Bayesian system identification by message passing on factor graphs

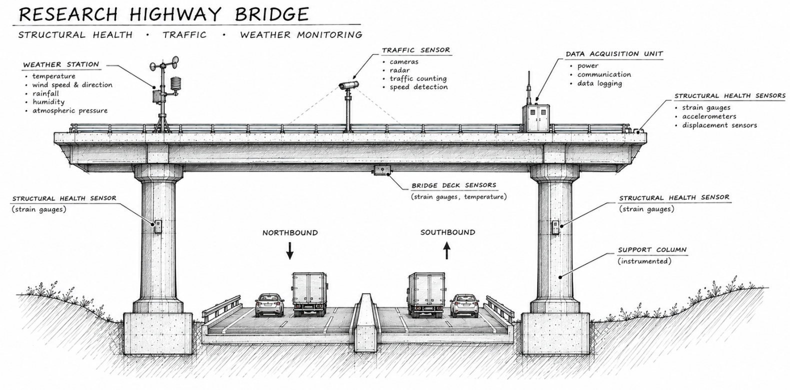
European Research Network on System Identification 2026

Wouter Kouw

Bayesian Intelligent Autonomous Systems lab | SPS | EE

TU/e EINDHOVEN UNIVERSITY OF TECHNOLOGY

System identification



https://wmkouw.github.io/ERNSI2026-talk/

 https://bwk.kuleuven.be/bwm/z24

System identification

As a driver, you may have many questions about this bridge:

System identification

As a driver, you may have many questions about this bridge:

- How old is this bridge?

System identification

As a driver, you may have many questions about this bridge:

- How old is this bridge?
- Can the bridge hold all this traffic?

System identification

As a driver, you may have many questions about this bridge:

- How old is this bridge?
- Can the bridge hold all this traffic?
- Have weather effects eroded it?

System identification

As a driver, you may have many questions about this bridge:

- How old is this bridge?
- Can the bridge hold all this traffic?
- Have weather effects eroded it?
- Is it safe for me to drive over it?

System identification

As a driver, you may have many questions about this bridge:

- How old is this bridge?
- Can the bridge hold all this traffic?
- Have weather effects eroded it?
- Is it safe for me to drive over it?

System identification

As a driver, you may have many questions about this bridge:

- How old is this bridge?
- Can the bridge hold all this traffic?
- Have weather effects eroded it?
- Is it safe for me to drive over it?

How can we answer such questions?

System identification

As a driver, you may have many questions about this bridge:

- How old is this bridge?
- Can the bridge hold all this traffic?
- Have weather effects eroded it?
- Is it safe for me to drive over it?

How can we answer such questions?

Claim: Uncertainty and information should propagate through a model and impose confidence on predictions and decisions.

Probabilistic model and Bayesian inference

Let $y_t \in \mathbb{R}$ be an observed output at time t, $\bar{y}_t$ a history, $u_t \in \mathbb{R}^{D_u}$ an observed input and $\theta = (\beta, \tau)$ the coefficients and precision.

Probabilistic model and Bayesian inference

Let $y_t \in \mathbb{R}$ be an observed output at time t, $\bar{y}_t$ a history, $u_t \in \mathbb{R}^{D_u}$ an observed input and $\theta = (\beta, \tau)$ the coefficients and precision.

A probabilistic autoregressive model with exogenous inputs is:

$$\underbrace{p(\boldsymbol{y}, \theta \mid \boldsymbol{u})}_{\text{joint distribution}} = \underbrace{p(\theta)}_{\text{prior}} \prod_{t=1}^{n} \underbrace{p(y_t \mid \bar{y}_t, u_t, \theta)}_{\text{likelihood}}$$

where $\boldsymbol{y} = (y_1, \ldots, y_n)$ and $\boldsymbol{u} = (u_1, \ldots, u_n)$.

Probabilistic model and Bayesian inference

Let $y_t \in \mathbb{R}$ be an observed output at time t, $\bar{y}_t$ a history, $u_t \in \mathbb{R}^{D_u}$ an observed input and $\theta = (\beta, \tau)$ the coefficients and precision.

A probabilistic autoregressive model with exogenous inputs is:

$$\underbrace{p(\boldsymbol{y}, \theta \mid \boldsymbol{u})}_{\text{joint distribution}} = \underbrace{p(\theta)}_{\text{prior}} \prod_{t=1}^{n} \underbrace{p(y_t \mid \bar{y}_t, u_t, \theta)}_{\text{likelihood}}$$

where $\boldsymbol{y} = (y_1, \ldots, y_n)$ and $\boldsymbol{u} = (u_1, \ldots, u_n)$.

Bayesian inference is about updating beliefs over latent variables:

$$\underbrace{p(\theta \mid \boldsymbol{y}, \boldsymbol{u})}_{\text{posterior}} = \frac{\overbrace{p(\boldsymbol{y} \mid \boldsymbol{u}, \theta)}^{\text{likelihood}}}{\underbrace{p(\boldsymbol{y} \mid \boldsymbol{u})}_{\text{model evidence}}} \underbrace{p(\theta)}_{\text{prior}}$$

Graphs

Probabilistic models can become challenging to manipulate manually.

Graphs

Probabilistic models can become challenging to manipulate manually.

Factor graphs visually represent the model's factorization structure.

Graphs

Probabilistic models can become challenging to manipulate manually.

Factor graphs visually represent the model's factorization structure.

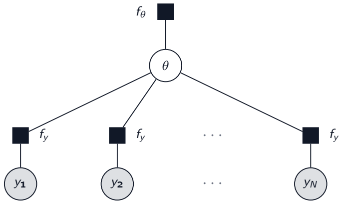
For example, $p(\boldsymbol{y}, \theta \mid \boldsymbol{u}) = \underbrace{p(\theta)}_{f_\theta} \prod_{t=1}^{n} \underbrace{p(y_t \mid \bar{y}_t, u_t, \theta)}_{f_y}$.

Graphs

Probabilistic models can become challenging to manipulate manually.

Factor graphs visually represent the model's factorization structure.

For example, $p(\boldsymbol{y}, \theta \mid \boldsymbol{u}) = \underbrace{p(\theta)}_{f_\theta} \prod_{t=1}^{n} \underbrace{p(y_t \mid \bar{y}_t, u_t, \theta)}_{f_y}$.



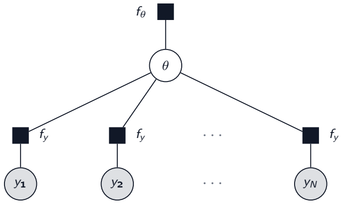
bipartite factor graph

Graphs

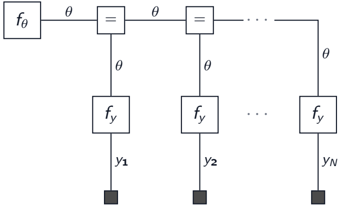
Probabilistic models can become challenging to manipulate manually.

