Probabilistic Fault Detection in Aircraft Engines Using Integrated DBN and MRF Models

Anand Patel 2561034
School Of Computer Science and Applied Mathematics
University of the Witwatersrand

***Abstract*—This project addresses early-stage fault detection in aircraft engines by using probabilistic graphical models. We propose an combined framework combining a Dynamic Bayesian Network (DBN) with a Markov Random Field (MRF) for sensor smoothing. The system operates on discretized sensor data simulated across various engine health scenarios. Our model captures component degradation and sensor reliability without requiring HMMs. Evaluated against baseline models, our method demonstrates improved robustness, fault distinction, and understandability.**

I. Introduction

Aircraft engines operate in dynamic, safety-critical environments where early detection of mechanical faults can reduce maintenance costs and prevent failure. Traditional threshold-based diagnostic systems often lack the flexibility to incorporate uncertainty, temporal dependencies, and sensor unreliability—factors that are present in real-world operation [1], [2].

This project focuses on the early detection of engine faults using probabilistic graphical models (PGMs), specifically a combination of a Dynamic Bayesian Network (DBN) and a Markov Random Field (MRF). The model aims to infer the health of core engine systems using noisy, discretized data from four sensors: EGT (Exhaust Gas Temperature), N2 (high-pressure spool speed), Oil Pressure, and Vibration (Vib1 and Vib2).

PGMs are fitting for this problem because they allow reasoning under uncertainty and work with structured, understandable inference. The DBN models the temporal evolution of component health, while the MRF ensures consistency in vibration sensor readings [3].

The key contributions of this work are:

- Development of an integrated DBN + MRF framework for aircraft engine fault detection.
- A probabilistic pipeline incorporating MRF-based smoothing, DBN inference, and threshold-based classification.
- Complete evaluation over fault scenarios, showing improved robustness compared to baseline models.

II. Related Work

This project focuses on the early detection of engine faults using probabilistic graphical models (PGMs), specifically we use a combination of a Dynamic Bayesian Network (DBN) and a Markov Random Field (MRF). The model aims to infer the health of core engine subsystems using noisy, discretized data from four main sensors: EGT (Exhaust Gas Temperature), N2 (high-pressure spool speed), Oil Pressure, and Vibration (Vib1 and Vib2).

Dynamic Bayesian Networks (DBNs) have been generally applied in health monitoring and fault diagnosis for their ability to model temporal dependencies and handle incomplete data. Amin et al. [4] demonstrate the success of DBNs in modeling complex systems where component states evolve over time and sensor readings are noisy. Their work on subsea system monitoring shows how DBNs can provide interpretable, probabilistic reasoning in mission-critical domains.

Markov Random Fields (MRFs) have been used for spatial smoothing and noise reduction in sensor tasks. In the context of fault diagnosis, MRFs are particularly effective when multiple redundant or parallel sensors (e.g., vibration sensors) output correlated behavior. By making sure we have local consistency, MRFs reduce the likelihood of false alarms due to short-lived sensor spikes [5].

While Hidden Markov Models (HMMs) and Kalman Filters are well known tools for time-series modeling, they lack the flexibility and multivariate inference capability of DBNs. Kalman Filters assume linear-Gaussian processes and are best suited for continuous data, while HMMs can model discrete state transitions but struggle with complex dependency structures across variables [6].

Our approach builds upon these foundations by integrating DBNs for temporal modeling and MRFs for spatial consistency. This hybrid structure allows for robust fault detection under uncertainty.

III. Methodology

A. Sensor Simulation and Discretization

We simulate 1800-step time-series data to reflect realistic operating conditions of an aircraft engine. The dataset spans five scenarios: Normal, Oil Leak, Bearing Wear, EGT Sensor Failure, and Vibration Sensor Failure. Each scenario generates timestamped readings for four main sensors: Exhaust Gas Temperature (EGT), N2 rotational speed, Oil Pressure, and two Vibration channels (Vib1 and Vib2).

The raw sensor data is converted into discrete categorical values representing three operational states: **Low**, **Medium**, and **High**. These thresholds are manually chosen to reflect typical ranges observed during cruise, climb, and descent operations. Discretization allows for simplified modeling in the DBN and more interpretable probability distributions.

