# Updates on SHEEP MODEL

**Elise Hinkle**

March 18, 2026

# Post-Fix Loss Curves – Variations

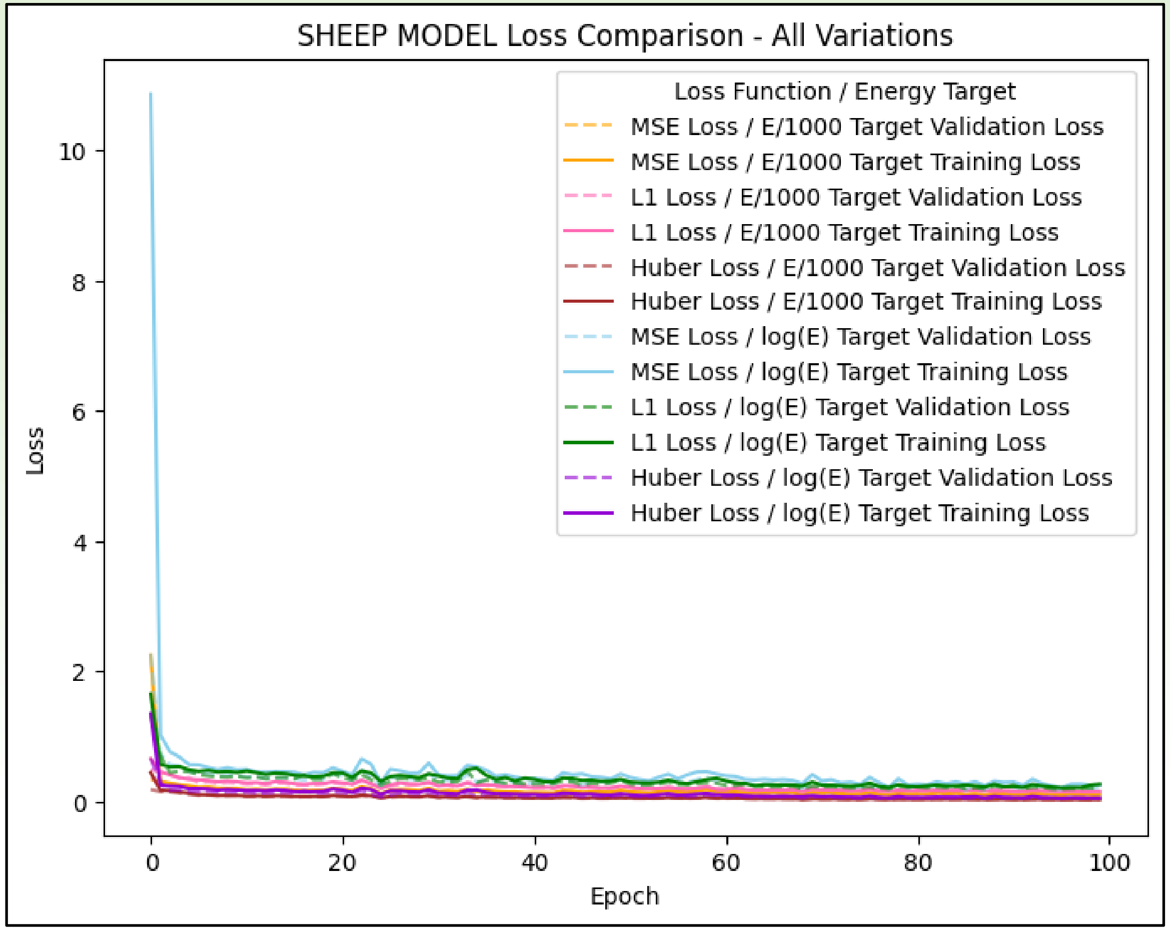
- Test on 100k event sample (80k train / 10k validation / 10k test)
  - All events electron showers between 0 and 2 GeV
- Train each variation for 100 epochs (full pass through train set)
- Constants:
  - Batch size = 3200 samples
  - Exponential moving average momentum = 0.01
  - Learning rate = 0.0001 (one test with increasing to 0.001 after epoch 20)
- Variables (7 variations total):
  - Loss function (MSE vs. L1 vs. Huber)
  - Target value (E/1000 vs. log(E))
  - One variation with MSE Loss | Target value E/1000 | LR increase

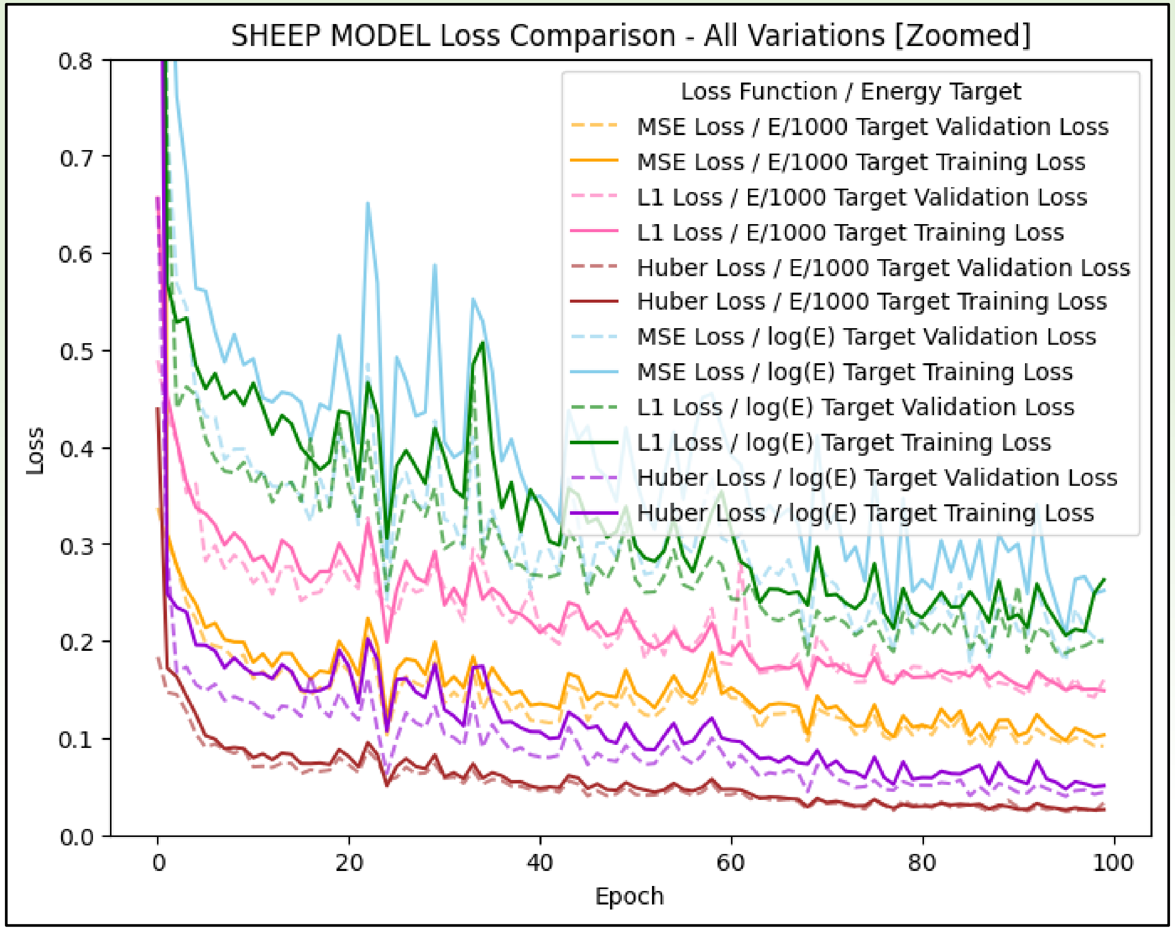
# Post-Fix Loss Curves – Loss Function Explanation

$\boldsymbol{y_i}$= true KE, $\boldsymbol{\hat{y}_i}$ = SHEEP prediction, $\boldsymbol{\delta}$= 1.0 (default value)

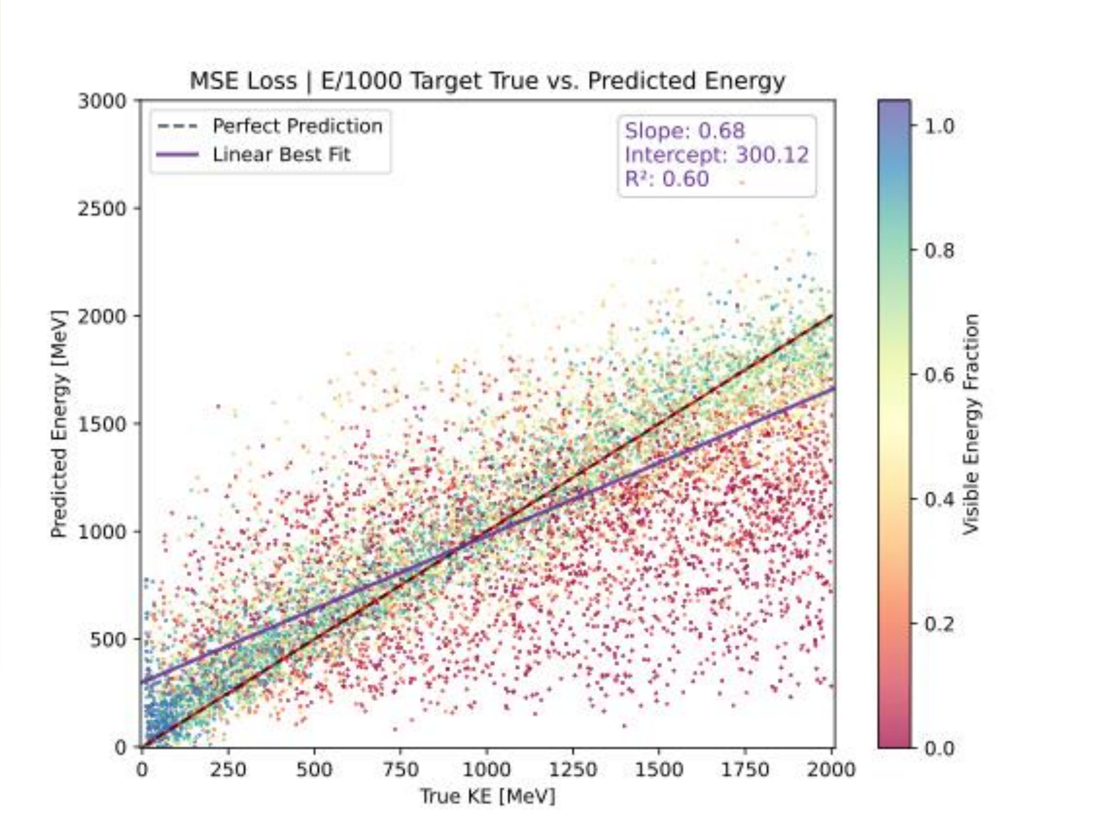
- **MSE Loss:** $\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2$
  - Most sensitive to large errors (faster convergence), but also outliers
- **L1 Loss:** $\frac{1}{N}\sum_{i=1}^{N}|\hat{y}_i - y_i|$
  - More robust to outliers than MSE, but not differentiable at 0
- **Huber Loss:** $\frac{1}{N}\sum_{i=1}^{N}\begin{cases}0.5(\hat{y}_i - y_i)^2, & \text{if } |\hat{y}_i - y_i| < \delta \\ \delta \cdot (|\hat{y}_i - y_i| - 0.5 * \delta), & \text{otherwise}\end{cases}$
  - Combination of MSE loss (close to 0 – differentiability) and L1 loss (for larger errors to improve robustness to outliers)