Factor graphs visually represent the model's factorization structure.

For example, $p(\boldsymbol{y}, \theta \mid \boldsymbol{u}) = \underbrace{p(\theta)}_{f_\theta} \prod_{t=1}^{n} \underbrace{p(y_t \mid \bar{y}_t, u_t, \theta)}_{f_y}$.



bipartite factor graph



Forney-style factor graph

Message passing as distributed inference

Suppose we have a joint distribution over three variables:

$$p(z_1, z_2, z_3) = p(z_1 \mid z_2)p(z_2 \mid z_3)p(z_3)$$

Message passing as distributed inference

Suppose we have a joint distribution over three variables:

$$p(z_1, z_2, z_3) = p(z_1 \mid z_2) p(z_2 \mid z_3) p(z_3)$$

We are interested in the marginal distributions of the edges:

$$p(z_2) = \iint p(z_1, z_2, z_3) \mathrm{d}z_1 \mathrm{d}z_3$$

Message passing as distributed inference

Suppose we have a joint distribution over three variables:

$$p(z_1, z_2, z_3) = p(z_1 \mid z_2) p(z_2 \mid z_3) p(z_3)$$

We are interested in the marginal distributions of the edges:

$$p(z_2) = \iint p(z_1, z_2, z_3) \mathrm{d}z_1 \mathrm{d}z_3$$

The factorization structure can be leveraged to distribute integrals:

$$p(z_2) = \underbrace{\int p(z_1 \mid z_2) \mathrm{d}z_1}_{\mu_{\mathbf{1} \to \mathbf{2}}(z_2)} \underbrace{\int p(z_2 \mid z_3) p(z_3) \mathrm{d}z_3}_{\mu_{\mathbf{3} \to \mathbf{2}}(z_2)}$$

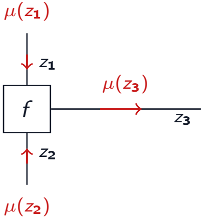
Message passing as distributed inference

Factor nodes integrate information into an outgoing message:

 Loeliger et al., 2007 (IEEE SPM).

Message passing as distributed inference

Factor nodes integrate information into an outgoing message:



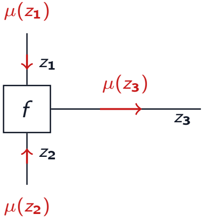
Sum-product message

$$\mu(z_i) = \int \prod_{j \neq i} \mu(z_j) f(z) \mathrm{d}z_{\setminus i}$$

Loeliger et al., 2007 (IEEE SPM).

Message passing as distributed inference

Factor nodes integrate information into an outgoing message:



Sum-product message

$$\mu(z_i) = \int \prod_{j \neq i} \mu(z_j) f(z) \mathrm{d}z_{\setminus i}$$

For example, a Gaussian factor node with a Gaussian message:

$$\mu(z_2) = \int \mathcal{N}(z_2 \mid z_1, \sigma_2^2) \mathcal{N}(z_1 \mid \mu_1, \sigma_1^2) \mathrm{d}z_1 = \mathcal{N}(z_2 \mid \mu_1, \sigma_2^2 + \sigma_1^2)\,.$$

Loeliger et al., 2007 (IEEE SPM).

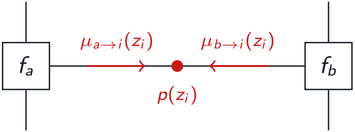
Message passing as distributed inference

Colliding messages update marginal distributions:

Loeliger et al., 2007 (IEEE SPM).

Message passing as distributed inference

Colliding messages update marginal distributions:



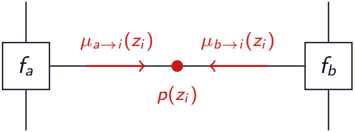
Marginal update

$$p(z_i) = \mu_{a\to i}(z_i)\ \mu_{b\to i}(z_i)$$

Loeliger et al., 2007 (IEEE SPM).

Message passing as distributed inference

Colliding messages update marginal distributions:



Marginal update

$$p(z_i) = \mu_{a\to i}(z_i)\ \mu_{b\to i}(z_i)$$

For example, two colliding Gaussian distributions yields:

$$p(z_i) = \mathcal{N}(z_i \mid \mu_a, \sigma_a^2) \cdot \mathcal{N}(z_i \mid \mu_b, \sigma_b^2) = \mathcal{N}\left(z_i \mid \lambda^{-1}\xi, \lambda^{-1}\right)$$

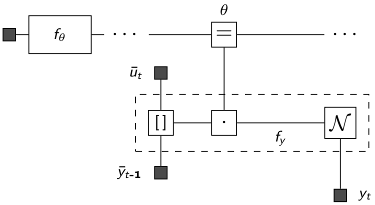
where $\lambda = \sigma_a^{-2} + \sigma_b^{-2}$ and $\xi = \sigma_a^{-2}\mu_a + \sigma_b^{-2}\mu_b$.

Loeliger et al., 2007 (IEEE SPM).

TU/e

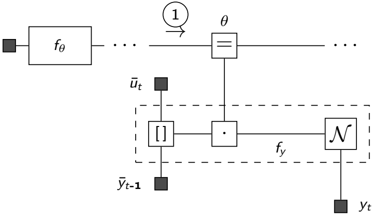
Message passing in AR model

Writing out the autoregressive model in more detail:



Message passing in AR model

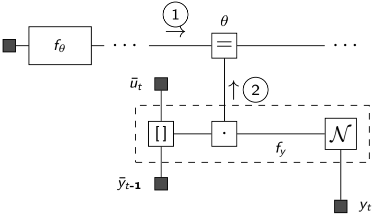
Writing out the autoregressive model in more detail:



- Message from prior distribution.

Message passing in AR model

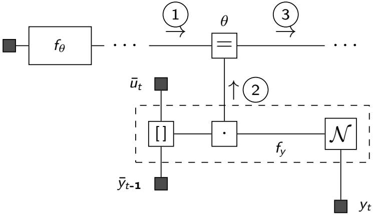
Writing out the autoregressive model in more detail:



- Message from prior distribution.
- Backwards message using data.

Message passing in AR model

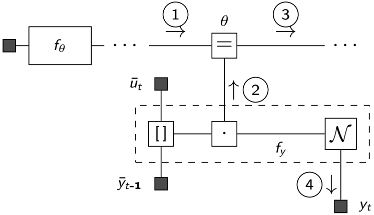
Writing out the autoregressive model in more detail:



- Message from prior distribution.
- Backwards message using data.
- Posterior becomes future prior.

Message passing in AR model

Writing out the autoregressive model in more detail:



- Message from prior distribution.
- Backwards message using data.
- Posterior becomes future prior.
- Posterior predictive distribution.

Generalizing Bayesian inference

Bayesian inference is a special case of free energy minimization.