To introduce realism, each fault scenario includes controlled noise and potential dropouts. For instance, sensor failure scenarios simulate corrupted or constant values while retaining plausible dynamics in unaffected sensors. This design enables the model to learn fault characteristics while still being robust to noise and incomplete data.

From an ethical standpoint, all simulations are artificial and do not rely on proprietary or sensitive operational data. This ensures full reproducibility and avoids misuse of sensitive aircraft information while still testing meaningful diagnostic behaviors.

The four sensors work as follows. ***EGT*** (Exhaust Gas Temperature) measures how hot the engine exhaust is — this can increase if the engine is working harder than usual or if there's internal friction from worn parts. ***N2*** tracks how fast the core part of the engine is spinning. It's a sign of how much power the engine is producing. ***Oil Pressure*** tells us whether the lubrication system is working properly — if the oil leaks or the pump fails, the pressure drops quickly. ***Vib1*** and ***Vib2*** measure vibration in different parts of the engine. High vibration usually means that rotating parts are wearing out or becoming unbalanced. By combining all these sensor readings, the system can figure out whether a fault is mechanical (like a worn bearing) or due to a failing sensor.

B. Markov Random Field (MRF)

To combat local inconsistencies between the two correlated vibration sensors (*Vib1* and *Vib2*), we implement a simple Markov Random Field (MRF) smoothing step prior to DBN inference.

The reasoning is that *Vib1* and *Vib2* monitor the same mechanical event (rotor vibration) and should therefore produce similar readings. However, due to sensor noise, electrical interference, or momentary disturbances, one sensor may spike or drop independently of the other. These uncorrelated events can mislead the DBN into diagnosing a false fault or ignoring a real one.

The MRF smoothing algorithm operates by applying pairwise consistency between the two sensor readings. For each timestep t, if the discretized values of *Vib1_raw* and *Vib2_raw* do not match, the smoother applies a correction based on a fixed penalty function. The updated value for each sensor is a weighted average influenced by the current value and its neighbor. This operation is repeated for a fixed number of iterations to converge toward a similar output without over-smoothing.

Once the smoothing is complete, the adjusted vibration signals are discretized into Low/Medium/High bins and passed to the DBN as observations. This preprocessing step has shown to reduce false high-frequency noise and improve the reliability of fault detection, particularly for the *Bearing Wear* scenario.

C. Dynamic Bayesian Network

The main component of our fault detection system is a Dynamic Bayesian Network (DBN), which models temporal dependencies between hidden subsystem states and observed sensor readings. Each timestep includes hidden nodes representing the health of the *Engine Core*, *Lubrication System*, and the *Sensor Health* of both EGT and Vibration channels. Observable nodes are the discretized values of EGT temperature, N2 RPM, Oil Pressure, Vib1, and Vib2.

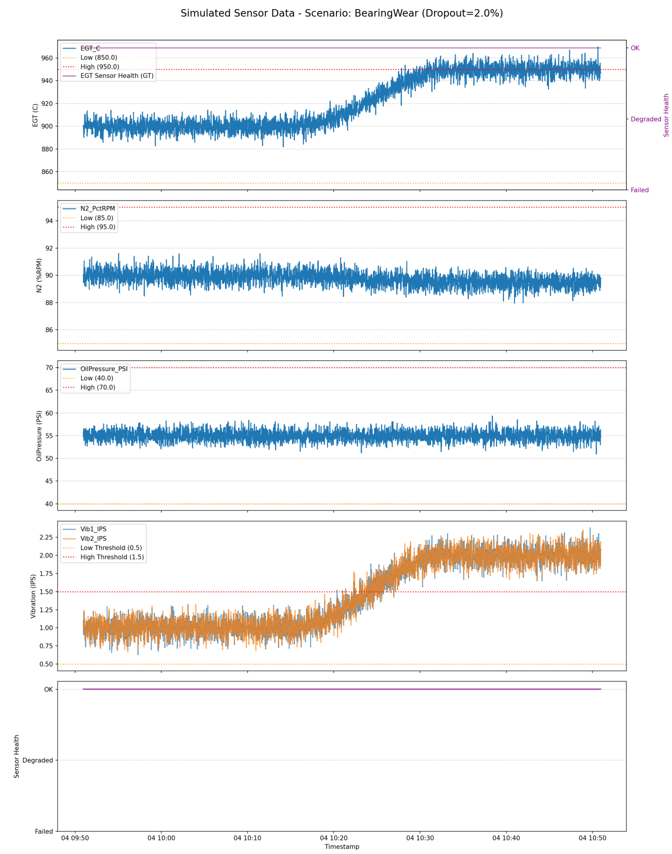

Fig. 1. Simulated raw sensor readings for the *Bearing Wear* scenario. Notice the gradual rise in vibration levels and slight EGT increase, aligned with the known degradation pattern.