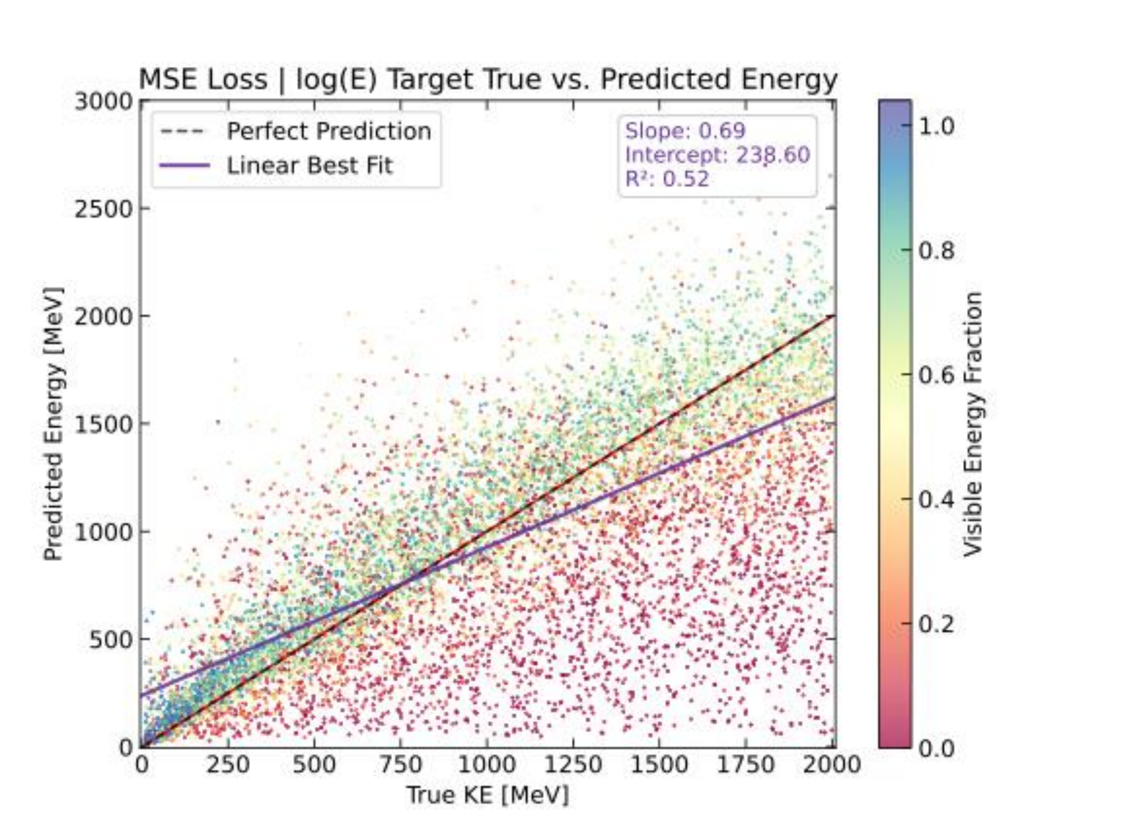
# Post-Fix Loss Curves – All Variations



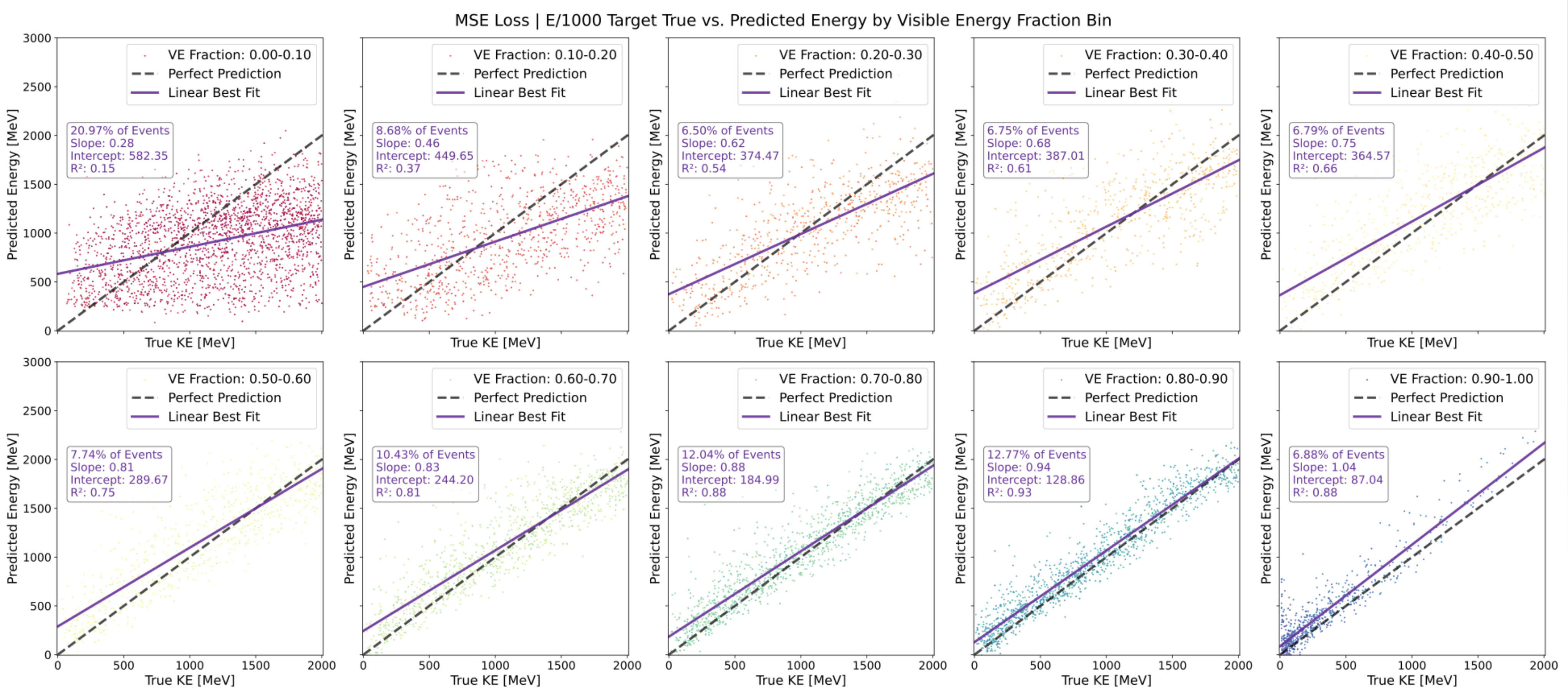



# MSE Loss, E/1000 Target vs. log(E) Target

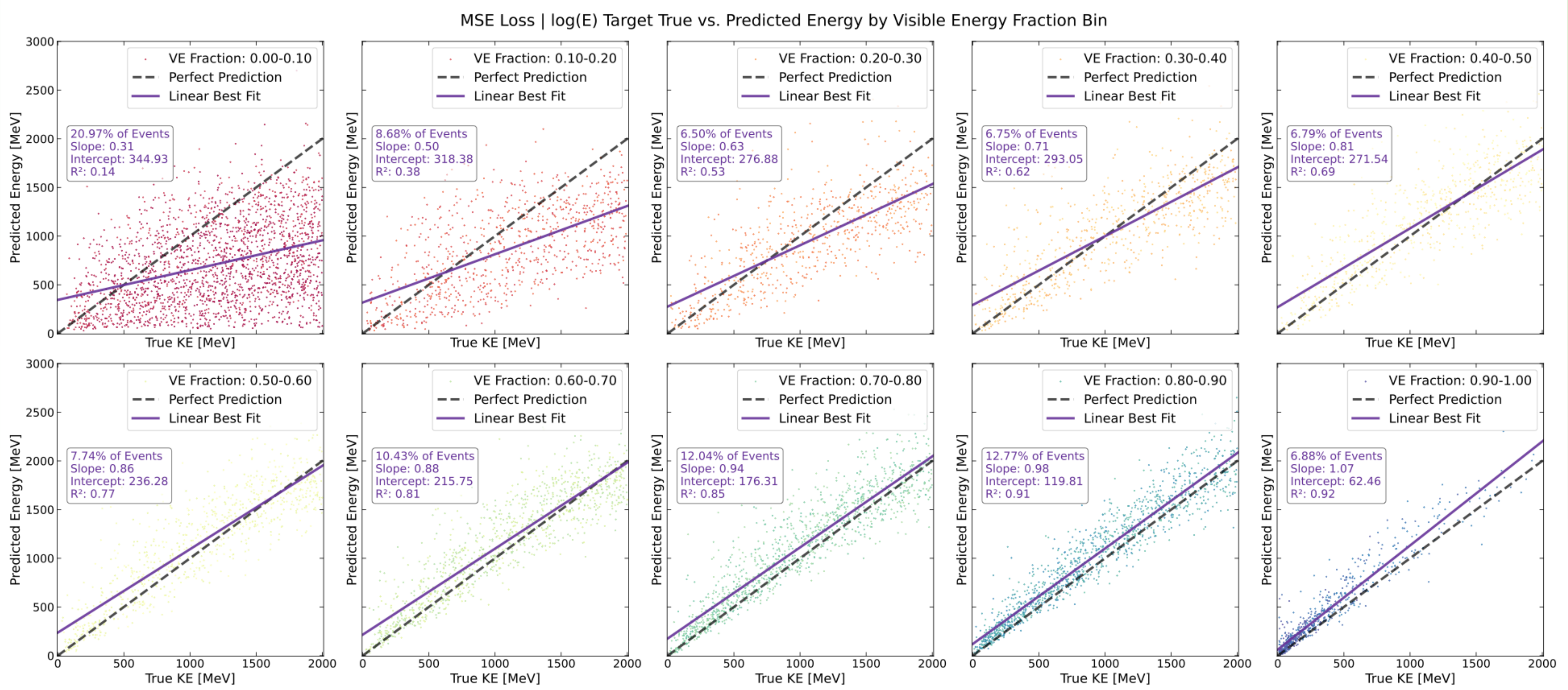


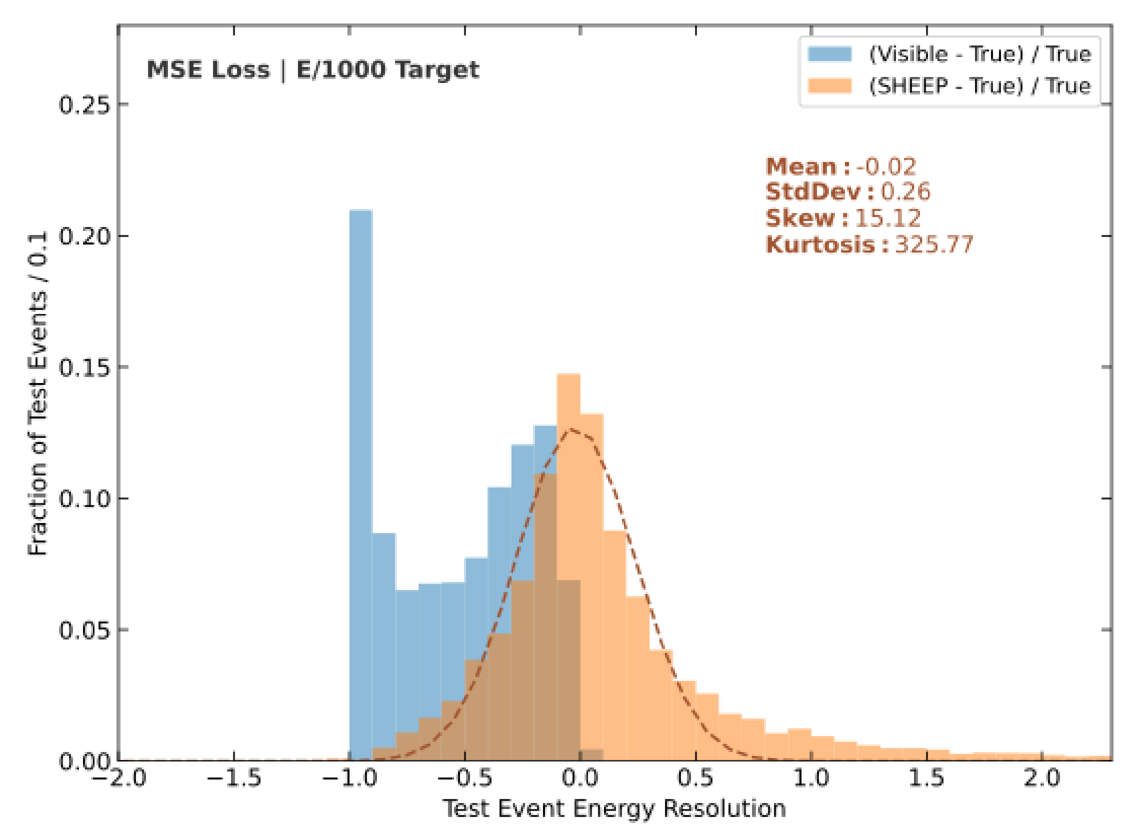
# MSE Loss, E/1000 Target vs. log(E) Target

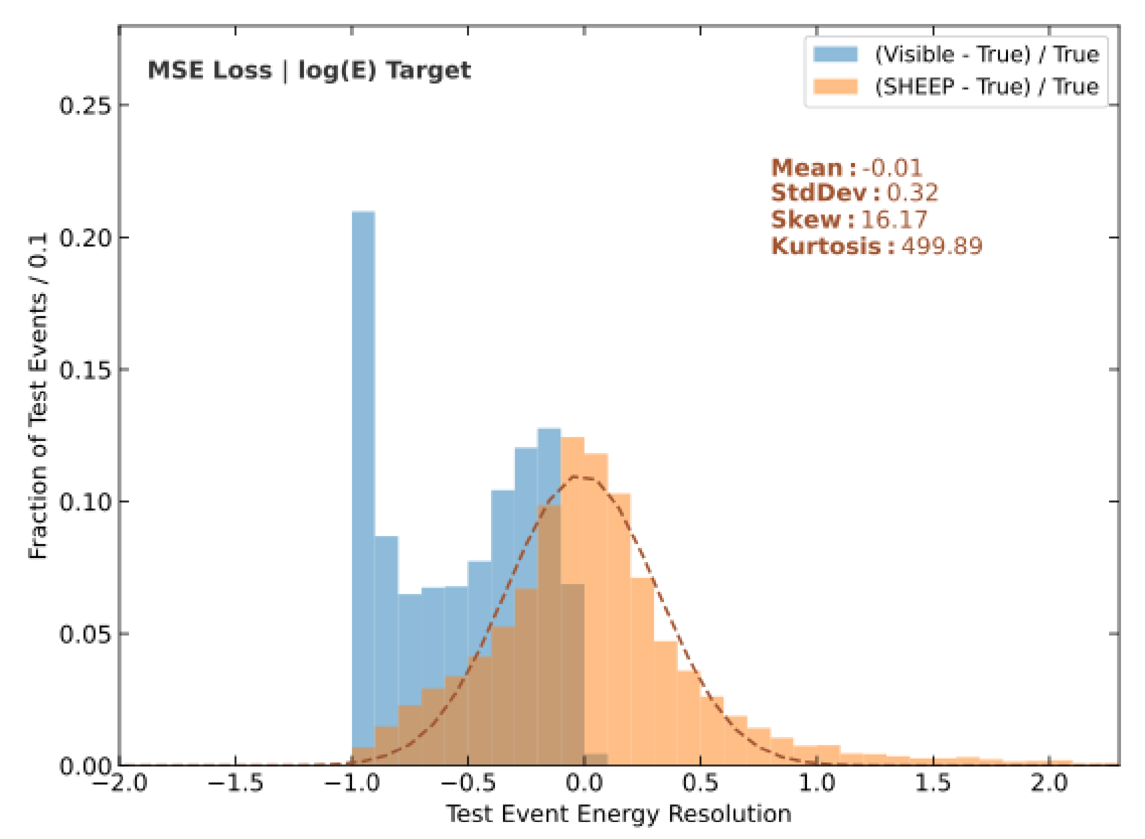


# MSE Loss, E/1000 Target vs. log(E) Target



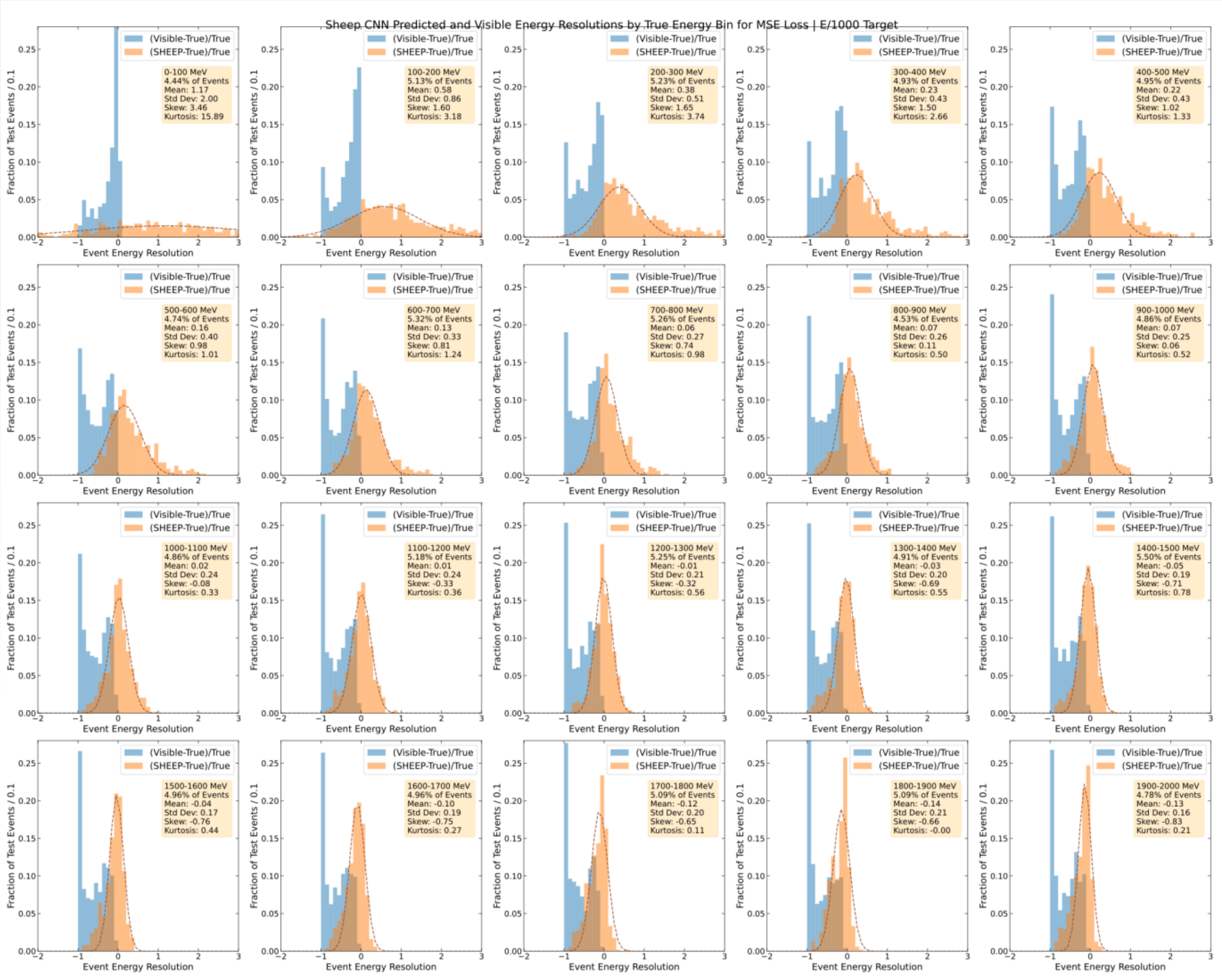
# MSE Loss, E/1000 Target vs. log(E) Target



