

Quantificação de incerteza e robustez em redes profundas

Cesar de Oliveira F. Silva

Setembro / 2026

Incerteza em redes profundas — por que importa?

- Em aplicações ambientais, decisões são tomadas sob risco:
 - cheias;
 - secas;
 - blooms algais;
 - eventos extremos de qualidade da água e clima.
- Uma previsão pontual sozinha não informa o grau de confiança da decisão.
- Por isso, o modelo deve produzir não apenas uma estimativa, mas também uma medida de **incerteza preditiva**.

Mensagem central

Em problemas ambientais, prever bem a média não basta: é preciso quantificar o risco associado à previsão.

Tipos principais de incerteza

- A incerteza preditiva costuma ser decomposta em dois componentes principais:
 - **aleatória (aleatoric);**
 - **epistêmica (epistemic).**
- Cada uma tem origem diferente e exige estratégias distintas de modelagem.

Intuição

- **Aleatória:** variabilidade intrínseca do sistema ou ruído de observação.
- **Epistêmica:** falta de conhecimento do modelo sobre o fenômeno.

Decomposição da incerteza

- **Incerteza aleatória:**

- ruído de medição;
- variabilidade natural;
- irreduzível com mais dados, em geral.

- **Incerteza epistêmica:**

- escassez de dados;
- extrapolação para novos regimes;
- desconhecimento dos parâmetros ou da estrutura do modelo.

Implicação prática

O ideal é modelar ambas: a primeira para representar variabilidade intrínseca e a segunda para sinalizar falta de confiança do modelo.

Ensembles de redes profundas

- Uma abordagem simples e forte é treinar vários modelos independentes.
- As diferenças entre previsões refletem parte da incerteza epistêmica.

Uso típico

- média das previsões → estimativa pontual;
- dispersão entre modelos → incerteza epistêmica.
- Vantagens:
 - bom desempenho empírico;
 - implementação conceitualmente simples;
 - robustez maior que um único modelo.
- Limitação principal: custo computacional elevado.

Dropout como aproximação Bayesiana

- Outra estratégia é manter o dropout ativo durante a inferência.
- Isso gera múltiplas previsões para a mesma entrada:

$$\hat{y}^{(1)}, \hat{y}^{(2)}, \dots, \hat{y}^{(T)}$$

- A técnica é conhecida como **MC Dropout**.

Interpretação

As várias *forward passes* aproximam uma distribuição preditiva, funcionando como uma aproximação variacional da incerteza epistêmica.

- Vantagem: fácil de implementar.
- Limitação: é uma aproximação, nem sempre bem calibrada.

Modelos probabilísticos diretos

- Em vez de prever apenas um valor, o modelo pode prever parâmetros de uma distribuição.

Exemplo Gaussiano

$$\hat{y} \sim \mathcal{N}(\mu_{\theta}(x), \sigma_{\theta}^2(x))$$

- A rede aprende:
 - média predita $\mu_{\theta}(x)$;
 - variância predita $\sigma_{\theta}^2(x)$.
- O treinamento pode usar **Negative Log-Likelihood (NLL)**.

Extensões

regressão quantílica, mixture density networks, distribuições não Gaussianas.

Construção de intervalos preditivos

- Intervalos podem ser obtidos de diferentes formas:
 - por quantis aprendidos diretamente;
 - por amostragem com ensembles;
 - por MC Dropout;
 - por modelos probabilísticos explícitos.
- O importante é que o intervalo reflita a incerteza real do problema.

Avaliação

- cobertura empírica vs nominal;
- largura média do intervalo;
- compromisso entre intervalo estreito e cobertura adequada.

Calibração preditiva

- Um modelo calibrado atribui probabilidades consistentes com frequências observadas.
- Exemplo: entre previsões com 80% de confiança, espera-se cerca de 80% de acerto ou cobertura.

Ferramentas

- reliability diagrams;
- Expected Calibration Error (ECE);
- análise de cobertura de intervalos.

Técnicas de ajuste

temperature scaling, isotonic regression e recalibração pós-treinamento.

Sharpness vs calibration

- Bons intervalos não devem ser apenas corretos em cobertura, mas também informativos.
- Dois objetivos entram em tensão:
 - **calibration:** intervalos compatíveis com a frequência observada;
 - **sharpness:** intervalos estreitos e úteis para decisão.