The DBN models both intra-slice dependencies (e.g., observation nodes depending on component and sensor health) and inter-slice transitions (e.g., health states evolving from t to $t+1$). For example, *EGT_C_Discrete* depends on both *Engine Core Health* and *EGT Sensor Health*, while *Oil Pressure* depends solely on *Lubrication System Health*.

Conditional probabilities were based on expert knowledge of how engine components typically behave. These mirror real-world expectations, such as: low oil pressure is highly indicative of lubrication failure, or that sensor degradation increases the likelihood of corrupted observations.

The model is built using `pgmpy`'s `DBNInference` engine, using exact forward inference (filtering) over time. At each step, the model receives discrete sensor evidence and outputs marginal probabilities over hidden health states.

D. Prediction Mapping and Baselines

The DBN outputs marginal probabilities for each hidden state at every timestep. To generate understandable fault predictions, we apply a rule-based mapping that translates these probabilities into discrete fault labels. For example:

- If $\mathbb{P}(\text{Lubrication System Health} = \text{Fail}) > 0.6$, predict *Oil Leak*.
- If $\mathbb{P}(\text{Engine Core Health} \in \{\text{Warn}, \text{Fail}\}) > 0.6$, predict *Bearing Wear*.
- If $\mathbb{P}(\text{EGT Sensor Health} = \text{Failed}) > 0.6$, predict *EGT Sensor Failure*.
- If $\mathbb{P}(\text{Vibration Sensor Health} = \text{Failed}) > 0.6$, predict *Vibration Sensor Failure*.
- Otherwise, classify as *Normal*.

These thresholds were chosen to level early detection while taking into false positives and were validated across multiple simulation runs.

To understand the performance of our full system, we compare against two baselines:

- **Vanilla DBN:** A simplified model with no sensor health nodes and no MRF smoothing. This tests the impact of our integrated sensor degradation modeling.
- **Rule-Based Classifier:** A manual logic system that directly classifies faults based on hard thresholds on raw sensor values (e.g., five consecutive low oil pressure readings). This was the non-probabilistic method baseline.

IV. Experiments and Evaluation

A. Experimental Setup

To evaluate our system's fault detection capabilities, we simulate time-series data covering five key scenarios: *Normal Operation*, *Oil Leak*, *Bearing Wear*, *EGT Sensor Failure*, and *Vibration Sensor Failure*. Each scenario consists of 1800 time steps of raw sensor readings generated via controlled stochastic processes.

We compare three models:

- **Full DBN + MRF:** Our system with vibration smoothing, sensor health modeling, and probabilistic temporal inference.
- **Vanilla DBN:** A simplified version of the DBN with no sensor health nodes and no MRF smoothing.
- **Rule-Based Classifier:** A deterministic baseline that flags faults based on threshold violations in raw data (e.g., 5 consecutive low oil pressure readings).

Each model is evaluated on the same simulation dataset. At each time step, the models generate a predicted label, which is then compared against the ground truth scenario label (for component faults) and the injected degradation label (for sensor faults, where applicable).

We measure the following classification metrics:

- **Accuracy:** Proportion of correct predictions over total predictions.
- **Precision:** True positives divided by total predicted positives, averaged across classes.