Generalizing Bayesian inference

Bayesian inference is a special case of free energy minimization.

The free energy functional of $p(\boldsymbol{y}, \theta \mid \boldsymbol{u})$ with variational model $q(\theta)$ is:

$$\mathcal{F}[q] = \int q(\theta) \log \frac{q(\theta)}{p(\boldsymbol{y}, \theta \mid \boldsymbol{u})} \mathrm{d}\theta \, .$$

Generalizing Bayesian inference

Bayesian inference is a special case of free energy minimization.

The free energy functional of $p(\boldsymbol{y}, \theta \mid \boldsymbol{u})$ with variational model $q(\theta)$ is:

$$\mathcal{F}[q] = \int q(\theta) \log \frac{q(\theta)}{p(\boldsymbol{y}, \theta \mid \boldsymbol{u})} \mathrm{d}\theta \,.$$

Decomposing reveals relation to Bayesian inference:

$$\mathcal{F}[q] = \overbrace{\int q(\theta) \log \frac{q(\theta)}{p(\theta \mid \boldsymbol{y}, \boldsymbol{u})} \mathrm{d}\theta}^{\text{approximation}} \overbrace{- \log p(\boldsymbol{y} \mid \boldsymbol{u})}^{\text{negative log-evidence}}$$

Generalizing Bayesian inference

Bayesian inference is a special case of free energy minimization.

The free energy functional of $p(\boldsymbol{y}, \theta \mid \boldsymbol{u})$ with variational model $q(\theta)$ is:

$$\mathcal{F}[q] = \int q(\theta) \log \frac{q(\theta)}{p(\boldsymbol{y}, \theta \mid \boldsymbol{u})} \mathrm{d}\theta \,.$$

Decomposing reveals relation to Bayesian inference:

$$\begin{aligned}
\mathcal{F}[q] &= \overbrace{\int q(\theta) \log \frac{q(\theta)}{p(\theta \mid \boldsymbol{y}, \boldsymbol{u})} \mathrm{d}\theta}^{\text{approximation}} \; \overbrace{- \log p(\boldsymbol{y} \mid \boldsymbol{u})}^{\text{negative log-evidence}} \\
&= \underbrace{\int q(\theta) [- \log p(\boldsymbol{y} \mid \boldsymbol{u}, \theta)] \mathrm{d}\theta}_{\text{prediction error}} + \underbrace{\int q(\theta) \log \frac{q(\theta)}{p(\theta)} \mathrm{d}\theta}_{\text{change in beliefs}}
\end{aligned}$$

Free energy on graphs

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ define a graph over a set of nodes $\mathcal{V}$ and edges $\mathcal{E}$.

 Yedidia, Freeman & Weiss, 2005 (IEEE TIT); Şenöz et al., 2021 (Entropy).

Free energy on graphs

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ define a graph over a set of nodes $\mathcal{V}$ and edges $\mathcal{E}$.

The variational model $q(z)$ is indexed by nodes and edges:

$$q(z) = \frac{\prod_{a \in \mathcal{V}} q_a(z_a)}{\prod_{i \in \mathcal{E}} q_i(z_i)}$$

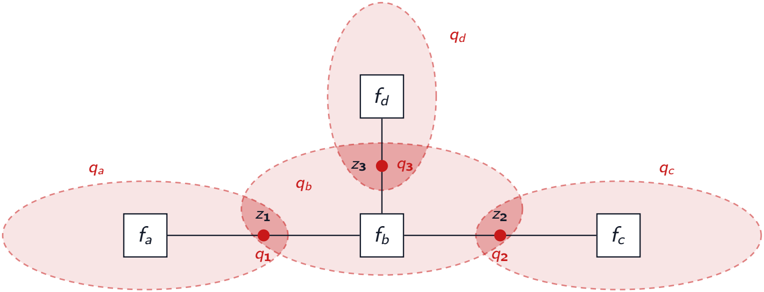
 Yedidia, Freeman & Weiss, 2005 (IEEE TIT); Şenöz et al., 2021 (Entropy).

Free energy on graphs

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ define a graph over a set of nodes $\mathcal{V}$ and edges $\mathcal{E}$.

The variational model $q(z)$ is indexed by nodes and edges:

$$q(z) = \frac{\prod_{a \in \mathcal{V}} q_a(z_a)}{\prod_{i \in \mathcal{E}} q_i(z_i)}$$



 Yedidia, Freeman & Weiss, 2005 (IEEE TIT); Şenöz et al., 2021 (Entropy).

Free energy on graphs

The Bethe free energy is a functional over node energies and edge entropies:

 Yedidia, Freeman & Weiss, 2005 (IEEE TIT); Şenöz et al., 2021 (Entropy).

Free energy on graphs

The Bethe free energy is a functional over node energies and edge entropies:

$$\mathcal{B}[q] = \sum_{a \in \mathcal{V}} \int q_a(z_a) \log \frac{q_a(z_a)}{f_a(z_a)} \mathrm{d}z_a + \sum_{i \in \mathcal{E}} - \int q_i(z_i) \log q_i(z_i) \mathrm{d}z_i \,,$$

Yedidia, Freeman & Weiss, 2005 (IEEE TIT); Şenöz et al., 2021 (Entropy).

Free energy on graphs

The Bethe free energy is a functional over node energies and edge entropies:

$$\mathcal{B}[q] = \sum_{a \in \mathcal{V}} \int q_a(z_a) \log \frac{q_a(z_a)}{f_a(z_a)} \mathrm{d}z_a + \sum_{i \in \mathcal{E}} - \int q_i(z_i) \log q_i(z_i) \mathrm{d}z_i \,,$$

subject to marginalization and normalization constraints

$$\int q_a(z_a)\, \mathrm{d}z_{a \setminus i} = q_i(z_i) \qquad \forall a, i \in \Delta a$$

$$\int q_i(z_i)\, \mathrm{d}z_i = 1 \qquad \forall i$$

 Yedidia, Freeman & Weiss, 2005 (IEEE TIT); Şenöz et al., 2021 (Entropy).

Free energy on graphs

The Bethe free energy is a functional over node energies and edge entropies:

$$\mathcal{B}[q] = \sum_{a \in \mathcal{V}} \int q_a(z_a) \log \frac{q_a(z_a)}{f_a(z_a)} \mathrm{d}z_a + \sum_{i \in \mathcal{E}} - \int q_i(z_i) \log q_i(z_i) \mathrm{d}z_i \,,$$

subject to marginalization and normalization constraints

$$\begin{aligned} \int q_a(z_a)\, \mathrm{d}z_{a \setminus i} &= q_i(z_i) && \forall a, i \in \Delta a \\ \int q_i(z_i)\, \mathrm{d}z_i &= 1 && \forall i \end{aligned}$$

Note that constraints are not the same as priors!

 Yedidia, Freeman & Weiss, 2005 (IEEE TIT); Şenöz et al., 2021 (Entropy).