Equilíbrio desejável

Intervalos muito largos podem ter boa cobertura, mas baixa utilidade operacional.

Análise de sensibilidade a ruído

- Um modelo robusto não deve reagir de forma instável a pequenas perturbações plausíveis.
- Em dados ambientais, isso inclui:
 - ruído espectral;
 - deslocamentos temporais;
 - falhas parciais de sensor;
 - pequenas perturbações meteorológicas.

Objetivo

Avaliar como previsão e incerteza mudam quando a entrada é perturbada.

- Isso ajuda a revelar fragilidade, sobreconfiança e comportamento não físico.

Robustez a mudança de domínio

- Em aplicações ambientais, a distribuição de treino raramente coincide perfeitamente com a de operação.
- Mudanças comuns:
 - nova região geográfica;
 - novo sensor;
 - nova estação ou regime climático;
 - mudança no processo gerador dos dados.

Conceitos

- **covariate shift:** muda $p(x)$;
- **concept drift:** muda $p(y|x)$.

Mudança de regime climático e generalização

- Um modelo treinado em clima histórico pode falhar sob novos extremos ou padrões persistentes de mudança.
- Isso é especialmente crítico em:
 - previsão hidrológica;
 - eventos extremos;
 - qualidade da água;
 - downscaling climático.

Estratégias

- incluir períodos multi-decadais;
- usar reanálises e cenários climáticos;
- monitorar drift continuamente;
- incorporar estrutura física ao modelo.

Stress testing de modelos ambientais

- Além do teste padrão, é útil submeter o modelo a cenários severos.
- Exemplos:
 - chuvas intensas;
 - secas prolongadas;
 - combinações incomuns de variáveis;
 - observações fora do suporte típico de treino.

Perguntas importantes

- A previsão degrada de forma plausível?
- A incerteza aumenta quando deveria?
- O modelo permanece estável fora do domínio usual?

Comparação com modelos físicos

- Modelos físicos lidam naturalmente com incertezas estruturais e paramétricas.
- Ensembles físicos, como ensembles climáticos ou hidrológicos, já produzem distribuições ou intervalos.
- Modelos baseados em dados são mais flexíveis, mas dependem fortemente da representatividade do conjunto de treino.

Comparação prática

- modelos físicos: maior interpretabilidade e estrutura causal;
- redes profundas: maior flexibilidade e capacidade de ajuste;
- híbridos: combinação promissora entre robustez estrutural e desempenho empírico.

Integração físico-estatística

- Redes neurais podem complementar modelos físicos, em vez de substituí-los.
- Estratégias úteis:
 - correção de viés;
 - modelagem de resíduos;
 - uso da saída física como entrada adicional;
 - propagação de incerteza física para o modelo estatístico.

Direção promissora

Modelos híbridos que combinem conhecimento físico, estimação probabilística e aprendizado profundo.

Métricas para avaliação de incerteza

- Métricas devem avaliar não apenas erro médio, mas a qualidade da distribuição predita.
- Exemplos importantes:
 - **CRPS**: qualidade global da distribuição preditiva;
 - **Brier Score**: eventos binários;
 - cobertura empírica de intervalos;
 - sharpness e calibration.

Princípio

Uma boa incerteza é simultaneamente **calibrada, informativa e sensível ao risco real**.

Boas práticas operacionais

- Sempre reportar incerteza junto com a previsão pontual.
- Monitorar drift de dados e degradação de calibração após deploy.
- Recalibrar periodicamente quando o domínio operacional mudar.
- Documentar:
 - hipóteses do modelo;
 - limites de validade;
 - cenários onde a confiança pode falhar.

Em operação

Incerteza útil é aquela que ajuda a tomar decisão, priorizar resposta e comunicar risco.

Resumo estratégico

- 1 Modele explicitamente a incerteza, não apenas o erro médio.
- 2 Distinga incerteza aleatória de epistêmica.
- 3 Avalie calibração, sharpness e robustez sob domain shift.
- 4 Faça stress testing em cenários extremos e fora do suporte de treino.
- 5 Compare com modelos físicos, estatísticos e híbridos.