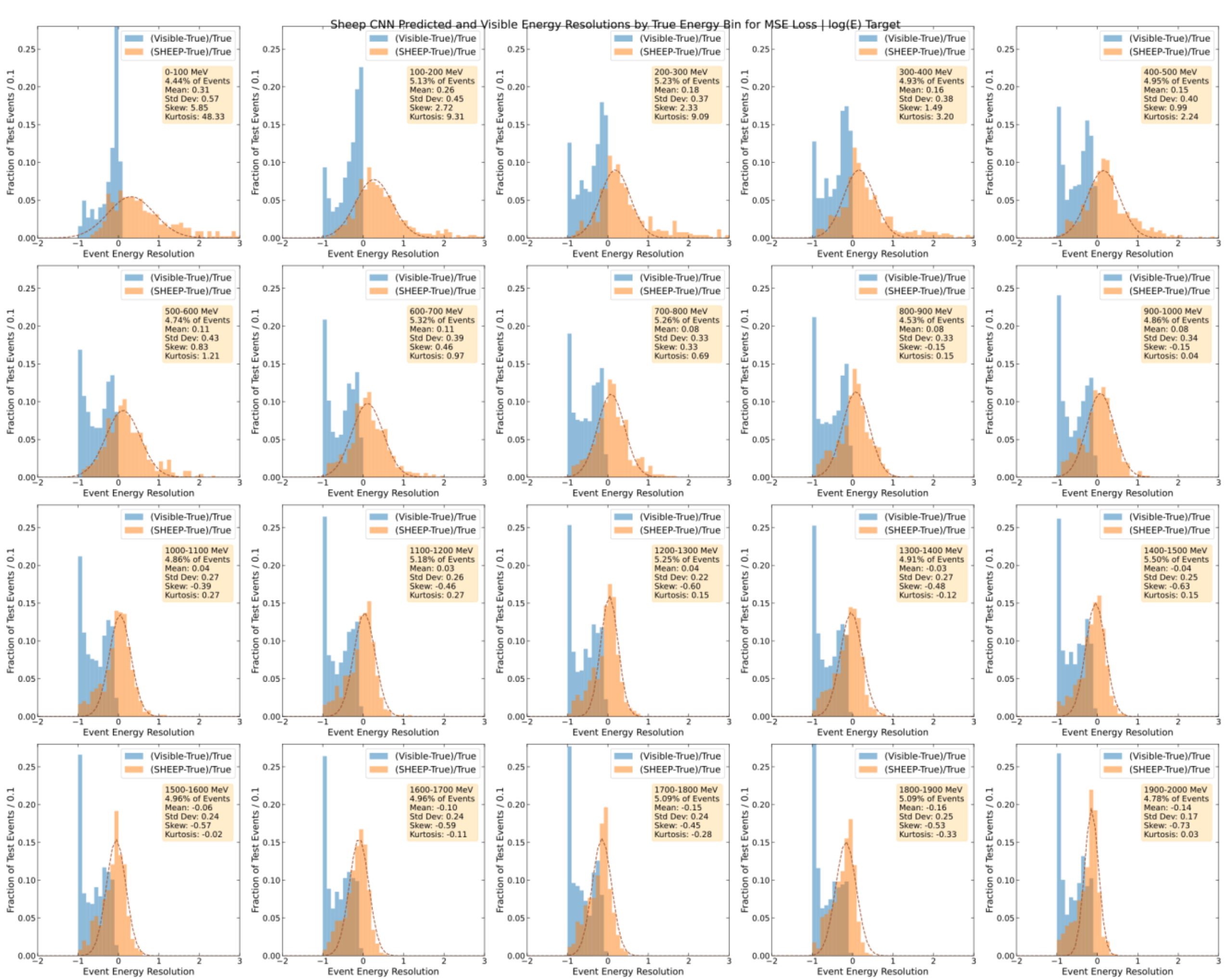



Best [Val Loss] Epoch: 87

# MSE Loss, E/1000 Target vs. log(E) Target

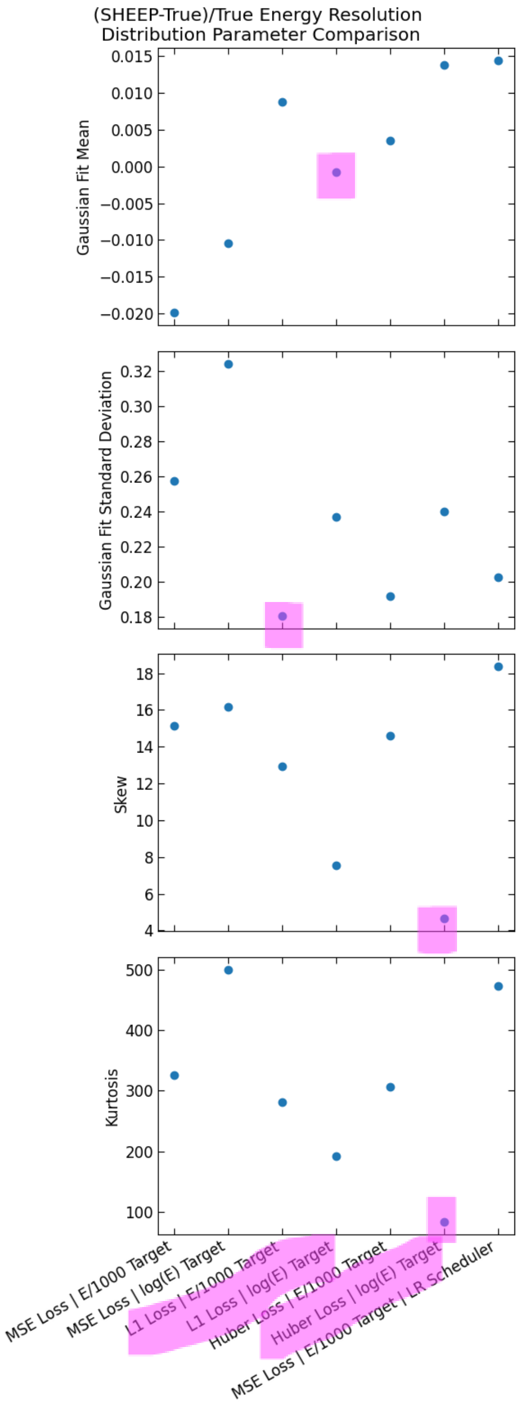


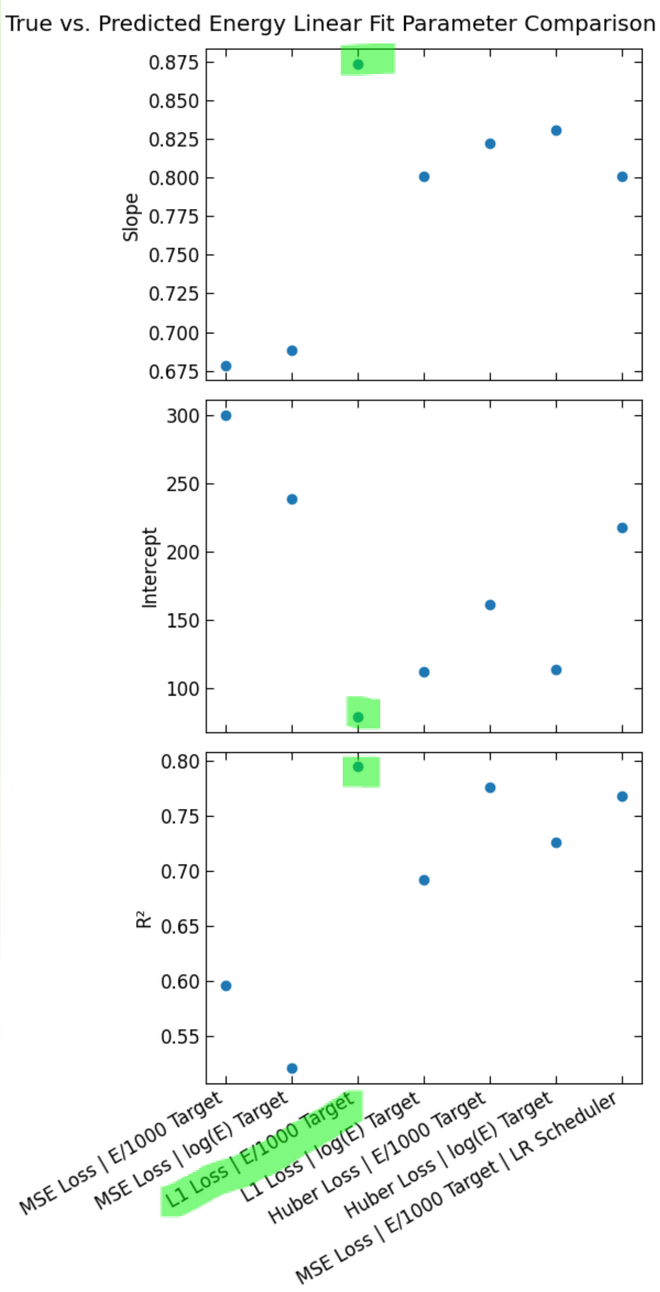
# MSE Loss, E/1000 Target vs. log(E) Target



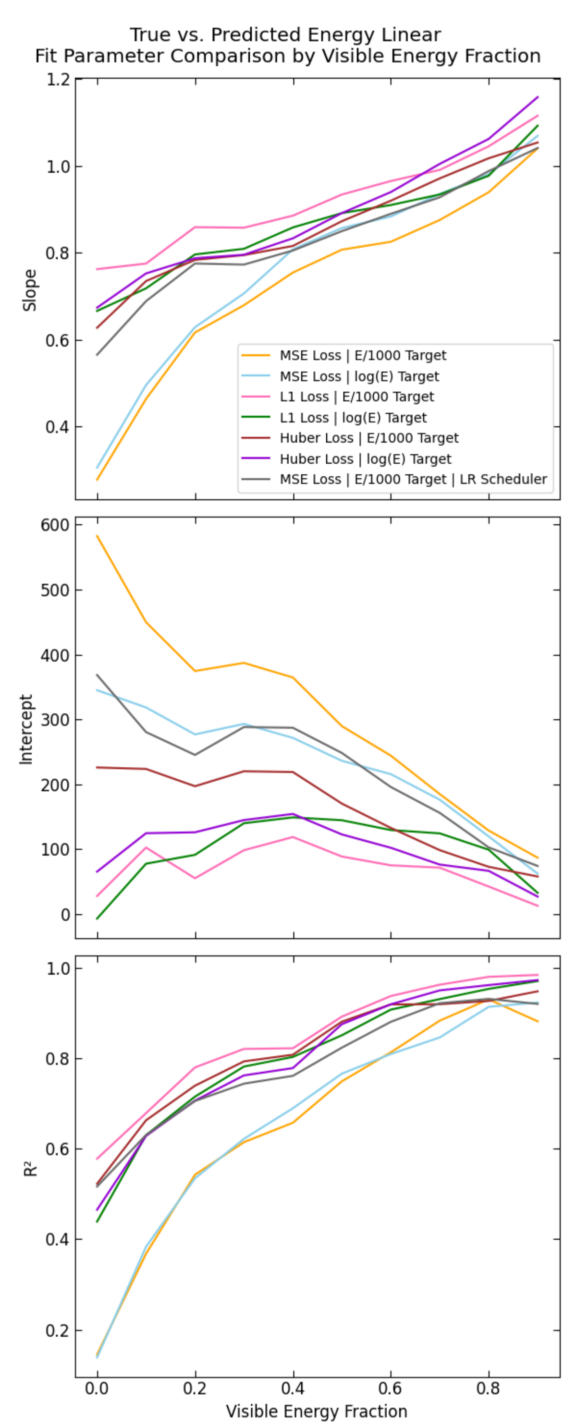
# Comparing Models

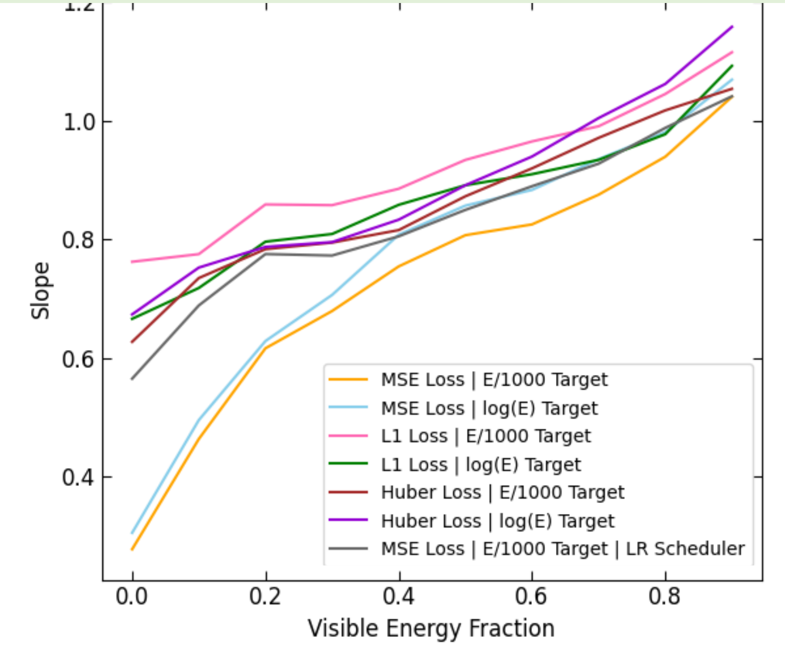
- ”Best Epoch” version varies considerably
  - Ranges from 68 (L1 | log(E)) to 98 (L1 | E/1000)

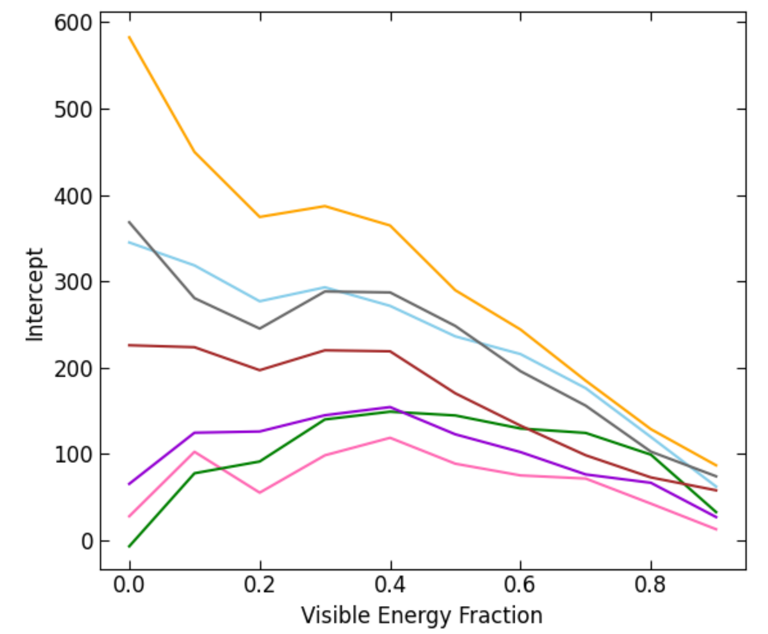


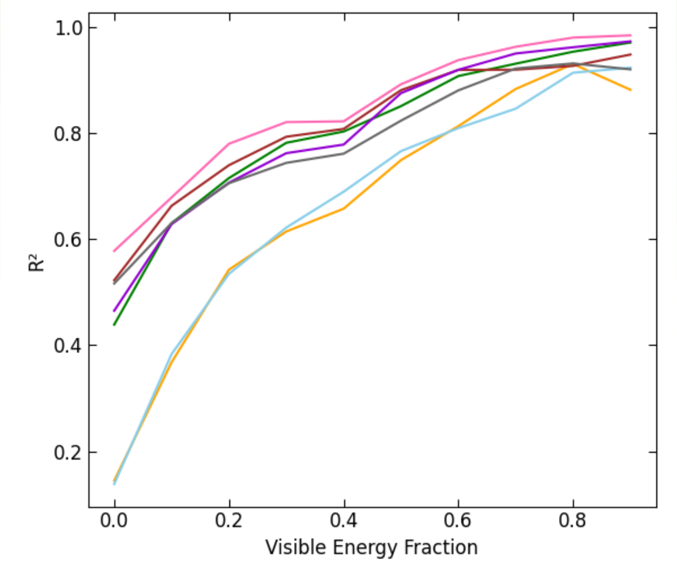
# Comparing Models – By Visible Energy Fraction



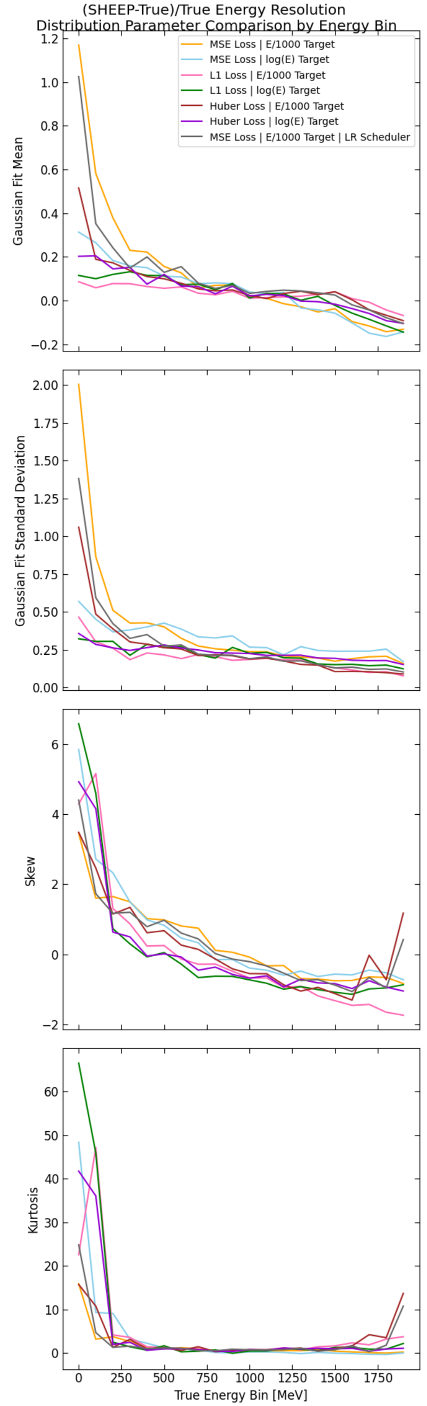



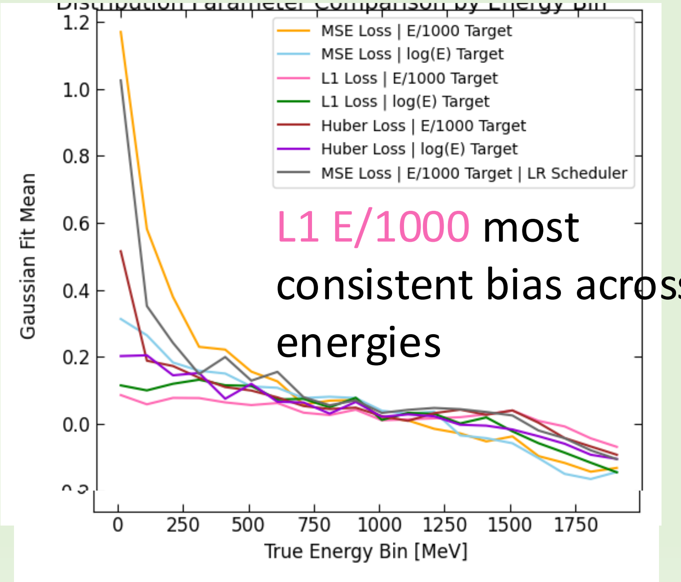




- Increasing visible energy fraction = closer to slope of 1, lower offset (intercept)
- High VE bins over predict