Minimizing the Bethe free energy functional

We form a Lagrangian out of the Bethe free energy and its constraints:

Zhang et al., 2021 (IEEE TWC); Şenöz et al., 2021 (Entropy).

Minimizing the Bethe free energy functional

We form a Lagrangian out of the Bethe free energy and its constraints:

$$\mathcal{L}[q] = \mathcal{B}[q] + \sum_{a \in \mathcal{V}} \int \lambda_{ai}(z_i)\Big(\int q_a(z_a)\mathrm{d}z_{a\setminus i} - q_i(z_i)\Big)\mathrm{d}z_i + \sum_{i \in \mathcal{E}} \lambda_i\Big(\int q_i(z_i)\mathrm{d}z_i - 1\Big).$$

 Zhang et al., 2021 (IEEE TWC); Şenöz et al., 2021 (Entropy).

Minimizing the Bethe free energy functional

We form a Lagrangian out of the Bethe free energy and its constraints:

$$\mathcal{L}[q] = \mathcal{B}[q] + \sum_{a \in \mathcal{V}} \int \lambda_{ai}(z_i) \Big(\int q_a(z_a) \mathrm{d}z_{a \setminus i} - q_i(z_i) \Big) \mathrm{d}z_i + \sum_{i \in \mathcal{E}} \lambda_i \Big(\int q_i(z_i) \mathrm{d}z_i - 1 \Big).$$

The optimal variational factors are its stationary points:

$$q_a(z) \propto f_a(z_a) \prod_{j \in \Delta a} \mu_{j \to a}(z_j) \qquad q_i(z_i) \propto \prod_{a \in \Delta i} \mu_{a \to i}(z_i),$$

Zhang et al., 2021 (IEEE TWC); Şenöz et al., 2021 (Entropy).

Minimizing the Bethe free energy functional

We form a Lagrangian out of the Bethe free energy and its constraints:

$$\mathcal{L}[q] = \mathcal{B}[q] + \sum_{a \in \mathcal{V}} \int \lambda_{ai}(z_i)\Big(\int q_a(z_a)\mathrm{d}z_{a \setminus i} - q_i(z_i)\Big)\mathrm{d}z_i + \sum_{i \in \mathcal{E}} \lambda_i\Big(\int q_i(z_i)\mathrm{d}z_i - 1\Big).$$

The optimal variational factors are its stationary points:

$$q_a(z) \propto f_a(z_a) \prod_{j \in \Delta a} \mu_{j \to a}(z_j) \qquad q_i(z_i) \propto \prod_{a \in \Delta i} \mu_{a \to i}(z_i),$$

where the messages arise as mechanism to enforce constraints

$$\mu_{a \to i}(z_i) = \int f_a(z_a) \prod_{j \in \Delta a \setminus i} \mu_{j \to a}(z_j)\, \mathrm{d}z_{a \setminus i}\,.$$

 Zhang et al., 2021 (IEEE TWC); Şenöz et al., 2021 (Entropy).

Variational message passing

When the expectation over messages cannot be solved or if we wish to impose conjugacy, then we can impose additional *local* constraints.

Yedidia, Freeman & Weiss, 2005 (IEEE TIT); Dauwels, 2007 (IEEE ISIT).

Variational message passing

When the expectation over messages cannot be solved or if we wish to impose conjugacy, then we can impose additional *local* constraints.

Variational message rule

For node a, if $q_a(z_a) = \prod_j q_j(z_j)$, the message is derived to be:

$$\mu_{a \to i}(z_i) \propto \exp\left(\mathbb{E}_{\prod_{j \neq i} q_j(z_j)}\big[\log f_a(z_a)\big]\right)$$

 Yedidia, Freeman & Weiss, 2005 (IEEE TIT); Dauwels, 2007 (IEEE ISIT).

Variational message passing

When the expectation over messages cannot be solved or if we wish to impose conjugacy, then we can impose additional *local* constraints.

Variational message rule

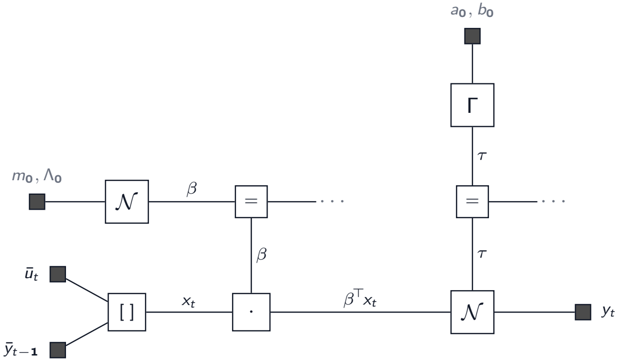
For node a, if $q_a(z_a) = \prod_j q_j(z_j)$, the message is derived to be:

$$\mu_{a \to i}(z_i) \propto \exp\left(\mathbb{E}_{\prod_{j \neq i} q_j(z_j)}\left[\log f_a(z_a)\right]\right)$$

Alternative constraints exist, such as moment matching.

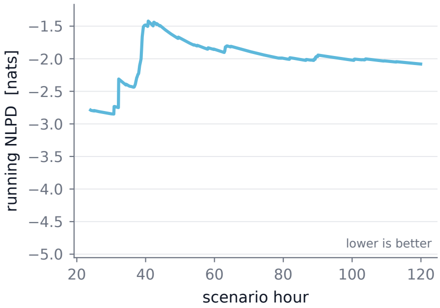
Yedidia, Freeman & Weiss, 2005 (IEEE TIT); Dauwels, 2007 (IEEE ISIT).

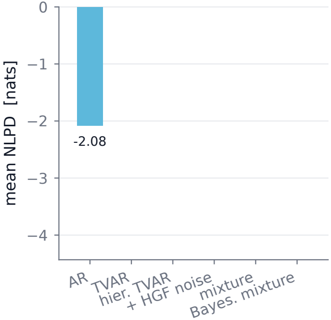
Demo 1: autoregressive model



 Nisslbeck & Kouw, 2025 (IEEE ECC).

Demo 1: does it forecast?





Composability

Factor graphs are modular computational graphs.

Loeliger, 2004 (IEEE SPM); Bagaev & de Vries, 2023 (Sci. Prog.).

Composability

Factor graphs are modular computational graphs.

Message passing rules can be stored and re-used.

- Deterministic operations, e.g., $+$, $-$, $\cdot$.

 Loeliger, 2004 (IEEE SPM); Bagaev & de Vries, 2023 (Sci. Prog.).

Composability

Factor graphs are modular computational graphs.

Message passing rules can be stored and re-used.

- Deterministic operations, e.g., $+$, $-$, $\cdot$.
- Exponential family-based conditional distributions, e.g., $\mathcal{N}$, $\mathcal{G}$.

 Loeliger, 2004 (IEEE SPM); Bagaev & de Vries, 2023 (Sci. Prog.).

