan-Sok Baek, Keiji Kyogoku, and Tsunamitsu Nakahara. An experimental study of transient traction characteristics between rail and wheel under low slip and low speed conditions. *Wear*, 265(9):1417–1424, 2008.

TM Beagley and C Pritchard. Wheel/rail adhesion—the overriding influence of water. *Wear*, 35(2):299–313, 1975.

TM Beagley, IJ McEwen, and C Pritchard. Wheel/rail adhesion—the influence of railhead debris. *Wear*, 33(1):141–152, 1975a.

TM Beagley, IJ McEwen, and C Pritchard. Wheel/rail adhesion—boundary lubrication by oily fluids. *Wear*, 31(1):77–88, 1975b.

Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, March 2004. ISBN 0521833787. URL http://www.worldcat.org/isbn/0521833787.

A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1):1–38, 1977. ISSN 00359246. doi: 10.2307/2984875. URL http://dx.doi.org/10.2307/2984875.

A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin. *Bayesian Data Analysis, Second Edition (Chapman & Hall/CRC Texts in Statistical Science)*. Chapman and Hall/CRC, 2 edition, July 2003. ISBN 158488388X. URL http://www.worldcat.org/isbn/158488388X.

T. Jaakkola and M. Jordan. Bayesian parameter estimation via variational methods. *Statistics and Computing*, 10(1):25–37, January 2000. doi: 10.1023/A:1008932416310. URL http://dx.doi.org/10.1023/A:1008932416310.

S. Kullback and R. A. Leibler. On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86, 1951. ISSN 00034851. doi: 10.2307/2236703. URL http://dx.doi.org/10.2307/2236703.

Thomas P. Minka. Expectation propagation for approximate Bayesian inference. In *UAI '01: Proceedings of the 17th Conference in Uncertainty in Artificial Intelligence*, pages 362–369, San Francisco, CA, USA, 2001. Morgan Kaufmann Publishers Inc. ISBN 1-55860-800-1. URL http://portal.acm.org/citation.cfm?id=720257.

B Safarinejadian, MB Menhaj, and M Karrari. Distributed variational bayesian algorithms for gaussian mixtures in sensor networks. *Signal Processing*, 90(4):1197–1208, 2010.

Dan Simon. *Optimal state estimation: Kalman, H-infinity, and nonlinear approaches*. Wiley-Interscience, 2006.

S Soylu. Electric vehicles modelling and simulations, 2011.

V Šmídl. Statistical methods for sensorless control of PMSM drives, 2013. Habilitation thesis, University of West Bohemia in Pilsen, Faculty of Electrical Engineering.

John Winn and Christopher M. Bishop. Variational message passing. *Journal of Machine Learning Research*, 6:661–694, 2005. ISSN 1532-4435. URL http://portal.acm.org/citation.cfm?id=1046920.1088695.

Weihua Zhang, Jianzheng Chen, Xuejie Wu, and Xuesong Jin. Wheel/rail adhesion and analysis by using full scale roller rig. *Wear*, 253(1):82–88, 2002.