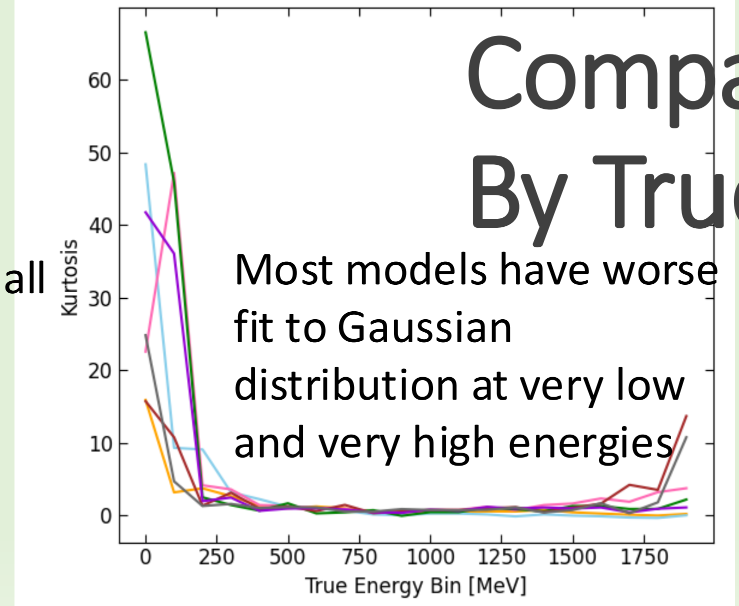


# Comparing Models – By True Energy Bin

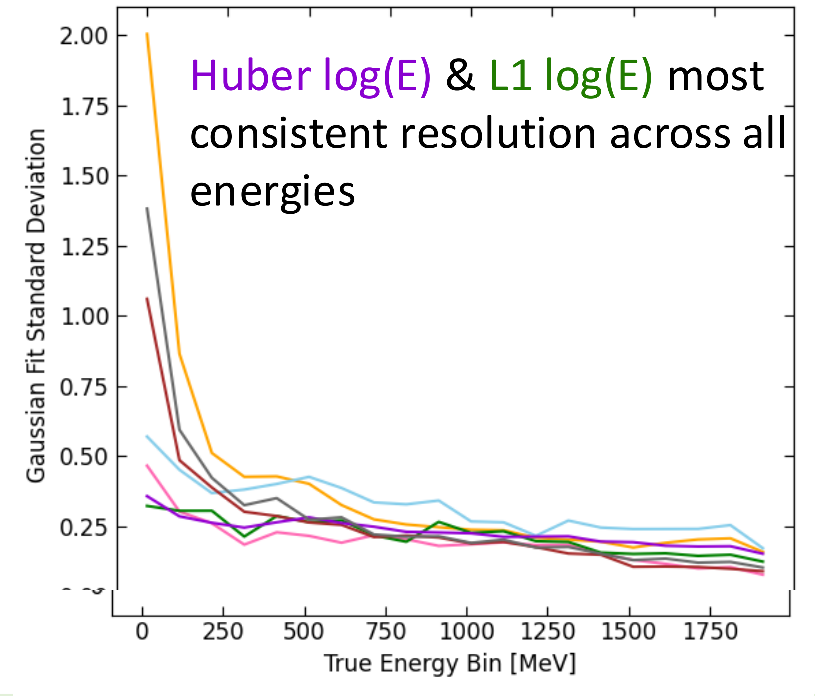


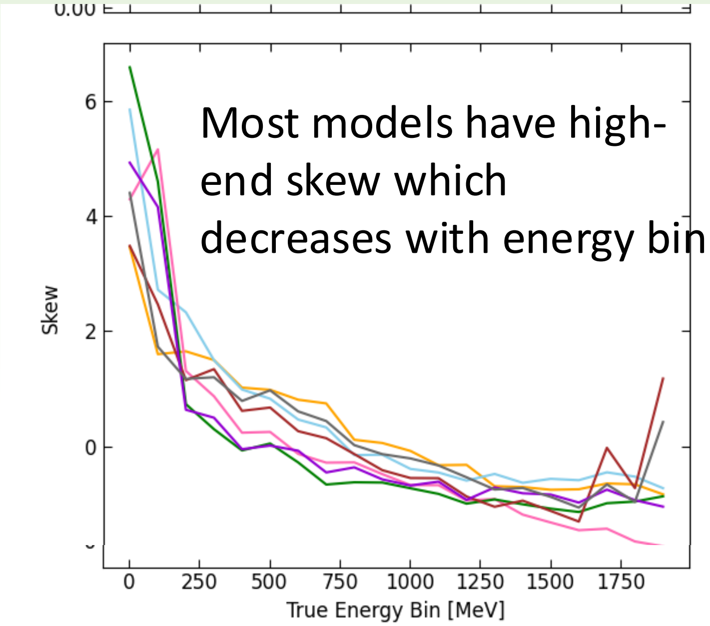

L1 E/1000 most consistent bias across all energies




Most models have worse fit to Gaussian distribution at very low and very high energies

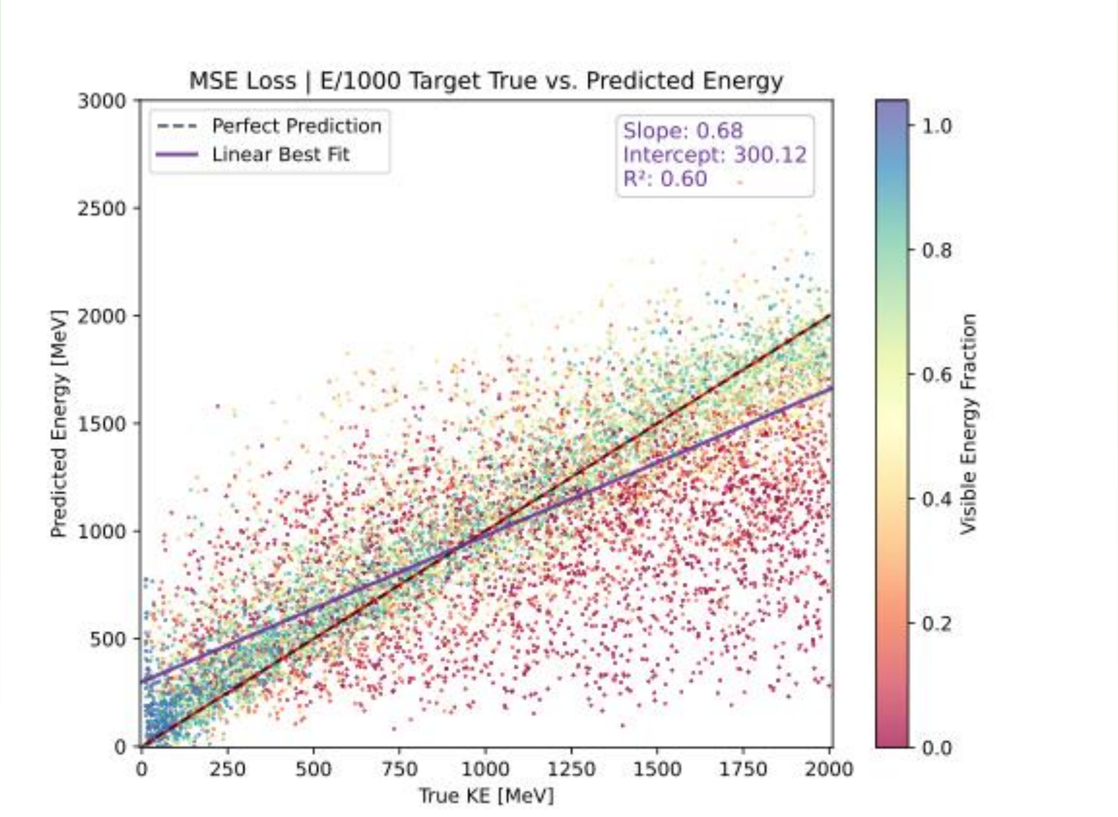


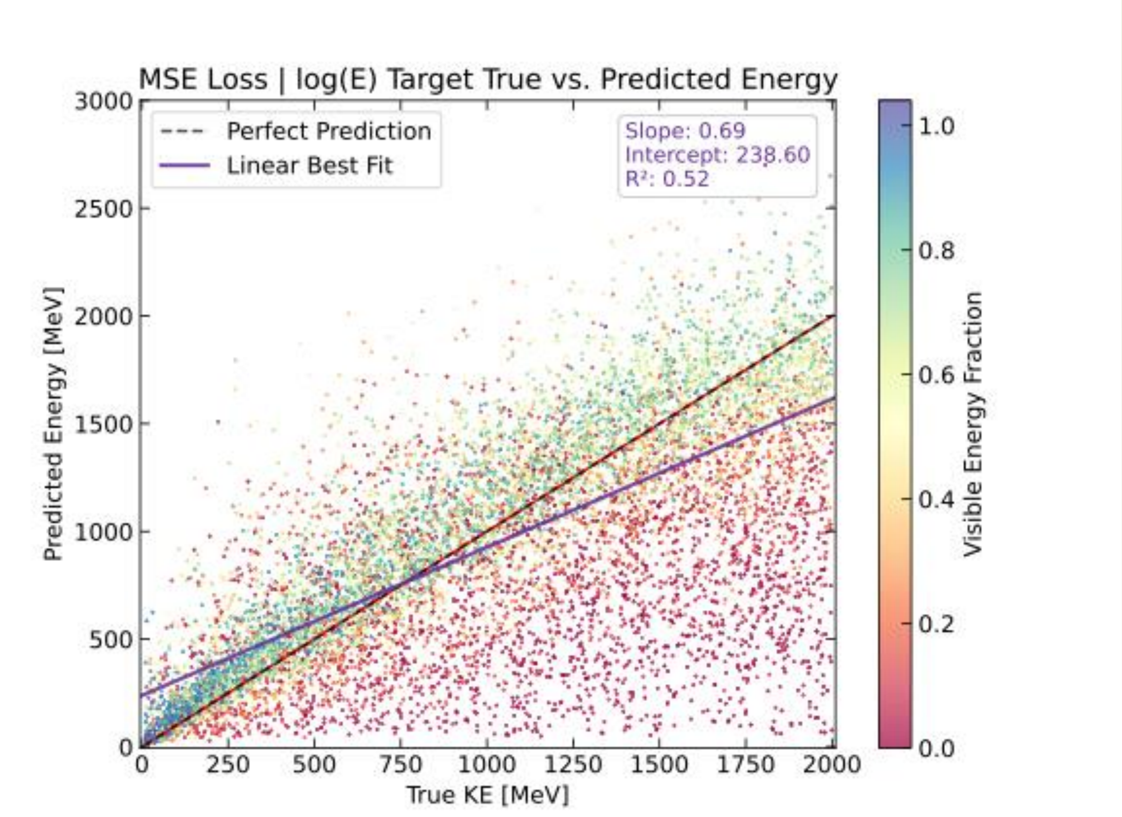

Huber log(E) & L1 log(E) most consistent resolution across all energies



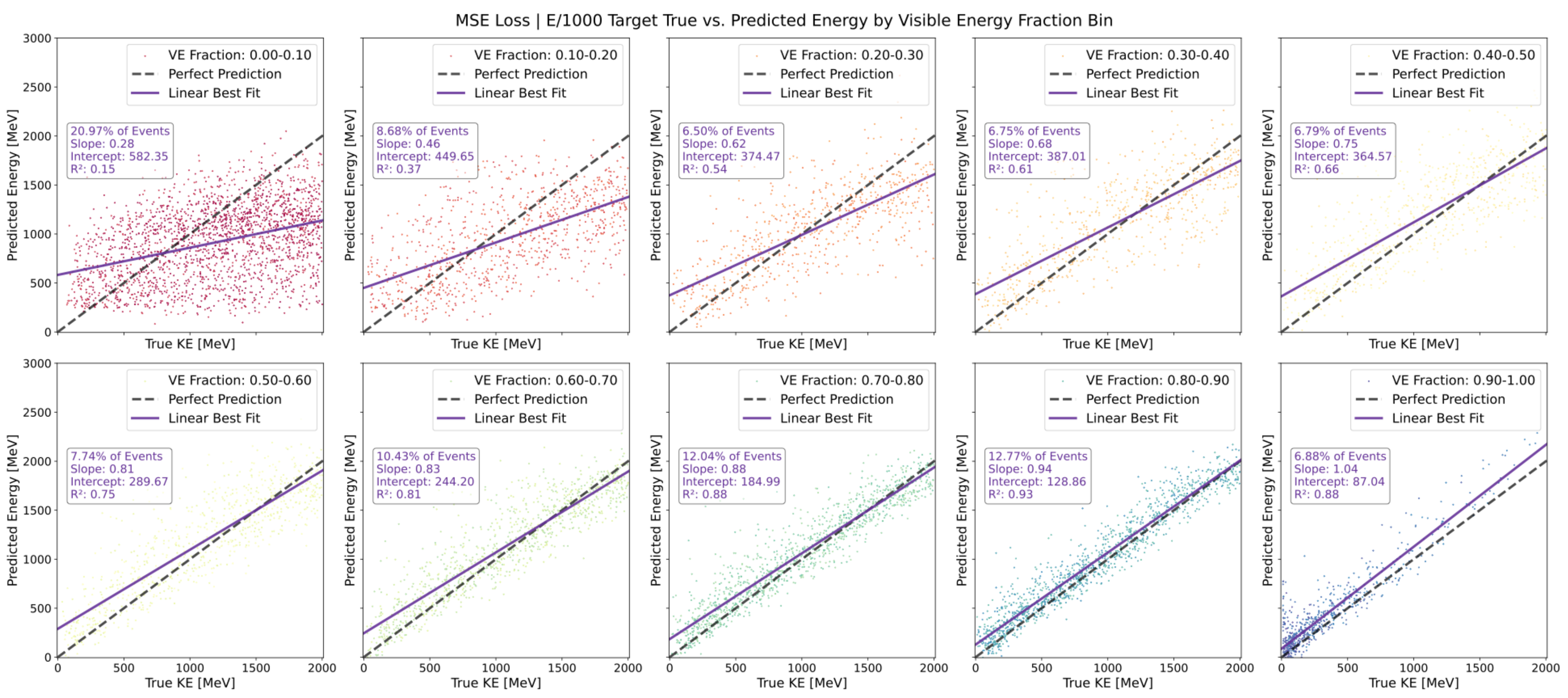

Most models have high-end skew which decreases with energy bin

# MSE Loss, E/1000 Target vs. log(E) Target

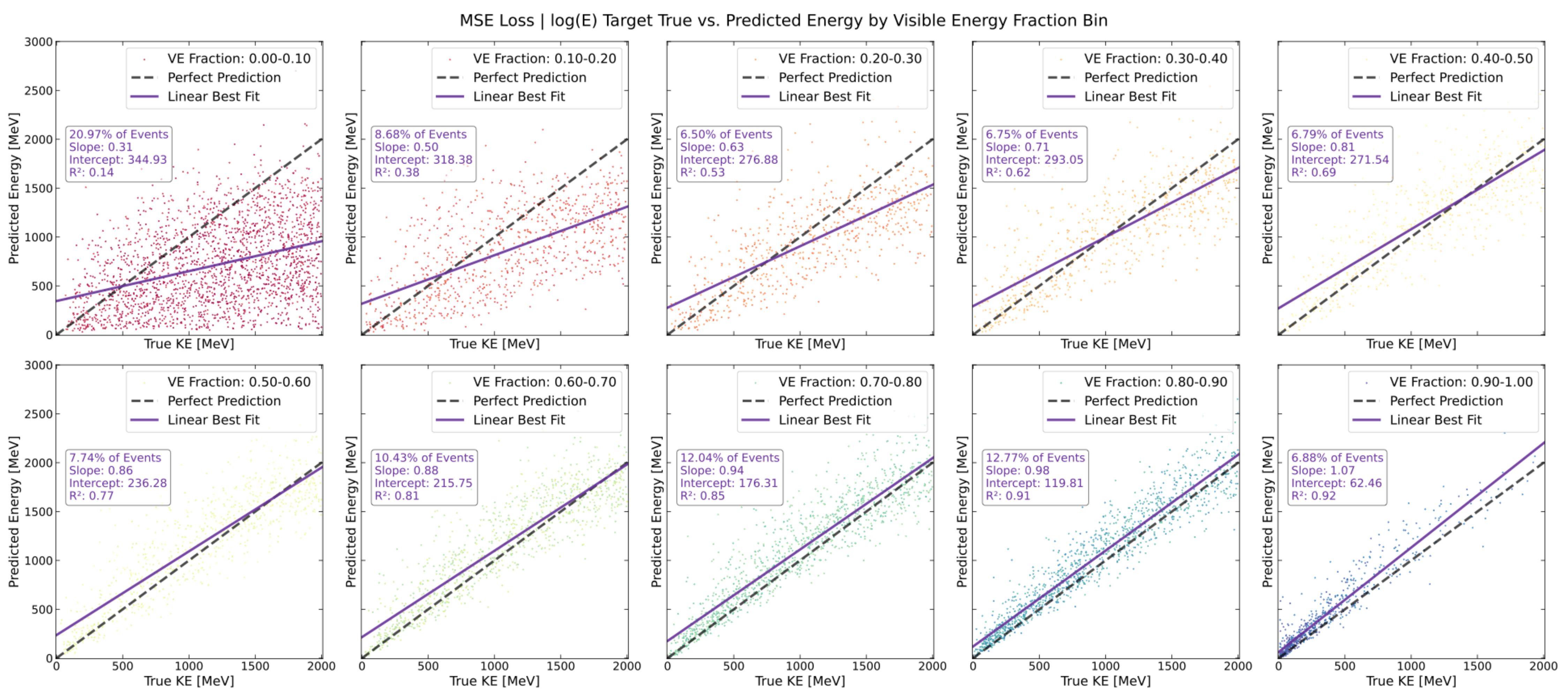


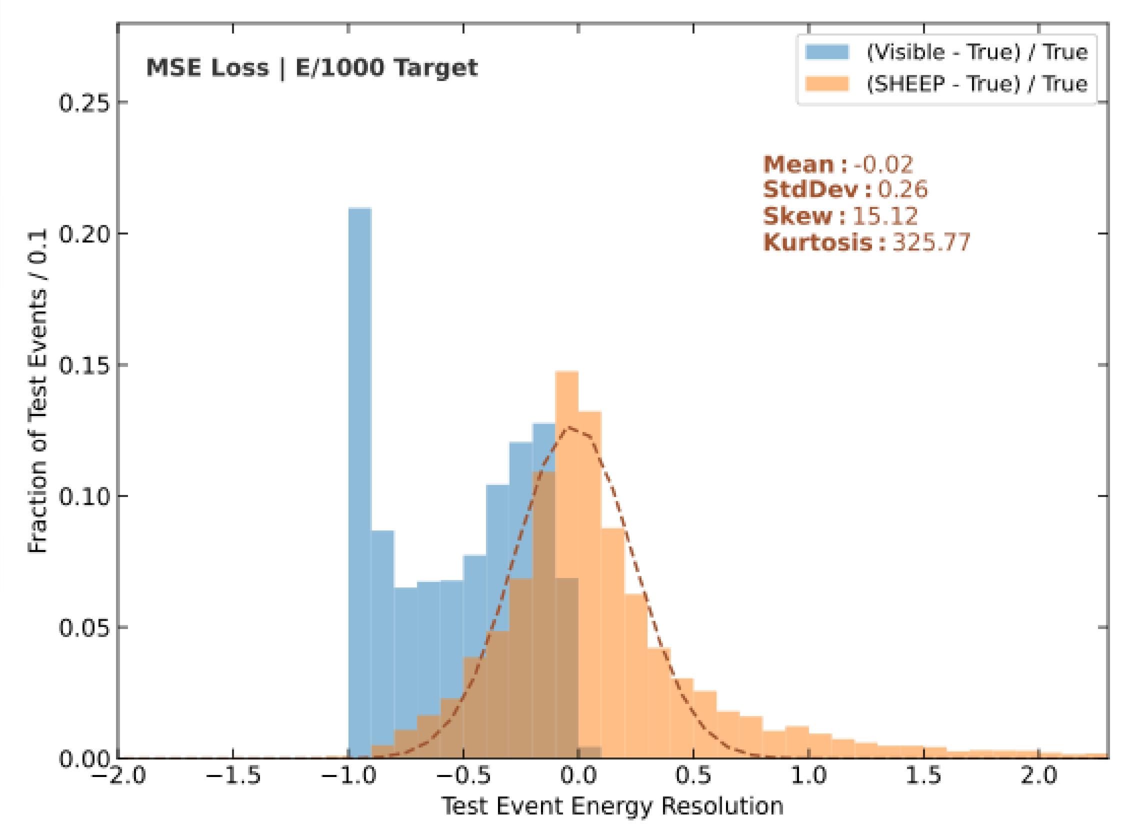
# MSE Loss, E/1000 Target vs. log(E) Target