Composability

Factor graphs are modular computational graphs.

Message passing rules can be stored and re-used.

- Deterministic operations, e.g., $+$, $-$, $\cdot$.
- Exponential family-based conditional distributions, e.g., $\mathcal{N}$, $\mathcal{G}$.
- Special compounds, e.g., mixtures, switches, non-parametric.

Loeliger, 2004 (IEEE SPM); Bagaev & de Vries, 2023 (Sci. Prog.).

Composability

Factor graphs are modular computational graphs.

Message passing rules can be stored and re-used.

- Deterministic operations, e.g., $+$, $-$, $\cdot$.
- Exponential family-based conditional distributions, e.g., $\mathcal{N}$, $\mathcal{G}$.
- Special compounds, e.g., mixtures, switches, non-parametric.

 Loeliger, 2004 (IEEE SPM); Bagaev & de Vries, 2023 (Sci. Prog.).

Composability

Factor graphs are modular computational graphs.

Message passing rules can be stored and re-used.

- Deterministic operations, e.g., $+$, $-$, $\cdot$.
- Exponential family-based conditional distributions, e.g., $\mathcal{N}$, $\mathcal{G}$.
- Special compounds, e.g., mixtures, switches, non-parametric.

This creates a design framework for a class of probabilistic models where inference is tractable, containing among others

- Kalman filters

Loeliger, 2004 (IEEE SPM); Bagaev & de Vries, 2023 (Sci. Prog.).

Composability

Factor graphs are modular computational graphs.

Message passing rules can be stored and re-used.

- Deterministic operations, e.g., $+$, $-$, $\cdot$.
- Exponential family-based conditional distributions, e.g., $\mathcal{N}$, $\mathcal{G}$.
- Special compounds, e.g., mixtures, switches, non-parametric.

This creates a design framework for a class of probabilistic models where inference is tractable, containing among others

- Kalman filters
- Hidden Markov models

 Loeliger, 2004 (IEEE SPM); Bagaev & de Vries, 2023 (Sci. Prog.).

Composability

Factor graphs are modular computational graphs.

Message passing rules can be stored and re-used.

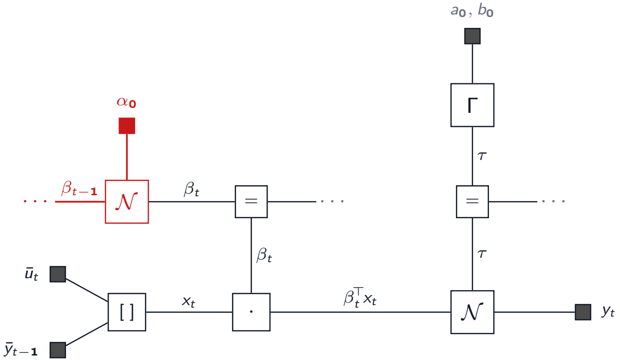
- Deterministic operations, e.g., $+$, $-$, $\cdot$.
- Exponential family-based conditional distributions, e.g., $\mathcal{N}$, $\mathcal{G}$.
- Special compounds, e.g., mixtures, switches, non-parametric.

This creates a design framework for a class of probabilistic models where inference is tractable, containing among others

- Kalman filters
- Hidden Markov models
- Gaussian process regression

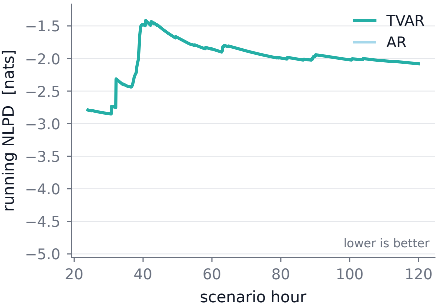
Loeliger, 2004 (IEEE SPM); Bagaev & de Vries, 2023 (Sci. Prog.).

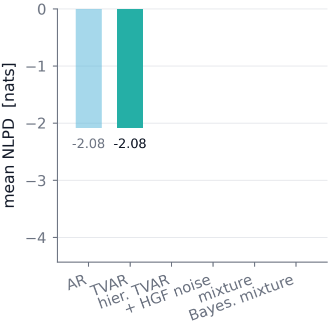
Demo 2: time-varying autoregressive model



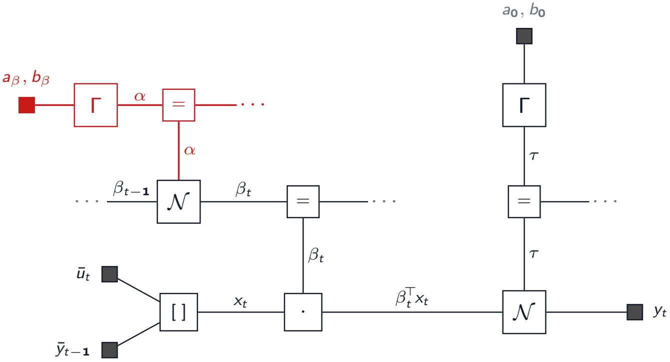
 Podusenko, Kouw & de Vries, 2020 (IEEE ISIT).

Demo 2: does it forecast?



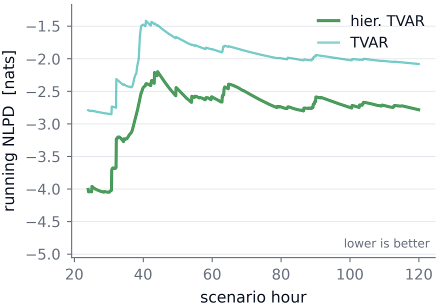


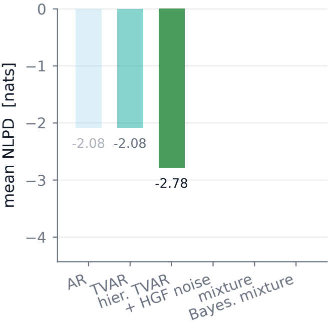
Demo 3: hierarchical autoregressive model



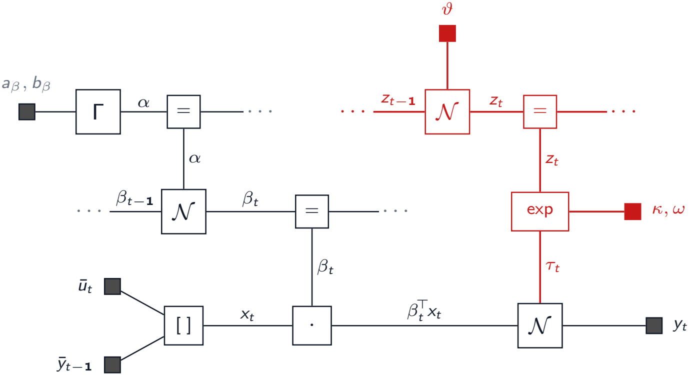
Podusenko, Kouw & de Vries, 2023 (Entropy).