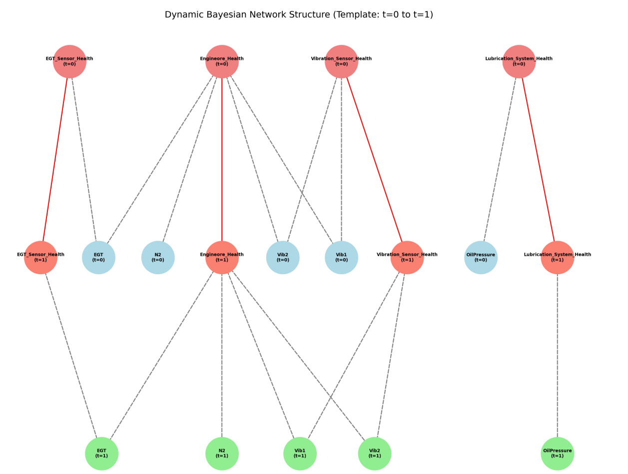

Fig. 2. Structure of the Dynamic Bayesian Network (DBN) used for engine fault diagnosis. Each node represents a variable at either time $t = 0$ or $t = 1$. **Red nodes** are hidden health states: core components (e.g., *EngineCore_Health*, *Lubrication_System_Health*) and sensor reliability (*EGT_Sensor_Health*, *Vibration_Sensor_Health*). **Blue nodes** represent discretized sensor observations at $t = 0$, while **green nodes** are observations at $t = 1$. Solid arrows denote temporal transitions, while dashed arrows denote dependencies within the same time slice.

- **Recall:** True positives divided by actual positives, averaged across classes.
- **F1-Score:** Harmonic mean of precision and recall.
- **Confusion Matrices:** Visual summaries of prediction distributions for qualitative comparison.

All metrics are computed using `scikit-learn`, and the analysis is repeated for both engine-level fault detection and sensor health classification (for models capable of such inference). Output metrics and reports are logged for each model automatically via `evaluation.py`.

B. Results

Table I shows the overall classification performance of the three evaluated models on engine-level fault detection. The Full DBN with MRF achieves high recall and precision for the key faults—*Oil Leak* and *Bearing Wear*—and outperforms the Vanilla DBN by a big margin. While the Rule-Based system achieves slightly higher overall accuracy, its fixed threshold logic offers limited flexibility and interpretability, and it lacks robustness in uncertain or noisy conditions.

TABLE I
Overall Weighted Performance Metrics for Engine Fault Detection

Model	Accuracy	Precision	Recall	F1 Score
Full DBN + MRF	0.851	0.756	0.851	0.796
Rule-Based	0.916	0.918	0.916	0.912
Vanilla DBN	0.178	0.775	0.178	0.135

The Full DBN is especially effective in capturing fault patterns over time. In the *Oil Leak* and *Bearing Wear* scenarios, it maintains high F1-scores due to its ability to integrate temporal dependencies and sensor reliability. In contrast, the Vanilla DBN fails to detect most faults and overfits to dominant classes, showing the need of both MRF smoothing and explicit sensor modeling.

The sensor health classification results further demonstrate the Full DBN's robustness to noisy input. Despite class imbalance, the model correctly identifies sensor failure states (F1 = 0.96) while squashing false alarms. This ability is absent in the baselines, which treat all sensor readings as equally reliable.

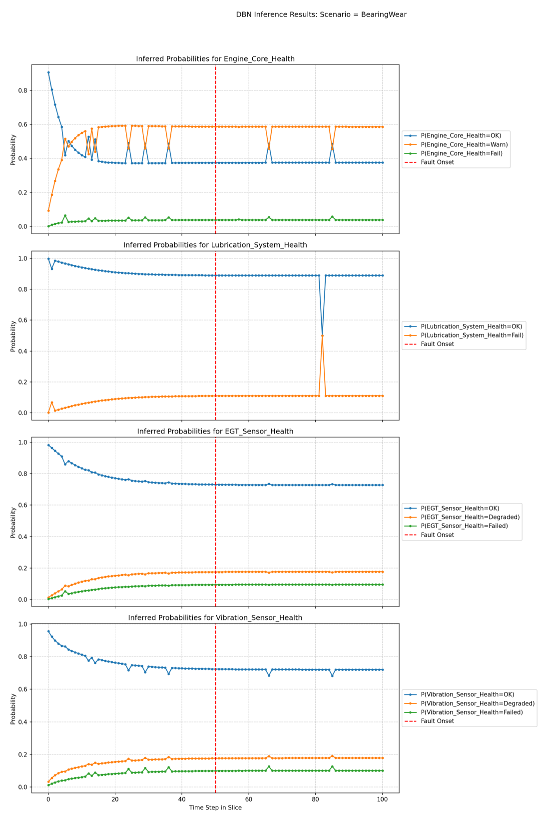

Fig. 3. Inference results for the *Bearing Wear* scenario. The DBN detects increasing fault probability in the engine core, while preserving stability in unrelated subsystems.