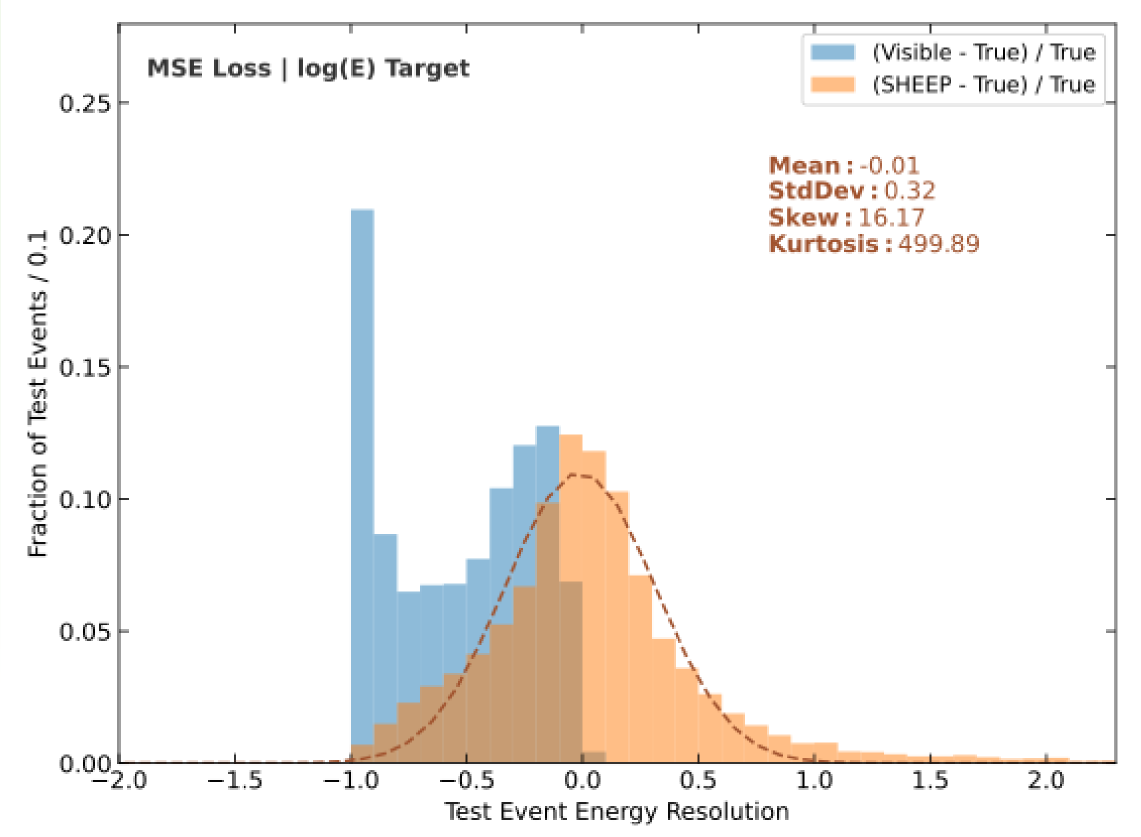


# MSE Loss, E/1000 Target vs. log(E) Target

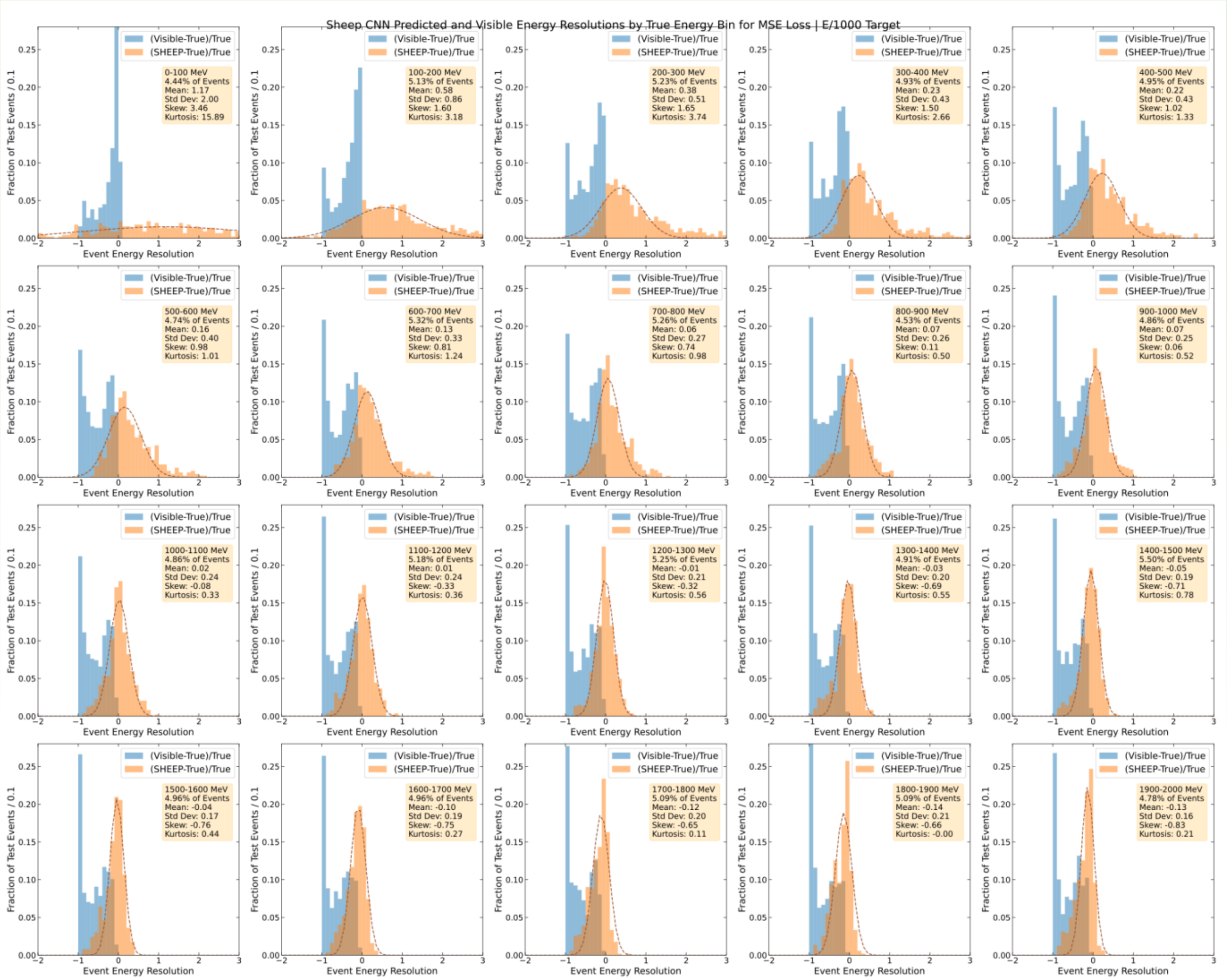


# MSE Loss, E/1000 Target vs. log(E) Target

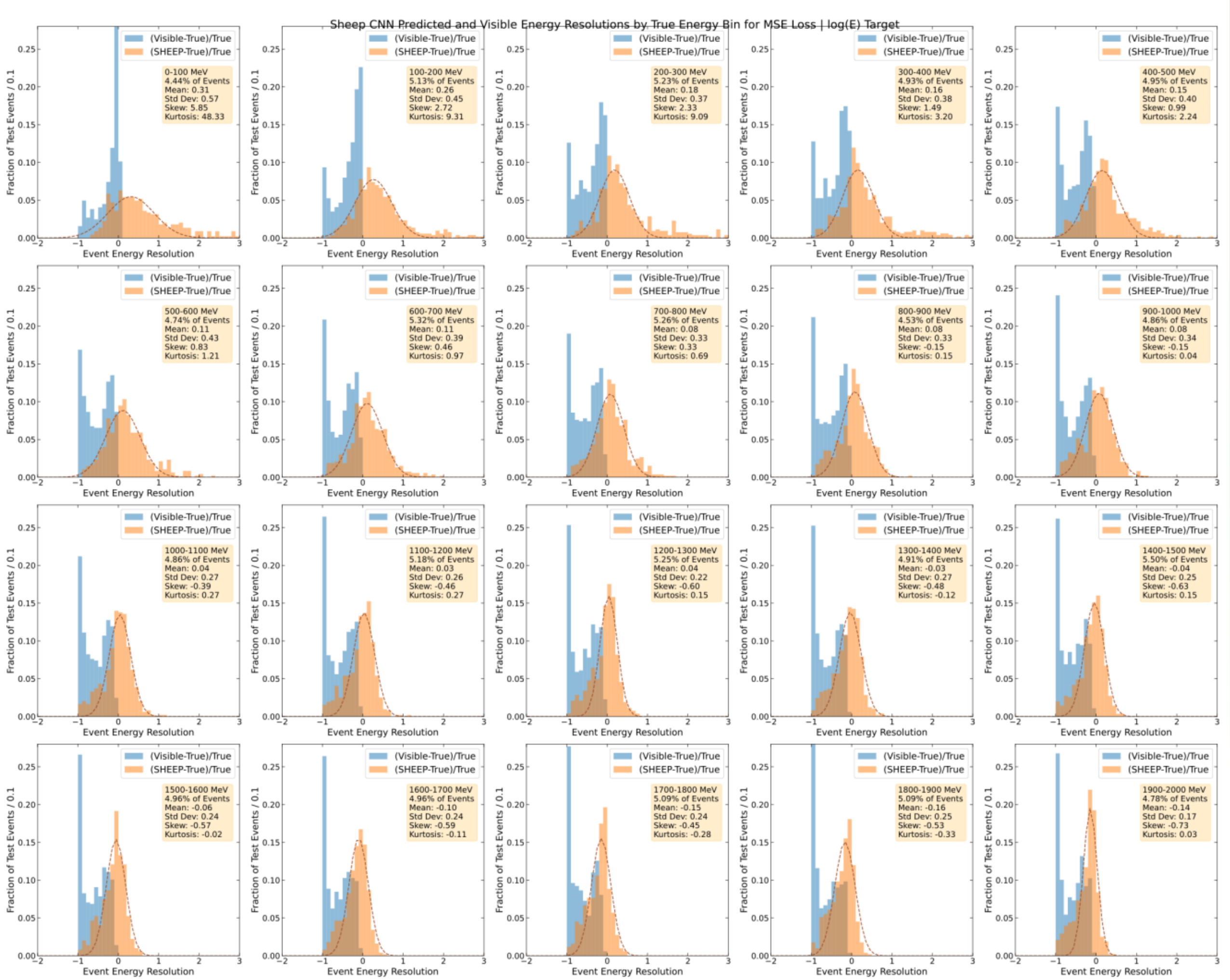


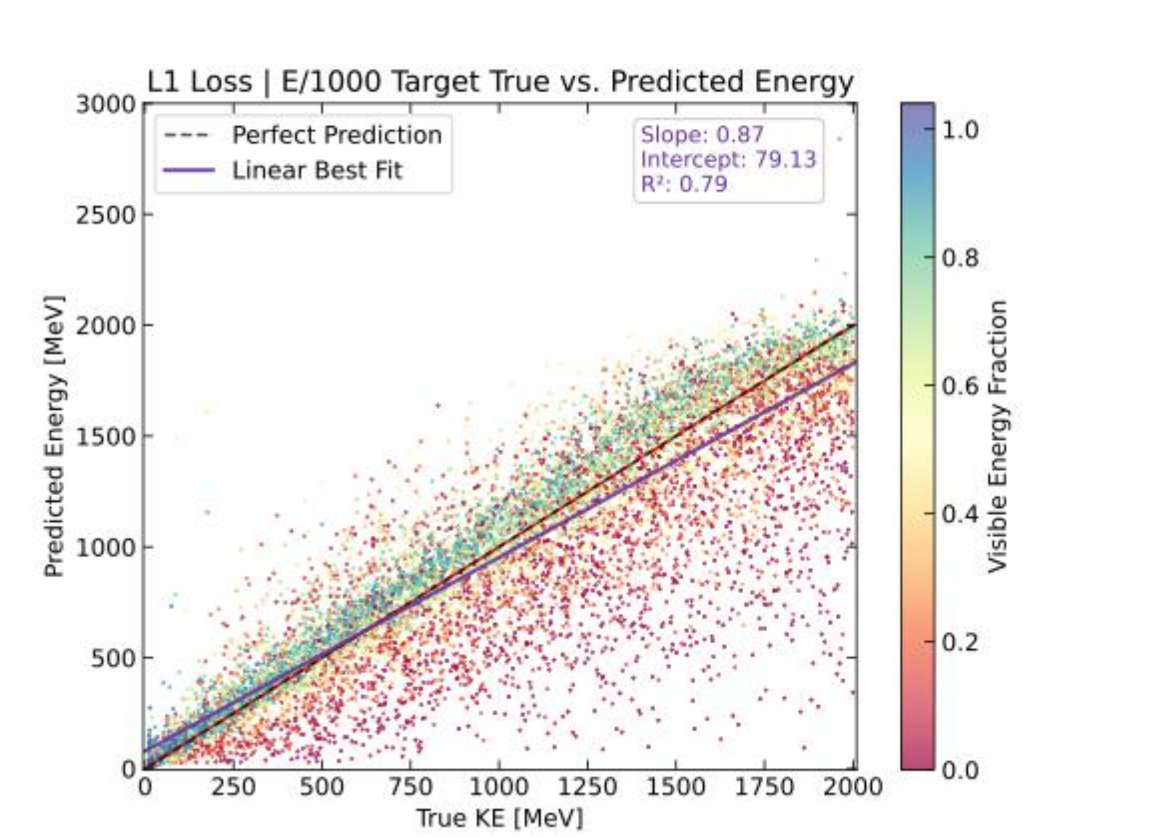
# MSE Loss, E/1000 Target vs. log(E) Target

