What are the challenges in using ML/AI for parameterization development?

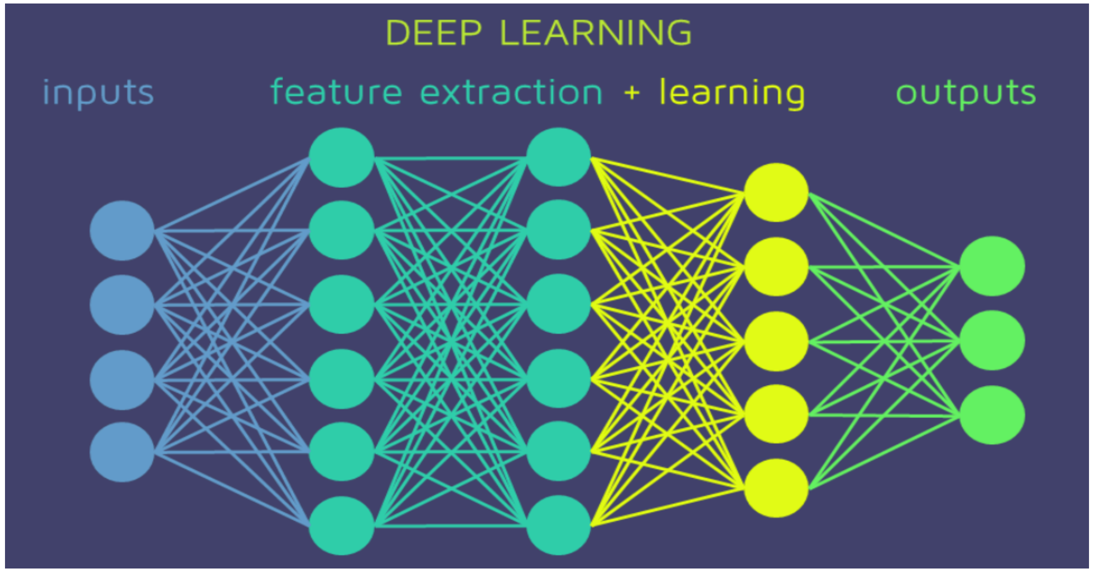
What are some opportunities to address these challenges?

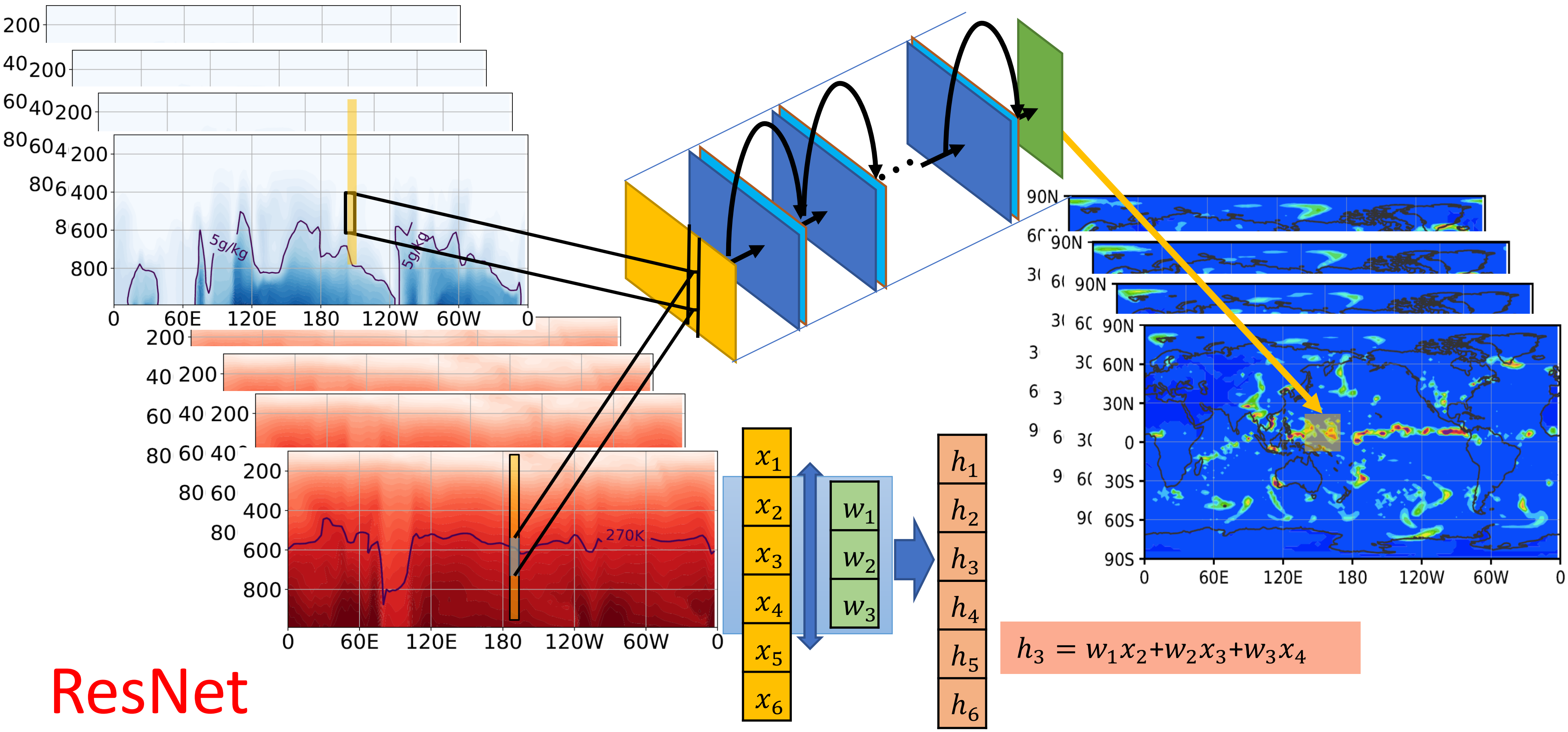
- **Can data-driven ML be used for parameterization development in Earth system models?**
    1. **Neural networks don't know physics laws. How to make them obey physics laws?**
    2. **How to ensure it can stably integrate with time in ESMs?**
- **How robust and interpretable are ML–based parameterizations?**
    1. **How to break the black box and peek inside?**
    2. **Are machine learned relationships between input and output physically meaningful?**
- **Should ML be used under current parameterization framework or be used to replace the current framework?**
    1. **How complex should ML-based emulators be (architecture, e.g. random forest, DNN, CNN, ResNet, …, depth etc.)?**
- **Opportunities?**