Demo 3: does it forecast?

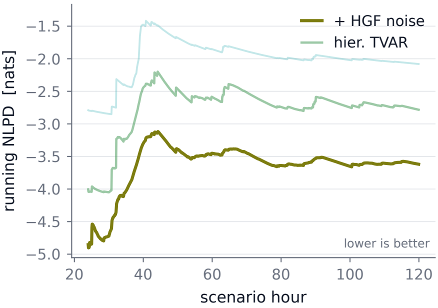


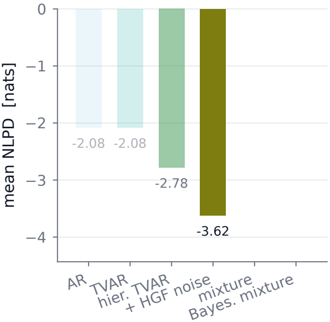


Demo 4: volatile HAR



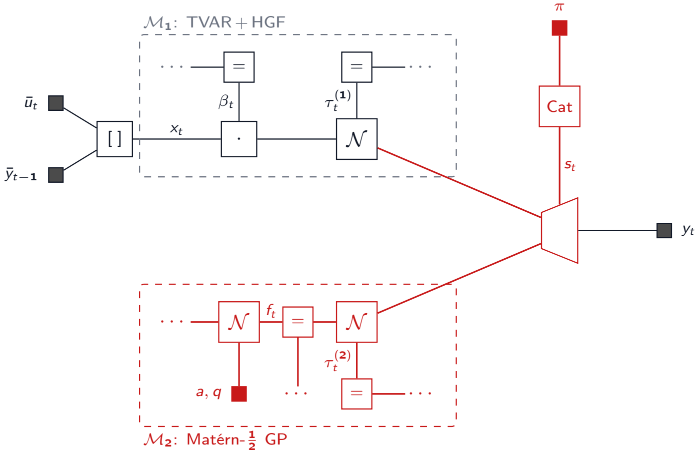
Demo 4: does it forecast?





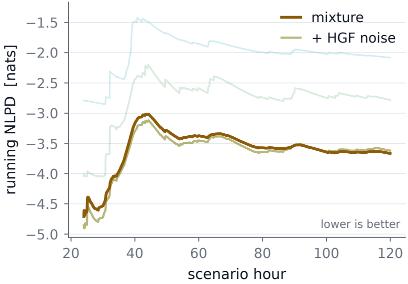
TU/e

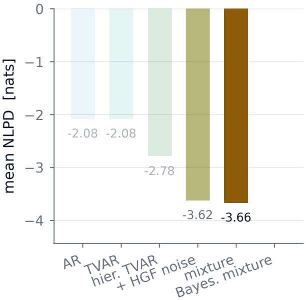
Demo 5: mixture of experts



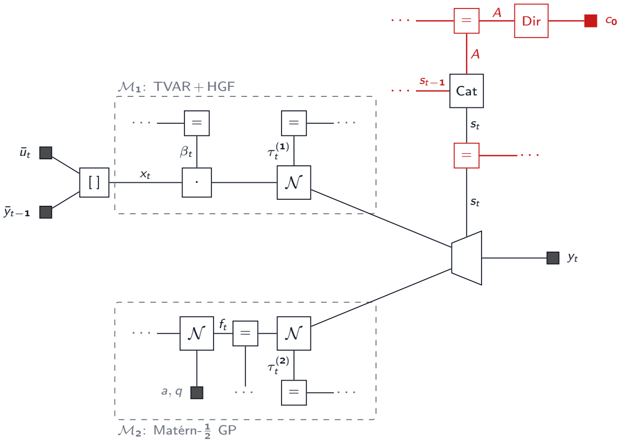
 Minka & Winn, 2008 (NeurIPS).

Demo 5: does it forecast?



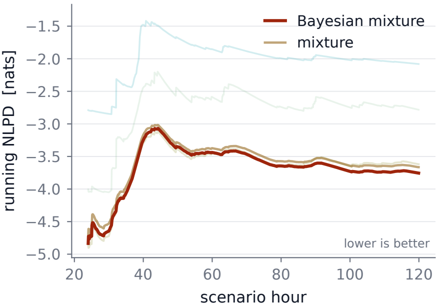


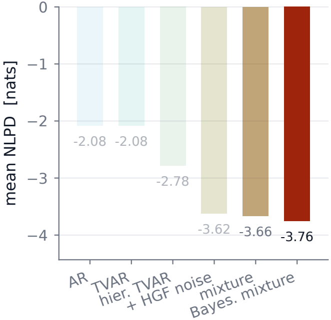
Demo 6: Bayesian mixture of experts



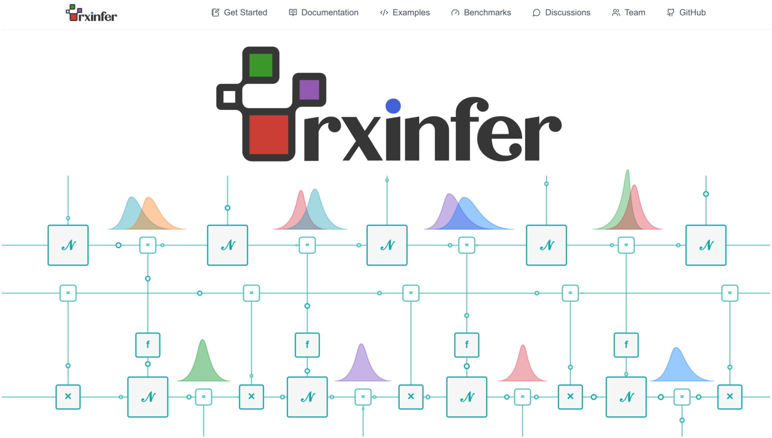
 Lukashchuk et al., 2026 (ProbNum); Van Erp, Nuijten & de Vries, 2024 (IWAI).

Demo 6: does it forecast?





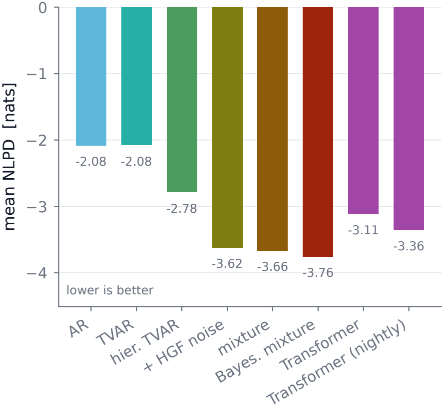
Composable means automatable

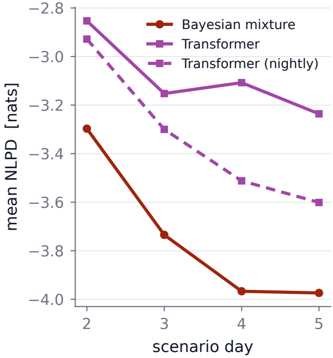


Bagaev & de Vries 2023 (Scientific Programming); Kouw 2026 (IWAI).