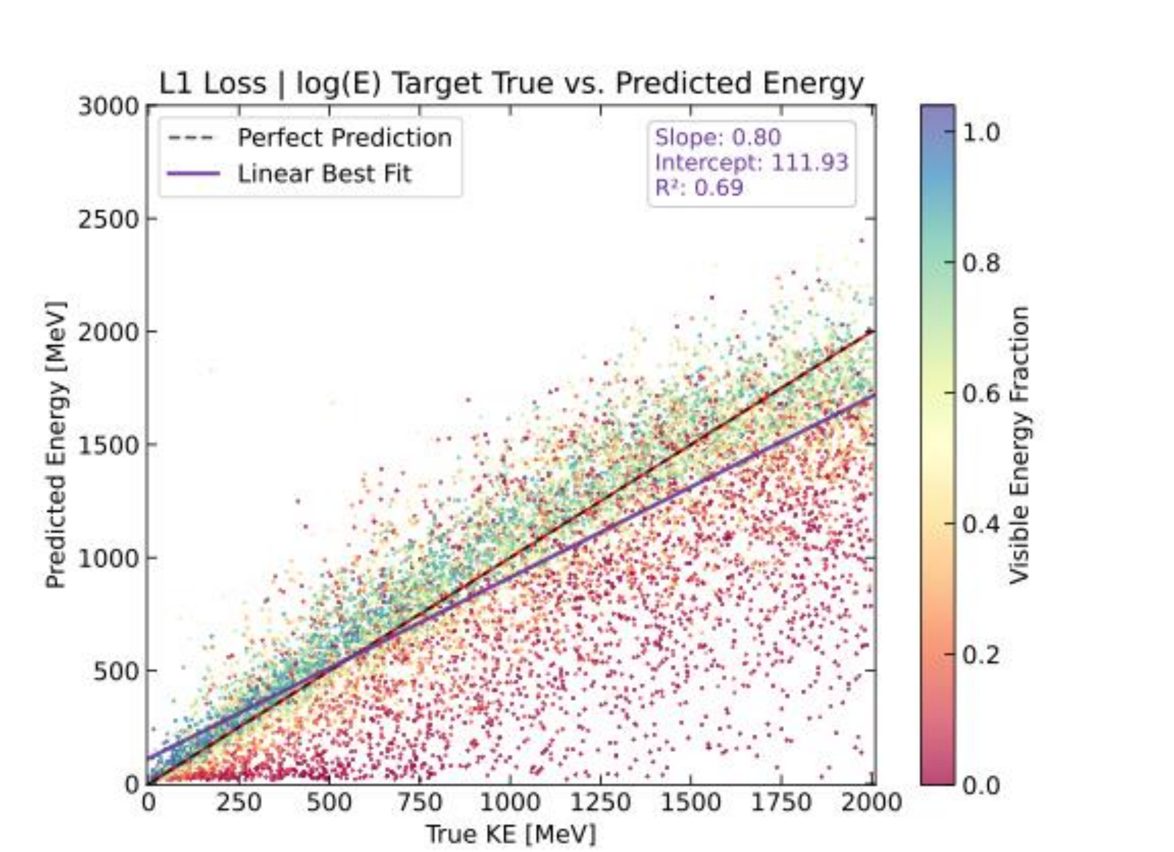


# MSE Loss, E/1000 Target vs. log(E) Target

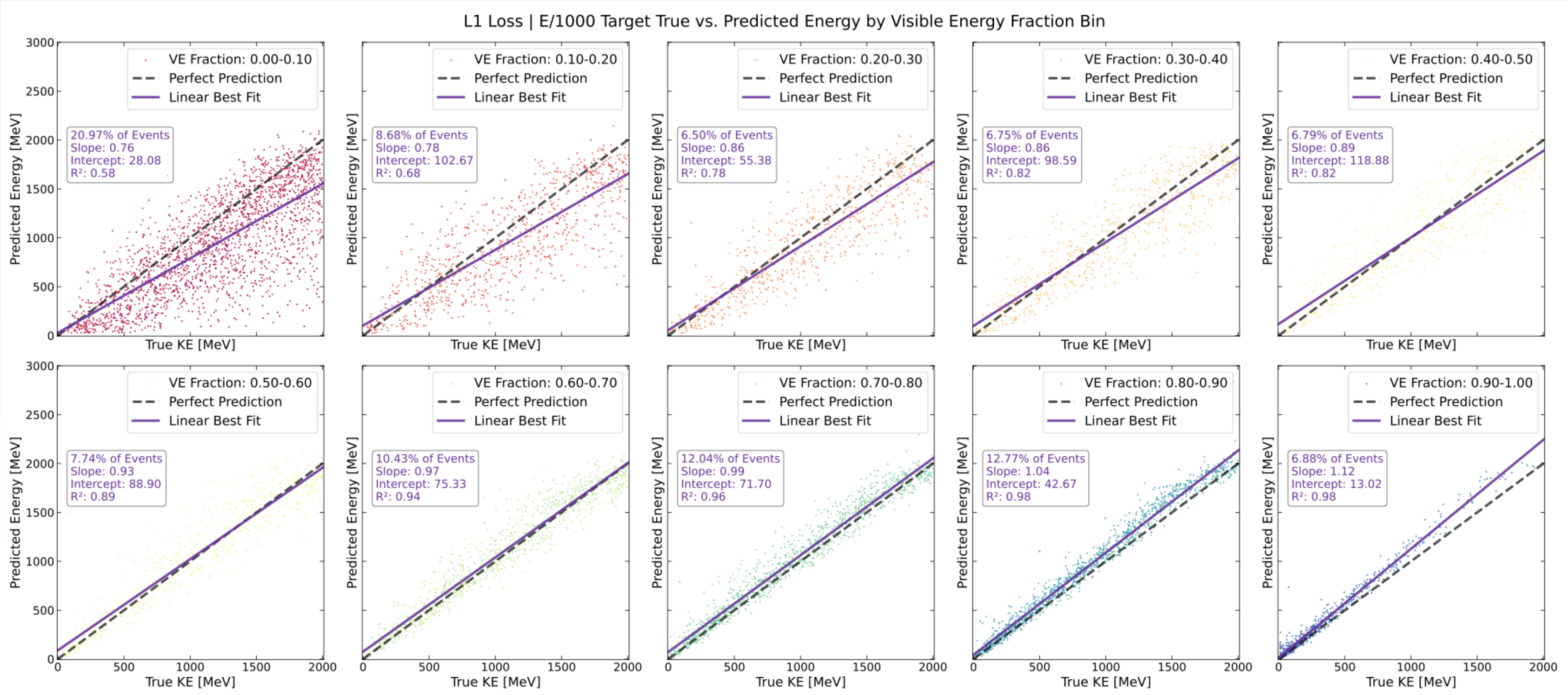


# L1 Loss, E/1000 Target vs. log(E) Target

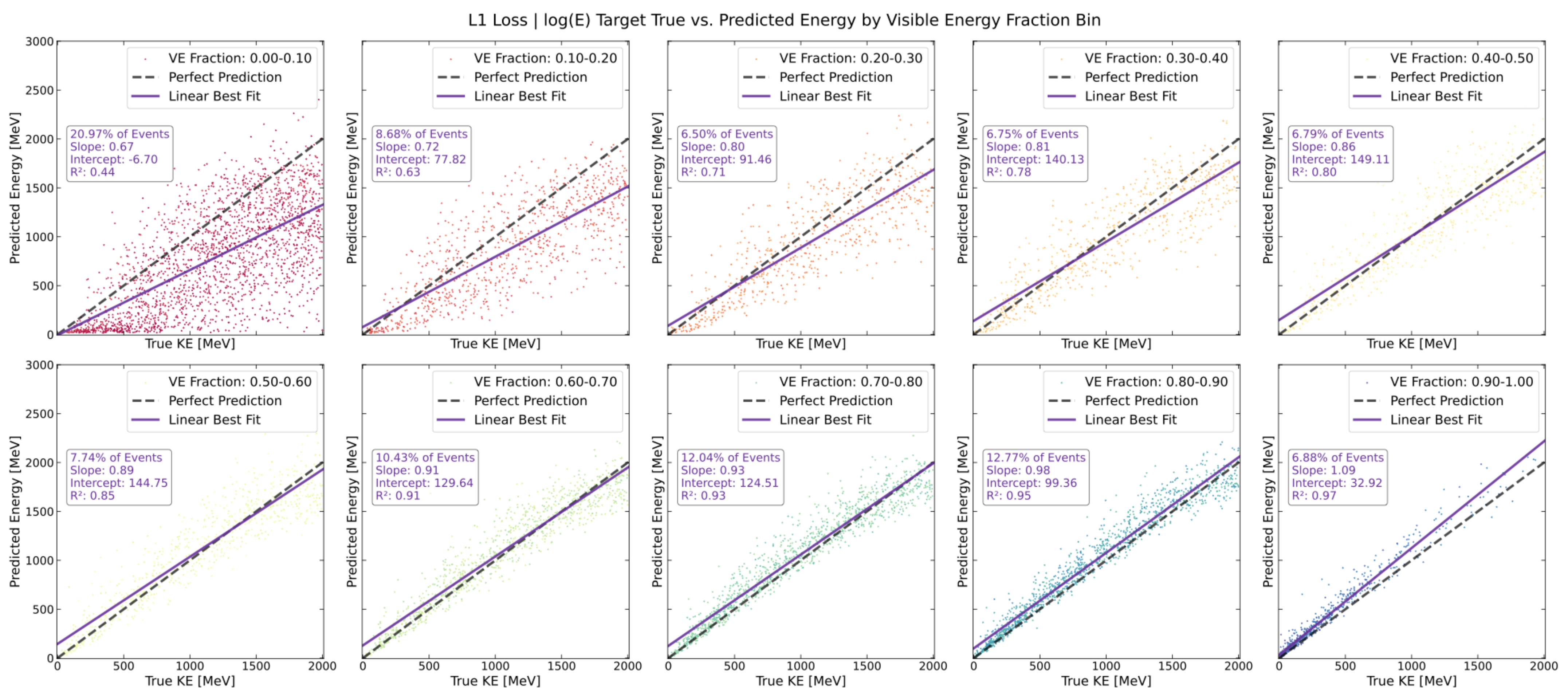


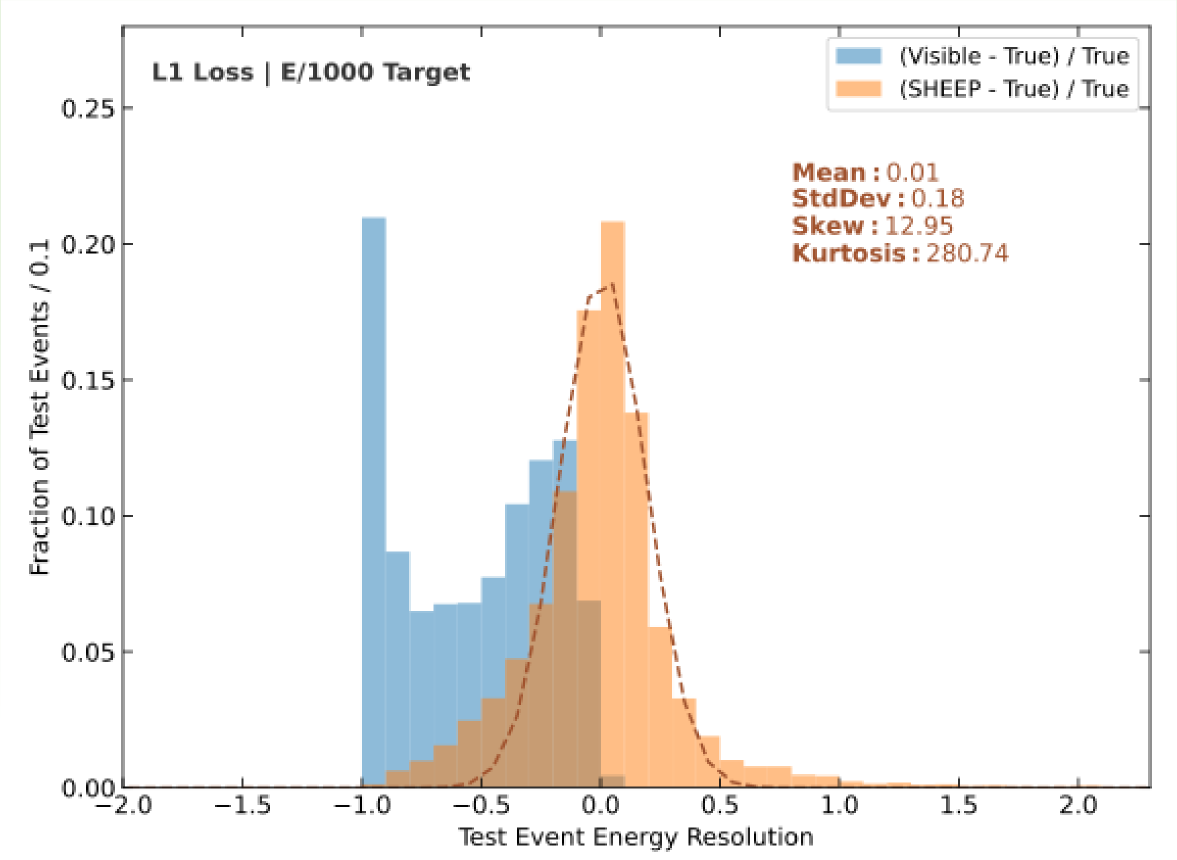
# L1 Loss, E/1000 Target vs. log(E) Target