    Big data (observations, MMFs, GCRMs)

    Advances in computer algorithms, particularly advanced NN's and hardware

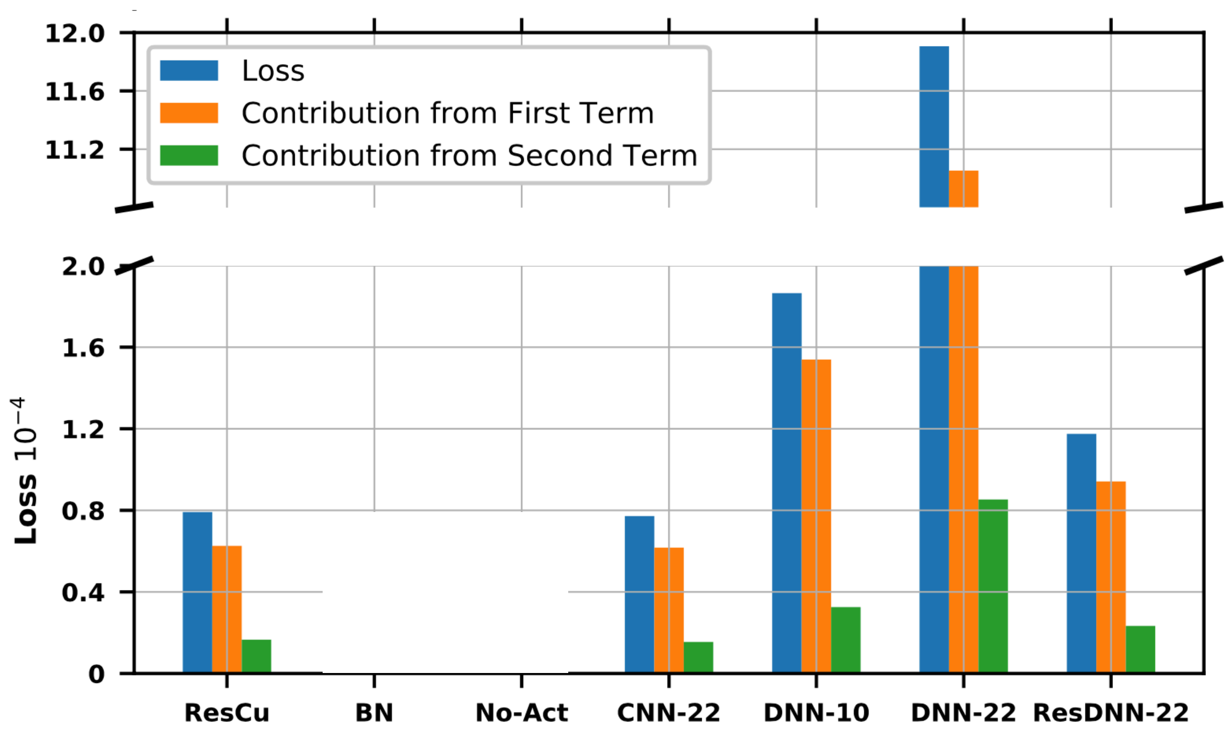
DNN


DEEP LEARNING
inputs
feature extraction + learning
outputs


200
400
600
800
0
60E
120E
180
120W
60W
5g/kg
270K
90N
60N
30N
0
30S
60S
90S
x_1
x_2
x_3
x_4
x_5
x_6
w_1
w_2
w_3
h_1
h_2
h_3
h_4
h_5
h_6
$h_3 = w_1x_2+w_2x_3+w_3x_4$