Comparing to a modern baseline

Can you make models that compete with modern deep networks?





 Vaswani et al., 2017 (NeurIPS).

Is it safe to drive over?

Suppose safe passage requires $y < 150\,\mathrm{mm/s^2}$ (r.m.s.).

Is it safe to drive over?

Suppose safe passage requires $y < 150\,\mathrm{mm/s^2}$ (r.m.s.).

- Leave Friday at 16:00 ($\pm$1500 vehicles per hour)

Is it safe to drive over?

Suppose safe passage requires $y < 150\,\text{mm/s}^2$ (r.m.s.).

- Leave Friday at 16:00 ($\pm$1500 vehicles per hour)
- A $6\,\text{m/s}$ breeze.

Is it safe to drive over?

Suppose safe passage requires $y < 150\,\mathrm{mm/s^2}$ (r.m.s.).

- Leave Friday at 16:00 ($\pm$1500 vehicles per hour)
- A 6 m/s breeze.
- The deck is at 7°C.

Is it safe to drive over?

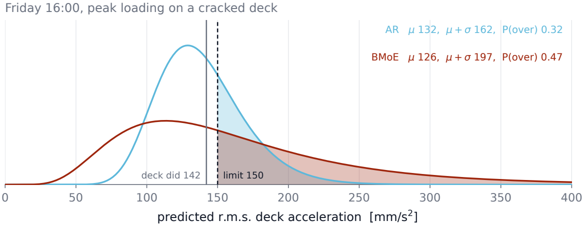
Suppose safe passage requires $y < 150\,\mathrm{mm/s^2}$ (r.m.s.).

- Leave Friday at 16:00 ($\pm$1500 vehicles per hour)
- A $6\,\mathrm{m/s}$ breeze.
- The deck is at 7°C.

Is it safe to drive over?

Suppose safe passage requires $y < 150\,\mathrm{mm/s^2}$ (r.m.s.).

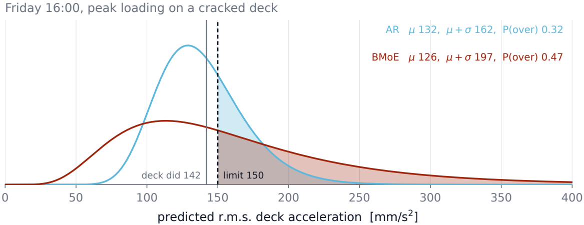
- Leave Friday at 16:00 ($\pm$1500 vehicles per hour)
- A $6\,\mathrm{m/s}$ breeze.
- The deck is at 7°C.



Is it safe to drive over?

Suppose safe passage requires $y < 150\,\mathrm{mm/s^2}$ (r.m.s.).

- Leave Friday at 16:00 ($\pm$1500 vehicles per hour)
- A 6 m/s breeze.
- The deck is at 7°C.



Mean stays below threshold but mean+std.dev. is above.

Is it safe to drive over?

Now, look at a later time point.

Is it safe to drive over?

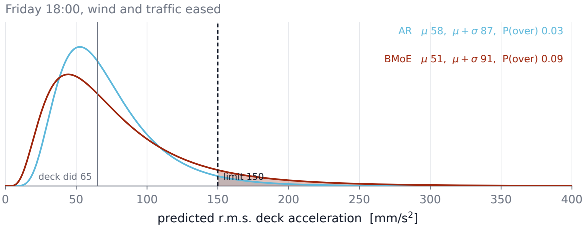
Now, look at a later time point.

- Leave Friday at 18:00 ($\pm$600 vehicles per hour)
- A 3 m/s breeze.
- The deck is at 12°C

Is it safe to drive over?

Now, look at a later time point.

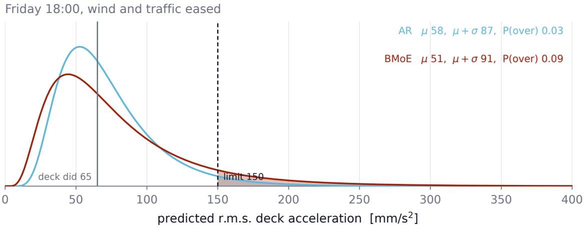
- Leave Friday at 18:00 ($\pm$600 vehicles per hour)
- A 3 m/s breeze.
- The deck is at 12°C



Is it safe to drive over?

Now, look at a later time point.

- Leave Friday at 18:00 ($\pm$600 vehicles per hour)
- A 3 m/s breeze.
- The deck is at 12°C



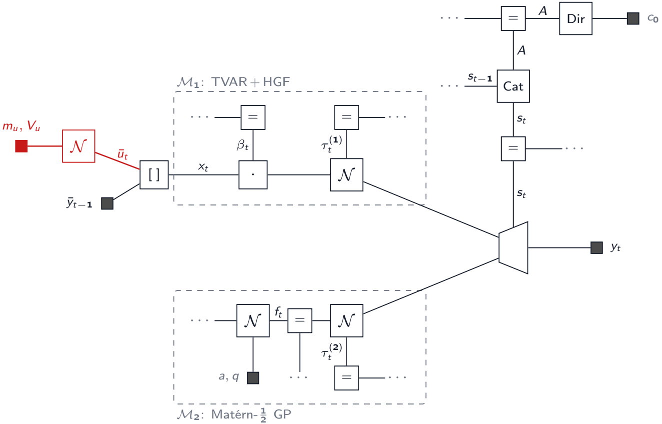
Mean + std. dev is below the threshold.

Is it safe to drive over?

What if we had more uncertainty over inputs?

Is it safe to drive over?

What if we had more uncertainty over inputs?



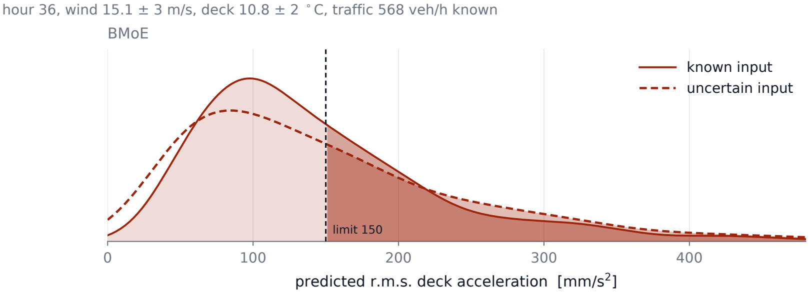
Is it safe to drive over?

What if we had more uncertainty over inputs?

Is it safe to drive over?

What if we had more uncertainty over inputs?

That uncertainty propagates through the model:



Predictive variance is increased, demanding more caution.

Discussion

- Convergence depends on graph structure.