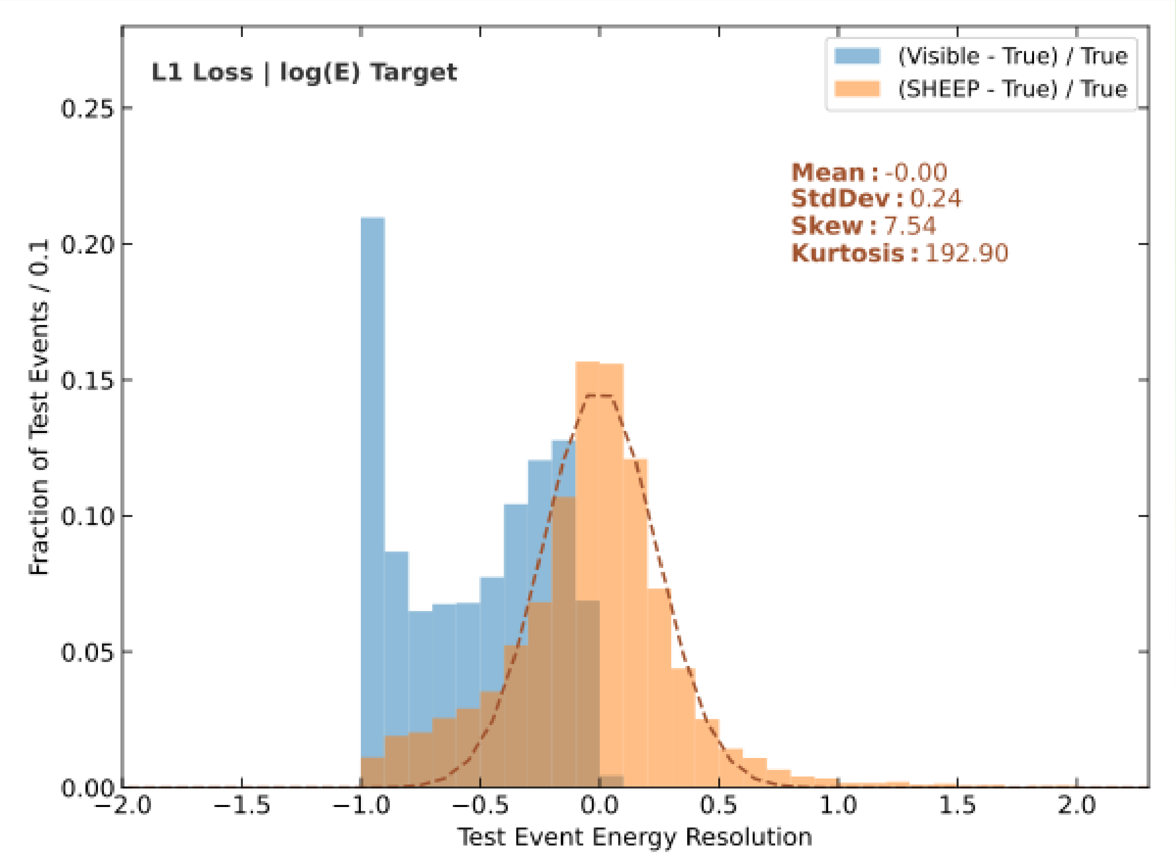


# L1 Loss, E/1000 Target vs. log(E) Target

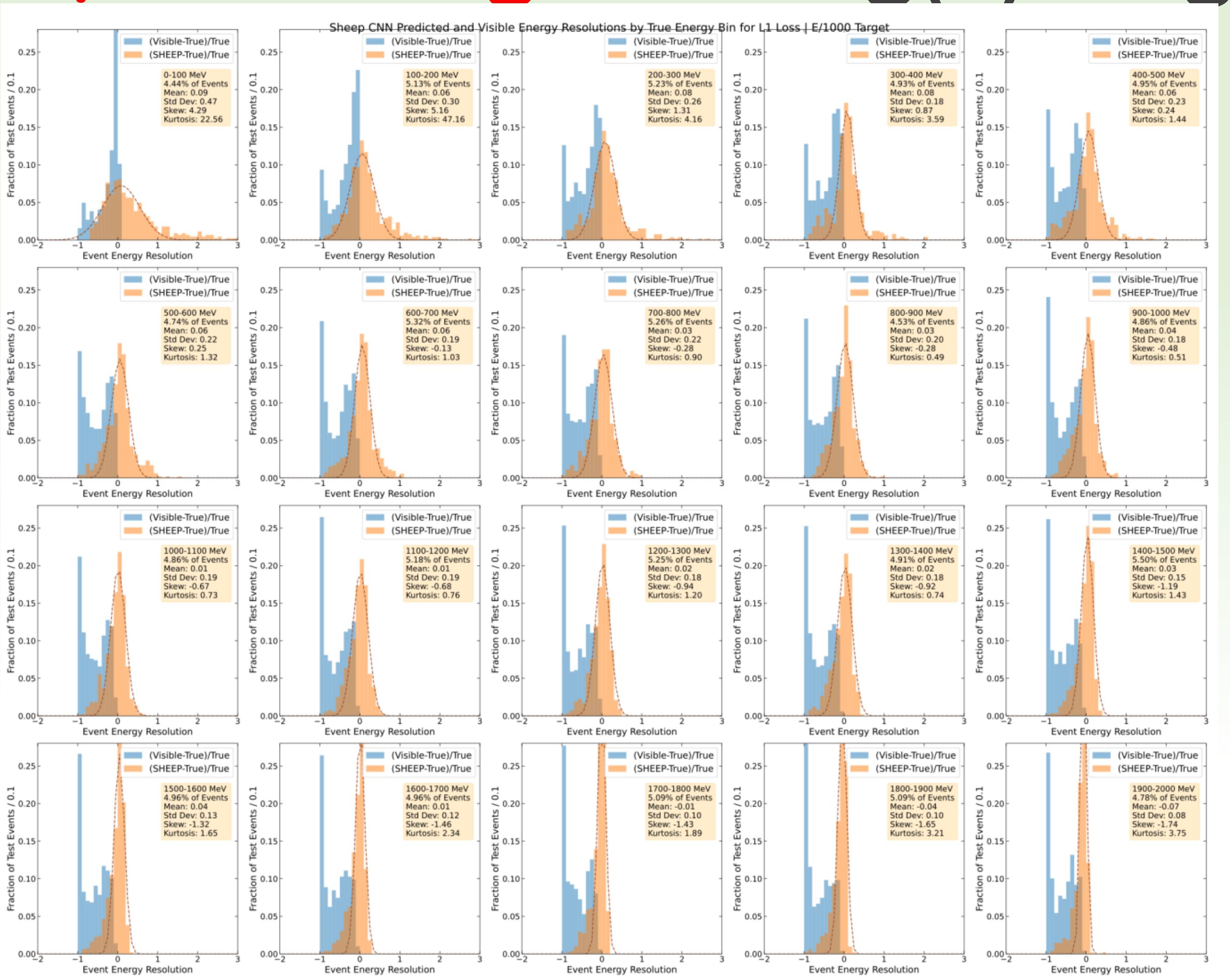


# L1 Loss, E/1000 Target vs. log(E) Target

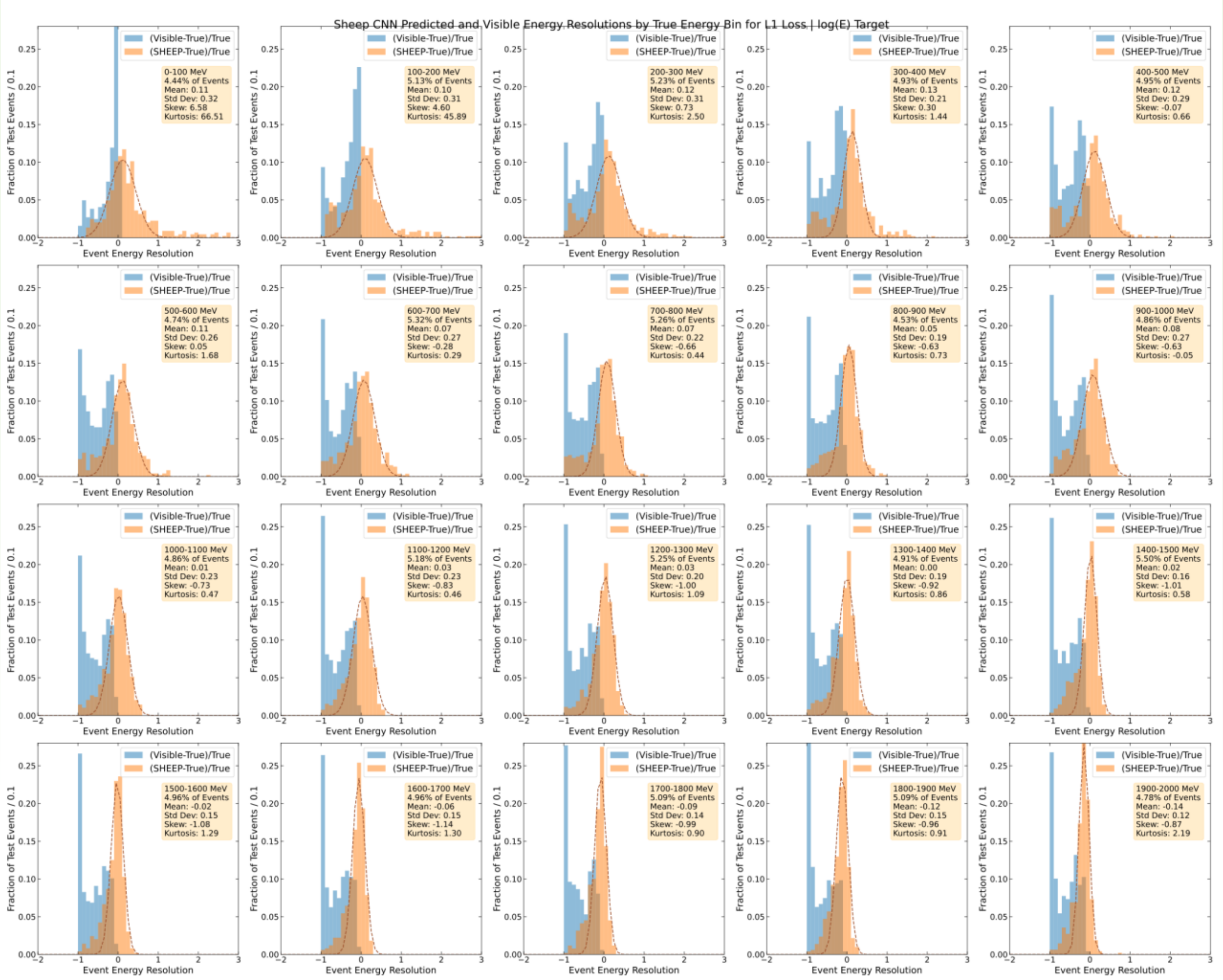


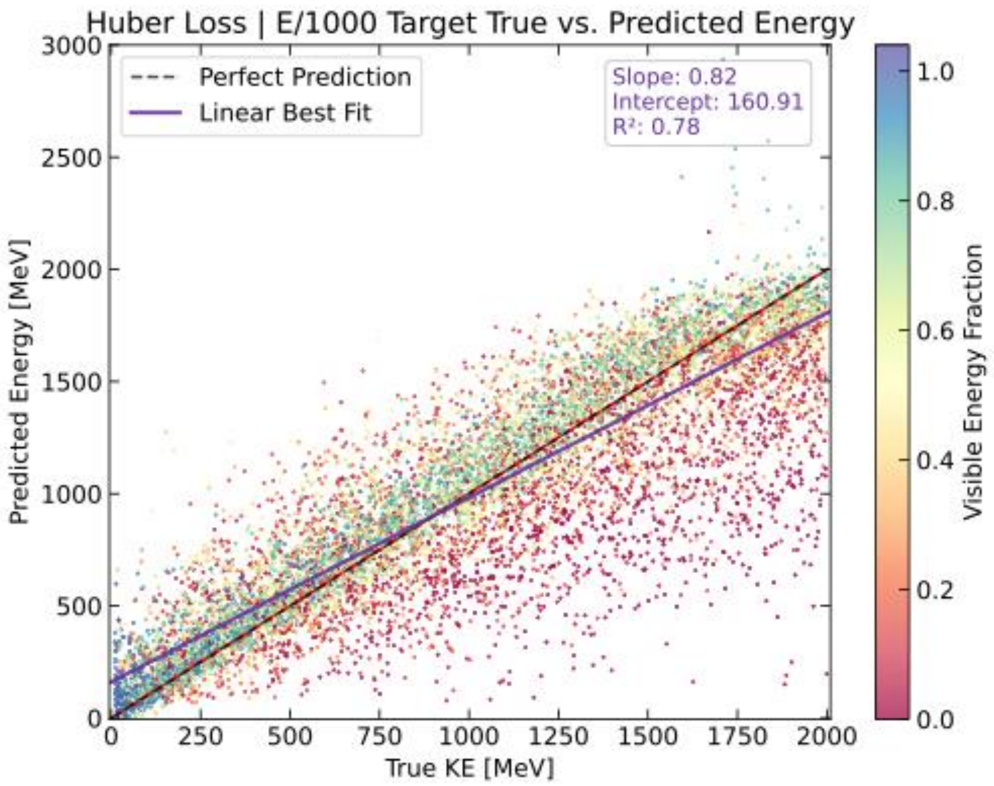
# L1 Loss, E/1000 Target vs. log(E) Target

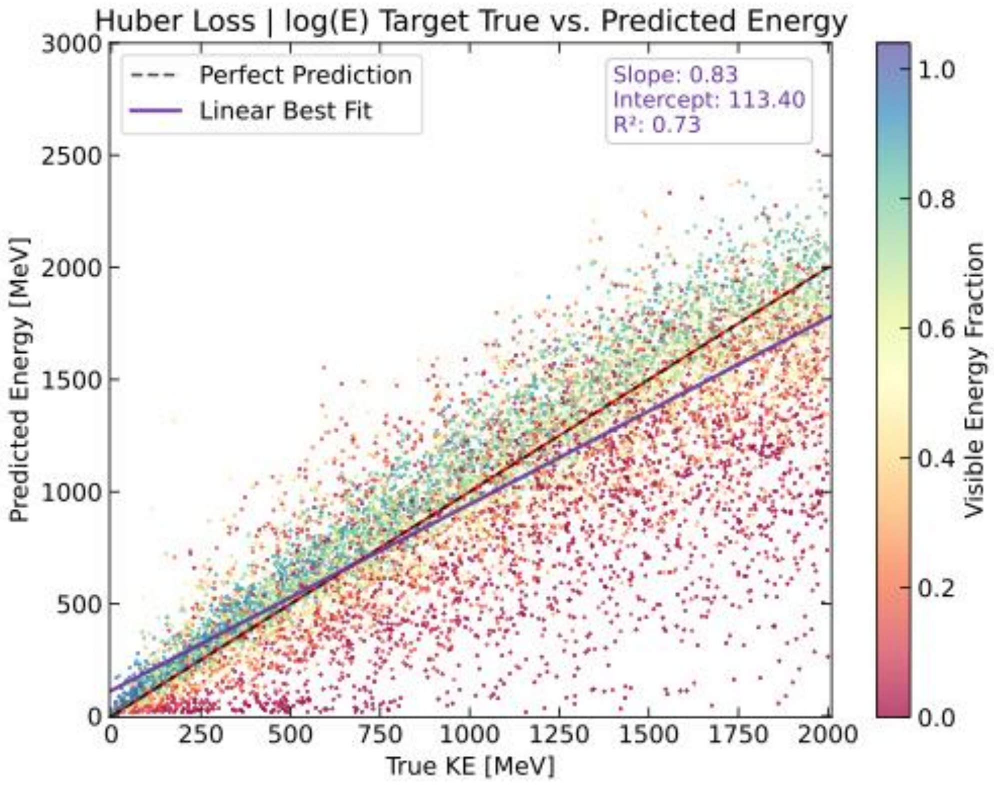


# L1 Loss, E/1000 Target vs. log(E) Target

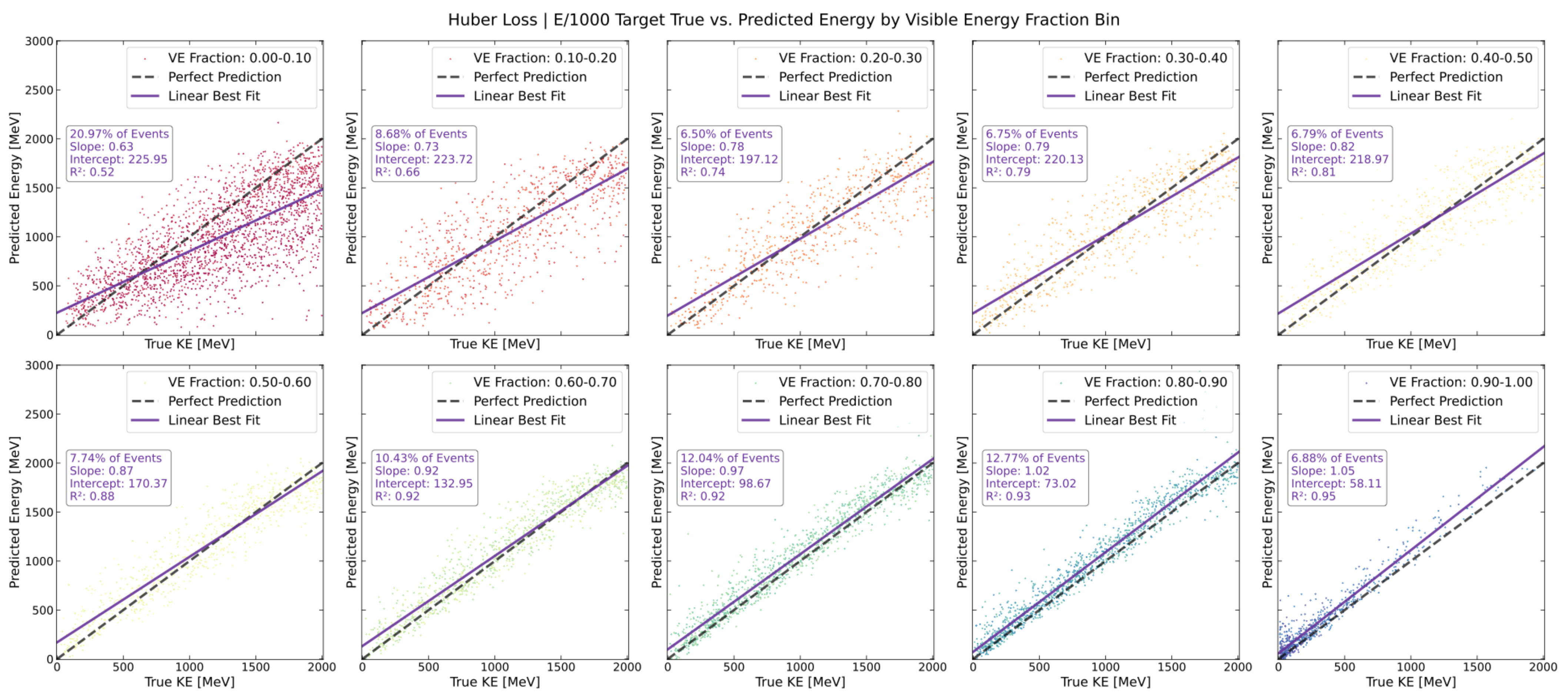


# Huber Loss, E/1000 Target vs. log(E) Target

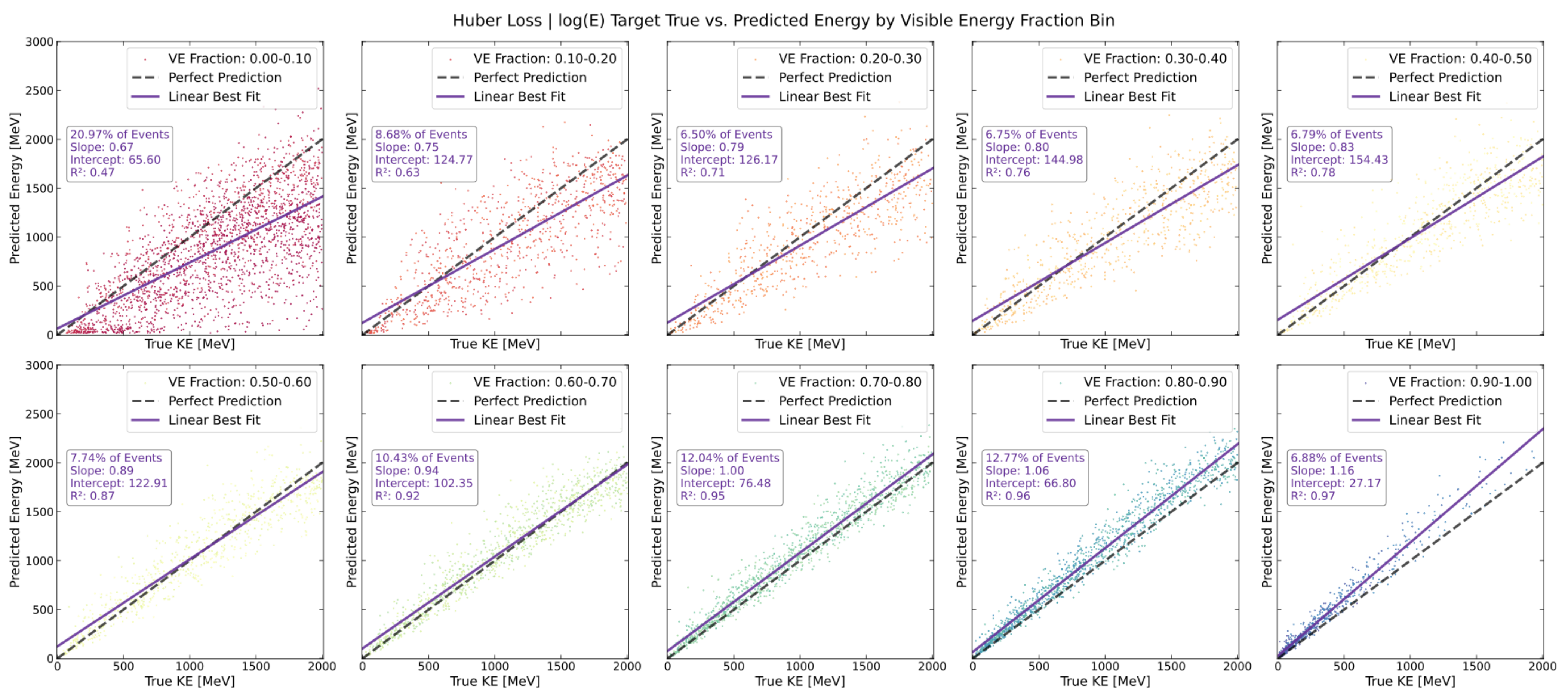


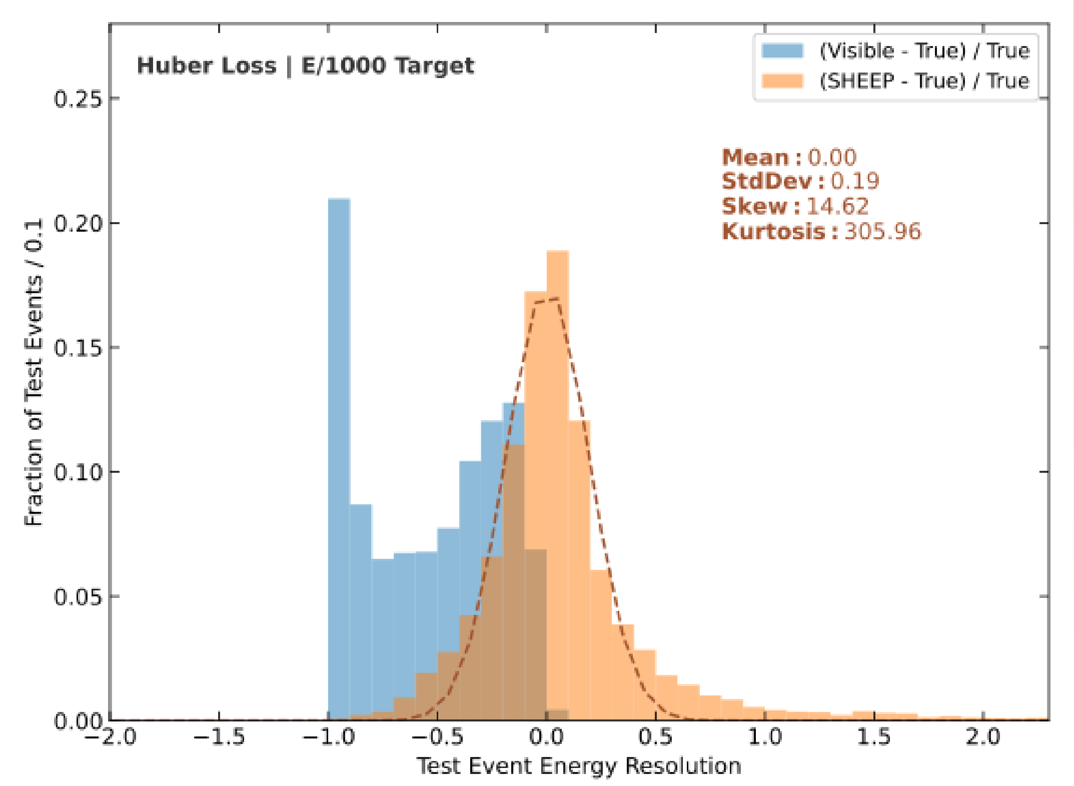
# Huber Loss, E/1000 Target vs. log(E) Target