Finally, inference plots across time confirm that the Full DBN transitions gradually from *Healthy* to *Warn/Fail* states, in line with how degradation is expected to manifest in real systems. These plots, along with confusion matrices, visually confirm the model's temporal awareness.

C. Robustness Analysis

The robustness of the Full DBN system is evaluated by analyzing performance under sensor failure conditions and noisy inputs. Two components of the model contribute to this robustness: the Markov Random Field (MRF) smoother and the explicit modeling of sensor health within the DBN.

The MRF smoother plays a key role in reducing instability in the vibration channels. In fault-free scenarios, the vibration readings may still contain stochastic changes due to noise or sampling variance. The MRF enforces local consistency between *Vib1* and *Vib2*, filtering out fault transitions. This has shown to reduce false positive detection of *Bearing Wear* when no mechanical degradation is present.

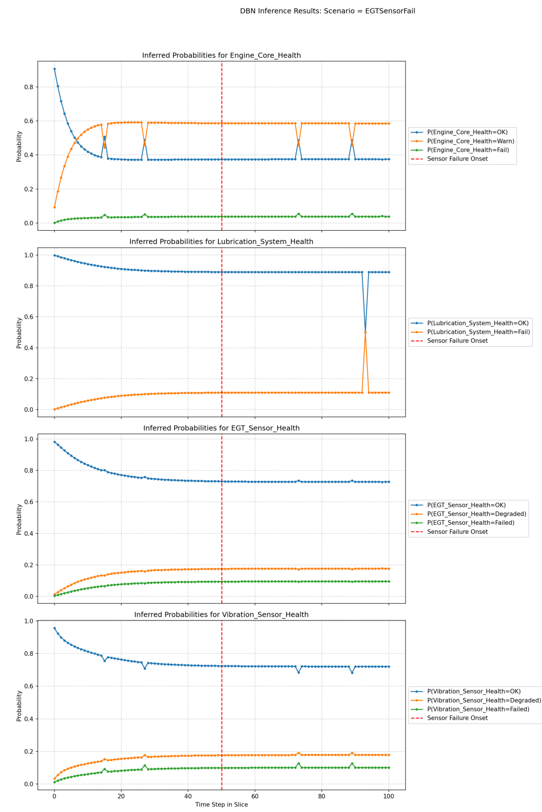

Fig. 4. Inference results for the *EGT Sensor Failure* scenario. The DBN correctly identifies the fault as a sensor failure, without erroneously raising core component health warnings.

The sensor health nodes within the DBN enable the model to reduce weight of evidence from faulty sensors. In the *EGT Sensor Failure* and *Vibration Sensor Failure* scenarios, the Full DBN maintains accurate component health classification, correctly related inconsistencies to sensor degradation rather than actual engine faults. In contrast, the Vanilla DBN—lacking any notion of sensor reliability—frequently misclassified these scenarios as real mechanical failures.

This robustness is particularly important in real-world applications where sensor malfunctions are common. By explicitly representing sensor state, the Full DBN avoids continuous misdiagnoses. These results highlight the importance of probabilistic modeling for safety-critical fault detection systems.

V. Discussion

The results demonstrate that the Full DBN architecture significantly outperforms the usual methods in both accuracy and interpretability. Unlike threshold-based classifiers, which treat all sensors as equally reliable, the DBN integrates expert-