ResNet

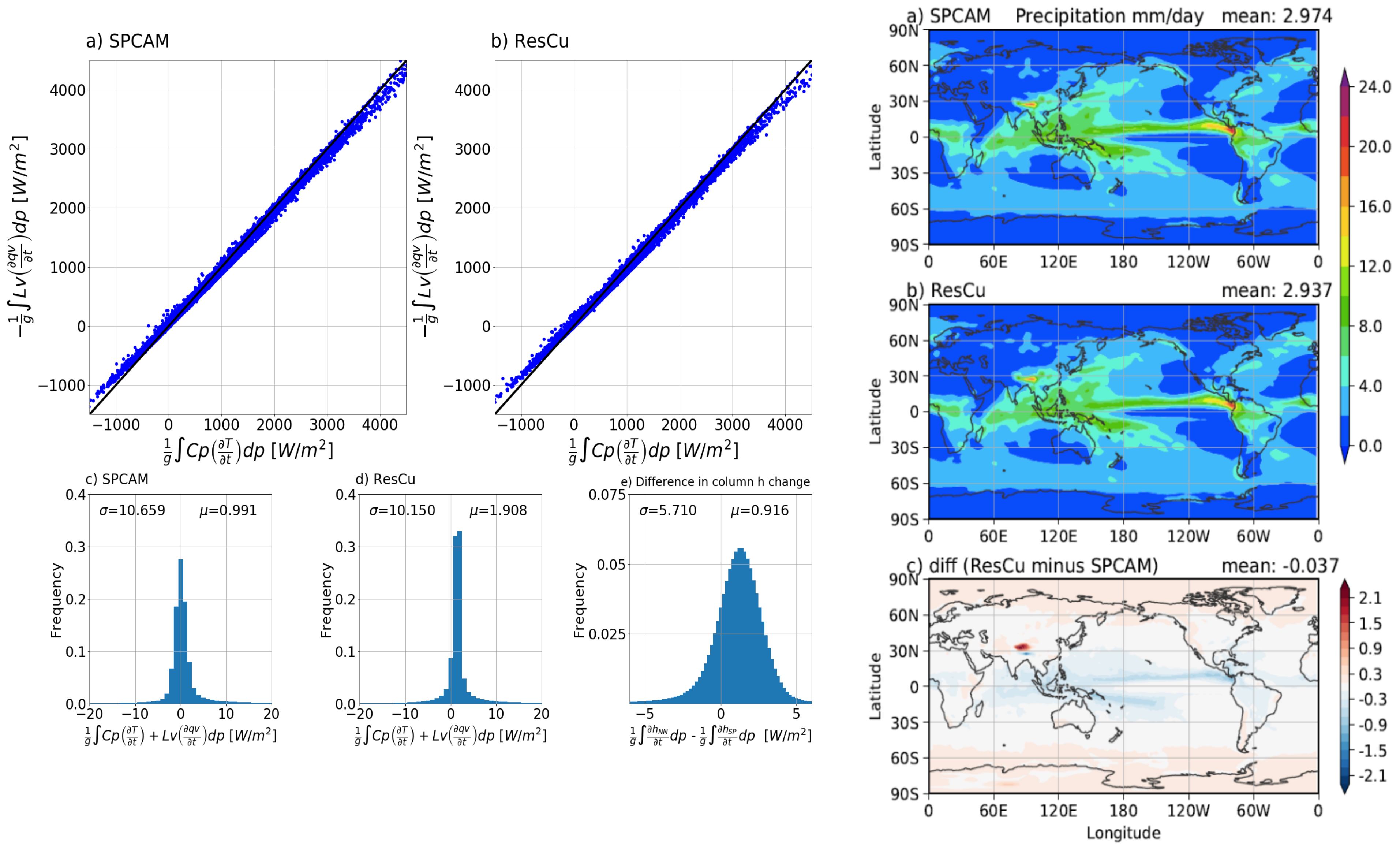
Sensitivity and Comparison of NN Structures



- Use of CNN reduces the loss by ~50% (compare ResCu and ResDNN-22)
- Advantage of ResNet vs. CNN is not significant for a 22-layer net (but is appreciable for deeper nets)
- DNN is relatively simpler, but less accurate (the loss is ~2x that of CNN)

$$loss = \|\hat{y} - y\|_2 + \lambda \left\| \frac{1}{g} \int_{pt}^{pb} \frac{\partial h_{SP}}{\partial t}\, dp - \frac{1}{g} \int_{pt}^{pb} \frac{\partial h_{NN}}{\partial t}\, dp \right\|_2$$

NN prediction can be very accurate



Precipitation difference is <0.5 mm/day