Khan & Rue 2023 (JMLR); Courts, Wills, Schön & Ninness 2023 (Automatica); Murai et al. 2024 (IEEE RA-L); MacKinlay et al. 2025 (ICLR).

Discussion

- Convergence depends on graph structure.
  - Sufficient conditions (e.g., walk-summability) exist.

Khan & Rue 2023 (JMLR); Courts, Wills, Schön & Ninness 2023 (Automatica); Murai et al. 2024 (IEEE RA-L); MacKinlay et al. 2025 (ICLR).

Discussion

- Convergence depends on graph structure.
  - Sufficient conditions (e.g., walk-summability) exist.
- Parameter updates are deterministic.

Khan & Rue 2023 (JMLR); Courts, Wills, Schön & Ninness 2023 (Automatica); Murai et al. 2024 (IEEE RA-L); MacKinlay et al. 2025 (ICLR).

Discussion

- Convergence depends on graph structure.
  - Sufficient conditions (e.g., walk-summability) exist.
- Parameter updates are deterministic.
  - Consistency can be analyzed, but updates are coupled.

Khan & Rue 2023 (JMLR); Courts, Wills, Schön & Ninness 2023 (Automatica); Murai et al. 2024 (IEEE RA-L); MacKinlay et al. 2025 (ICLR).

Discussion

- Convergence depends on graph structure.
  - Sufficient conditions (e.g., walk-summability) exist.
- Parameter updates are deterministic.
  - Consistency can be analyzed, but updates are coupled.
- (Strong) calibration for exact Bayesian inference.

Khan & Rue 2023 (JMLR); Courts, Wills, Schön & Ninness 2023 (Automatica); Murai et al. 2024 (IEEE RA-L); MacKinlay et al. 2025 (ICLR).

Discussion

- Convergence depends on graph structure.
  - Sufficient conditions (e.g., walk-summability) exist.
- Parameter updates are deterministic.
  - Consistency can be analyzed, but updates are coupled.
- (Strong) calibration for exact Bayesian inference.
  - Open question for variational message passing.

Khan & Rue 2023 (JMLR); Courts, Wills, Schön & Ninness 2023 (Automatica); Murai et al. 2024 (IEEE RA-L); MacKinlay et al. 2025 (ICLR).

Discussion

- Convergence depends on graph structure.
  - Sufficient conditions (e.g., walk-summability) exist.
- Parameter updates are deterministic.
  - Consistency can be analyzed, but updates are coupled.
- (Strong) calibration for exact Bayesian inference.
  - Open question for variational message passing.
- Other applications:

Khan & Rue 2023 (JMLR); Courts, Wills, Schön & Ninness 2023 (Automatica); Murai et al. 2024 (IEEE RA-L); MacKinlay et al. 2025 (ICLR).

Discussion

- Convergence depends on graph structure.
  - Sufficient conditions (e.g., walk-summability) exist.
- Parameter updates are deterministic.
  - Consistency can be analyzed, but updates are coupled.
- (Strong) calibration for exact Bayesian inference.
  - Open question for variational message passing.
- Other applications:
  - Active sensing.

Khan & Rue 2023 (JMLR); Courts, Wills, Schön & Ninness 2023 (Automatica); Murai et al. 2024 (IEEE RA-L); MacKinlay et al. 2025 (ICLR).

Discussion

- Convergence depends on graph structure.
    - Sufficient conditions (e.g., walk-summability) exist.
- Parameter updates are deterministic.
    - Consistency can be analyzed, but updates are coupled.
- (Strong) calibration for exact Bayesian inference.
    - Open question for variational message passing.
- Other applications:
    - Active sensing.
    - Simultaneous learning and control.

Khan & Rue 2023 (JMLR); Courts, Wills, Schön & Ninness 2023 (Automatica); Murai et al. 2024 (IEEE RA-L); MacKinlay et al. 2025 (ICLR).

Discussion

- Convergence depends on graph structure.
  - Sufficient conditions (e.g., walk-summability) exist.
- Parameter updates are deterministic.
  - Consistency can be analyzed, but updates are coupled.
- (Strong) calibration for exact Bayesian inference.
  - Open question for variational message passing.
- Other applications:
  - Active sensing.
  - Simultaneous learning and control.
  - Probabilistic numerics.

Khan & Rue 2023 (JMLR); Courts, Wills, Schön & Ninness 2023 (Automatica); Murai et al. 2024 (IEEE RA-L); MacKinlay et al. 2025 (ICLR).

With thanks to BIASlab, and collaborators.


Bert de Vries


Thijs van de Laar


Albert Podusenko


İsmail Şenöz


Dmitry Bagaev


Bart van Erp


Tim Nisslbeck


Mykola Lukashchuk

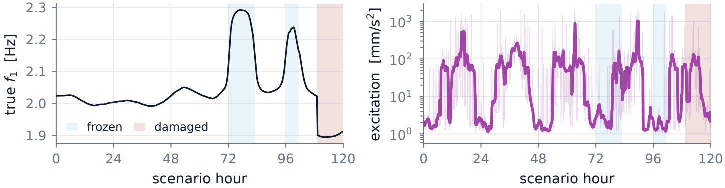

Wouter Nuijten


Ajith Anil Meera

Funded by

The simulated bridge

A 30 m single span, two bending modes at 2.0 and 6.0 Hz with damping 1.5% and 2%, one accelerometer at $0.3L$ sampled at 20 Hz.



- Two modes give two conjugate pole pairs, $\prod_n (1 - 2r_n \cos\phi_n z^{-1} + r_n^2 z^{-2})$ with $r_n = \exp(-\zeta_n \omega_n / f_s)$, so the deck *is* an AR(4).
- The poles move for three reasons. Temperature, about -0.3% per °C. Freezing, which lifts f_1 by 10%. Damage on Friday at 13:00, which drops it by 6%.
- The amplitude moves for two, traffic from 50 to 1500 vehicles an hour and a storm on Tuesday.

Loopy belief propagation

Fixed points of belief propagation are exactly the stationary points of $\mathcal{B}$, and the stable ones are its local minima.

- **On a tree.** $\mathcal{B}$ is exact, its minimum is $-\log p(\boldsymbol{y})$, and two sweeps reach it. No iteration.
- **With loops.** $\mathcal{B}$ is non-convex and is neither an upper nor a lower bound. Several fixed points may exist and the iteration can oscillate. It is convex with at most one cycle; beyond that, contraction conditions, damping, or a double-loop scheme that descends by construction and pays per iteration for it.
- **In the Gaussian case**, if it converges the means are exact, the variances generally too small.

Yedidia, Freeman & Weiss 2005 (IEEE TIT); Heskes,2006 (JAIR); Yuille 2002 (Neural Comp.); Mooij & Kappen 2007 (IEEE TIT); Weiss & Freeman 2001 (Neural Comp.).

TU/e