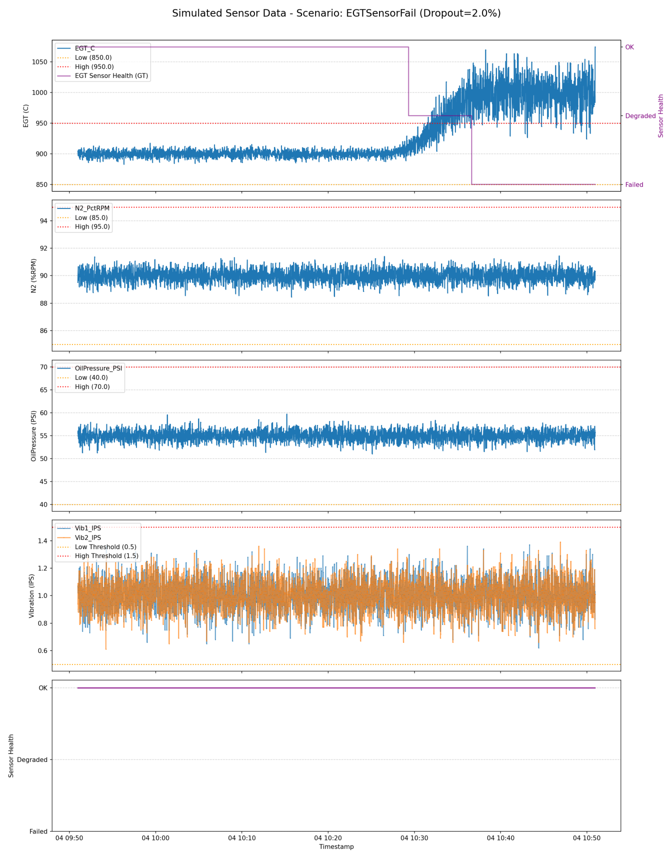

Fig. 5. Raw sensor readings for the *EGT Sensor Failure* scenario. While EGT appears erratic post-failure, other sensors remain stable. The DBN uses sensor health modeling to prevent false mechanical fault classification.

informed structure and explicitly models sensor reliability. This enables it to correctly attribute abnormal readings to sensor failures rather than engine faults, preserving diagnostic correctness.

The use of temporal transitions allows the DBN to smooth its beliefs over time and resist false state changes, unlike static models which often overreact to short-lived noise. The MRF preprocessing step increases the reliability of vibration data, especially under borderline or noisy conditions. Together, these features give a system capable of fault detection with high precision while maintaining low false positive rates for critical components.

However, several limitations are still present. The DBN structure and CPTs were manually defined and not learned from data. This limits adaptability and may bias the model toward expected behaviors.

The system is designed with a conservative bias: it favors detecting faults early, even at the risk of occasionally misclassifying normal states. This reflects the safety-critical nature of aviation, where early intervention is preferable to missed detections. Nonetheless, this tradeoff may reduce operator trust if false alarms are too frequent, and future work should explore calibrated probabilistic thresholds or operator-in-the-loop feedback systems.

VI. Conclusion

We developed a probabilistic fault detection framework for aircraft engines based on a Dynamic Bayesian Network (DBN) combined with a Markov Random Field (MRF) smoother. The model processes simulated sensor data to infer component health and identify faults such as *Oil Leak* and *Bearing Wear*. Key contributions include explicit modeling of sensor health, temporal belief propagation, and vibration smoothing for improved reliability.

Experimental results demonstrate that the Full DBN achieves significantly higher fault detection performance than a Vanilla DBN or Rule-Based baseline. The model maintains robustness under sensor failure conditions and suppresses false positives using structured probabilistic reasoning. It achieves this without any learned parameters, relying solely on expert-informed CPTs.

Future work includes extending the system to real-world data, incorporating structural learning for CPTs and network topology, and reintroducing HMM or Kalman-based modules for continuous sensor monitoring. These additions would enable a more adaptive and generalizable system for safety-critical diagnostics.

References

[1] E. Mast and J. Reed, "Bayesian belief networks for fault identification in aircraft gas turbine engines," *Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering*, vol. 219, no. 6, pp. 481–486, 2005.

[2] Y. Liu, H. Zhang, and Y. Wang, "Bayesian network–based fault diagnostic system for aircraft environmental control system," *Nuclear Science and Engineering*, vol. 196, no. 5, pp. 1234–1245, 2022.

[3] C. Zheng, X. Pan, X. Chen, X. Yang, X. Xin, and L. Su, "An object-based markov random field model with anisotropic penalty for semantic segmentation of high spatial resolution remote sensing imagery," *Remote Sensing*, vol. 11, no. 23, p. 2878, 2019.

[4] S. Amin and J. Wang, "A dynamic bayesian network based methodology for fault diagnosis of subsea systems," *Ocean Engineering*, vol. 186, p. 106118, 2019.

[5] J. Diebel and S. Thrun, "An application of markov random fields to range sensing," in *Advances in neural information processing systems*, 2005, pp. 291–298.

[6] S. Roweis and Z. Ghahramani, "A unifying review of linear gaussian models," *Neural computation*, vol. 11, no. 2, pp. 305–345, 1999.