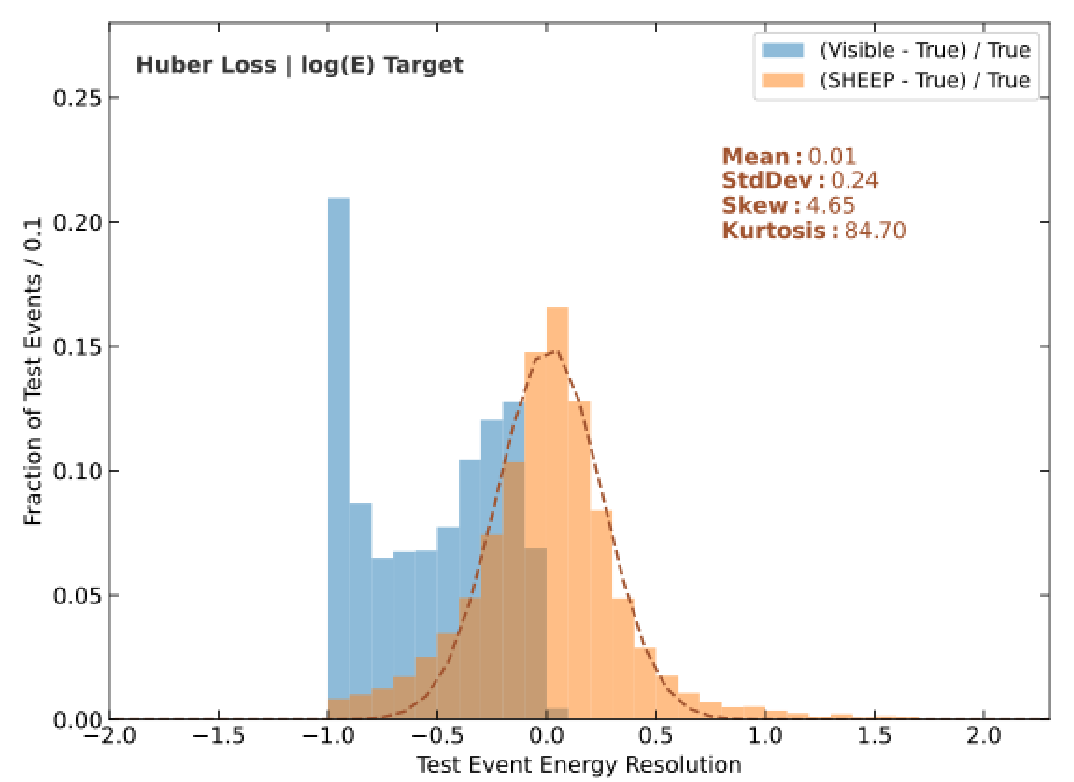


# Huber Loss, E/1000 Target vs. log(E) Target

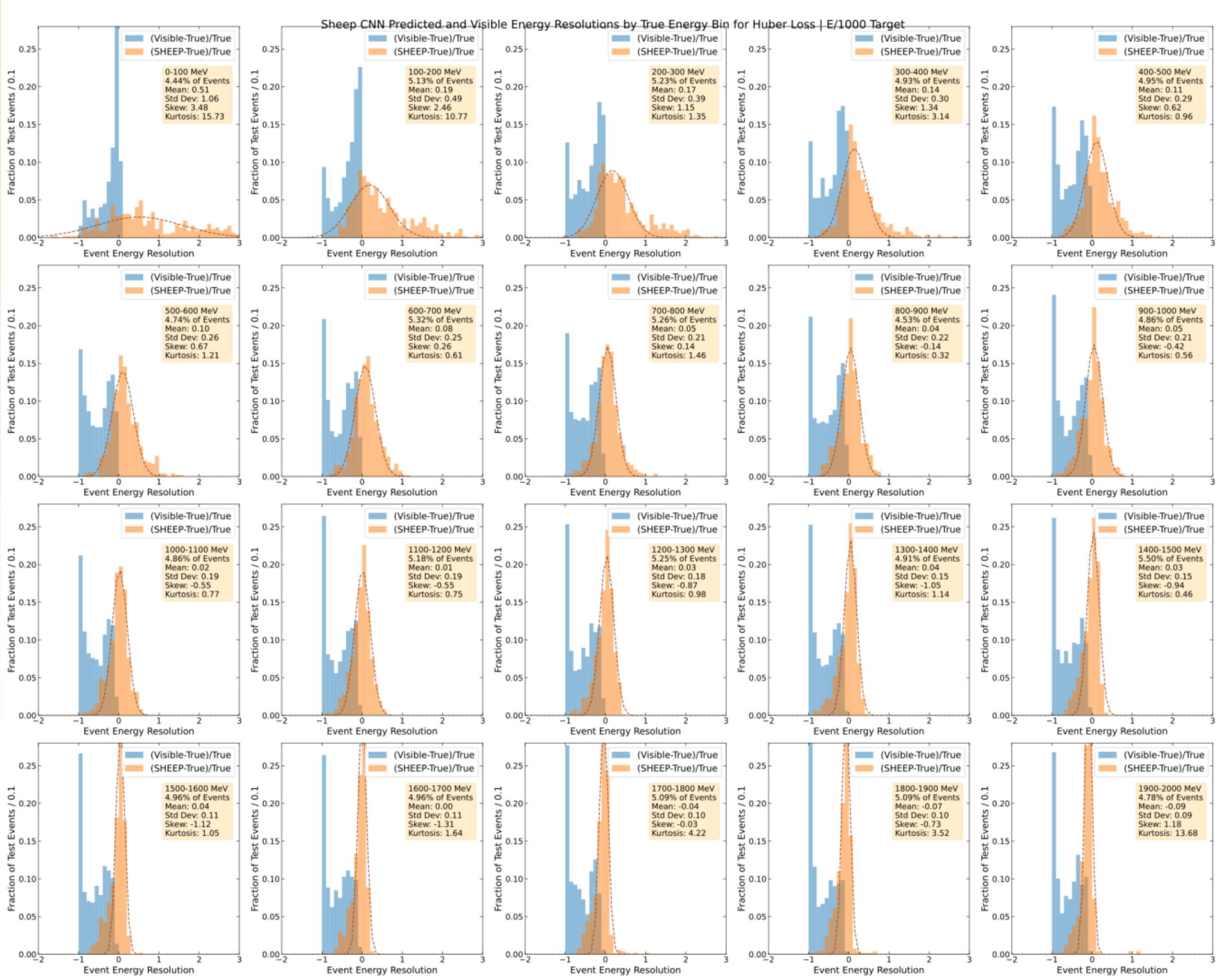


# Huber Loss, E/1000 Target vs. log(E) Target

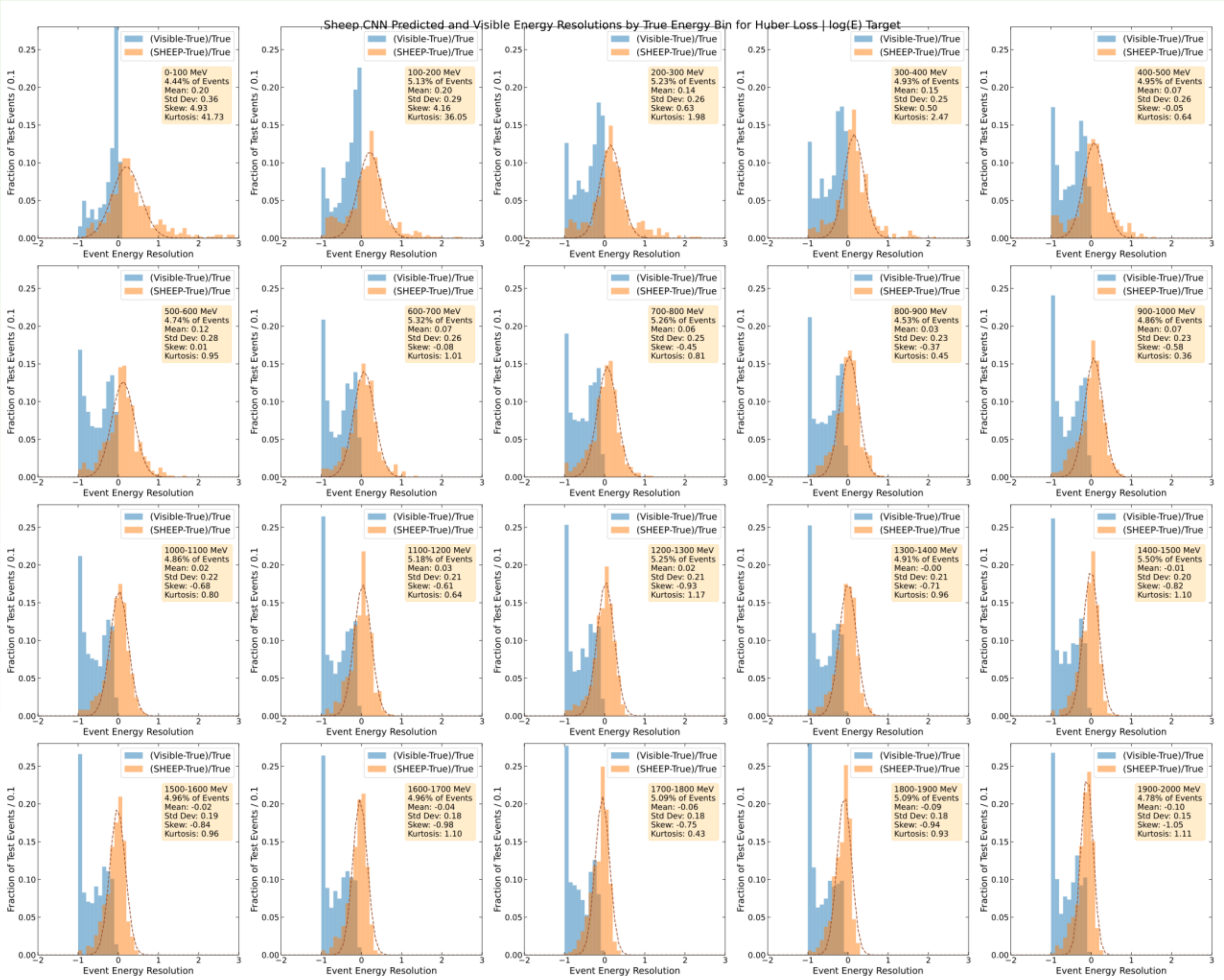


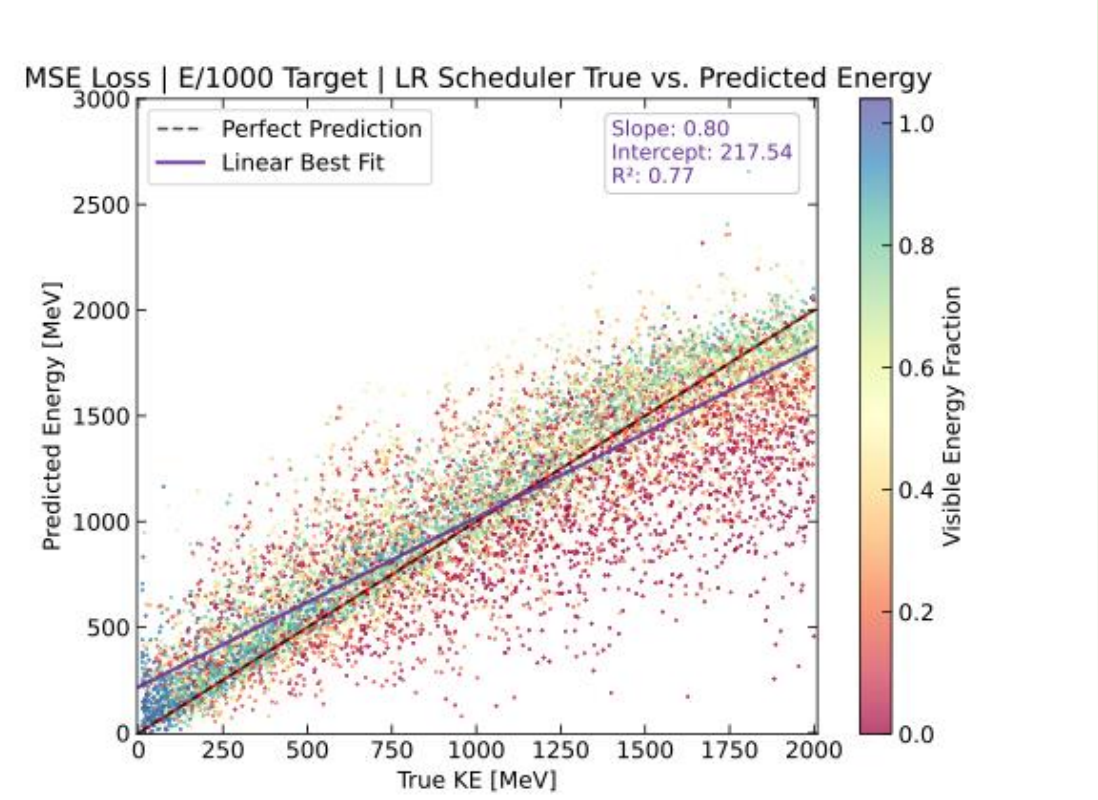
# Huber Loss, E/1000 Target vs. log(E) Target

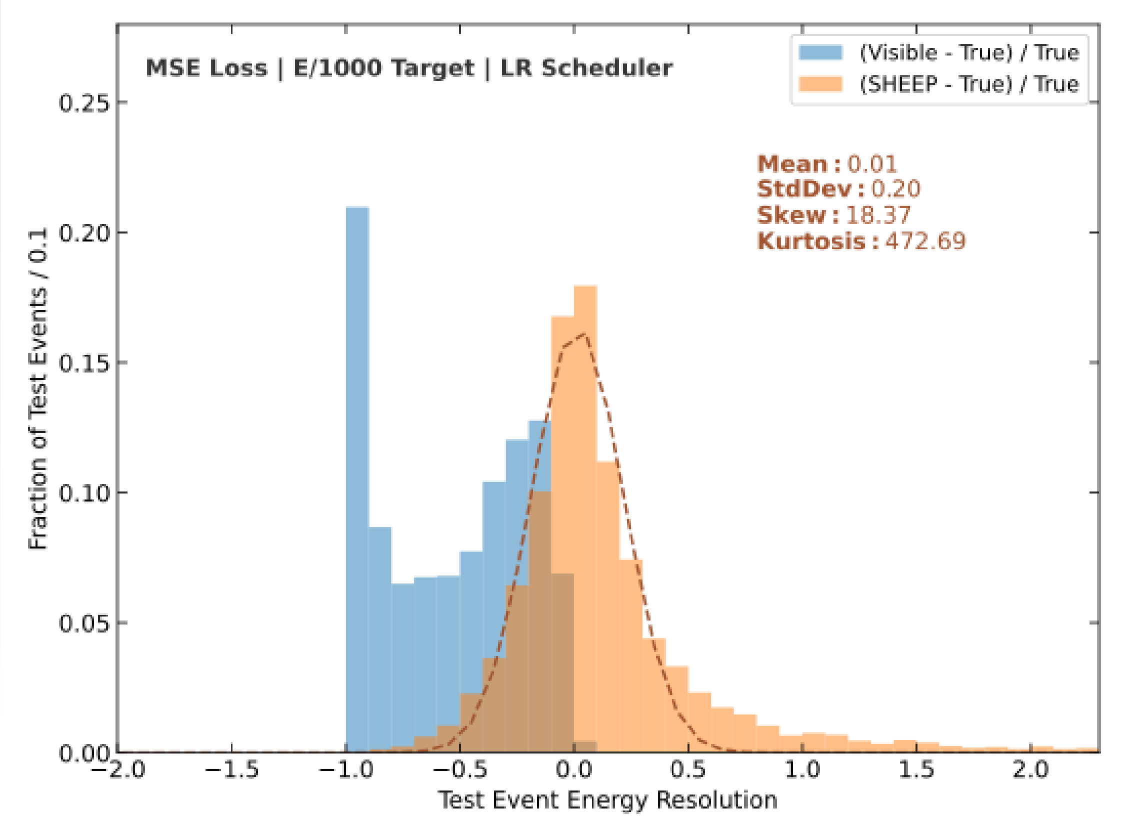


# Huber Loss, E/1000 Target vs. log(E) Target

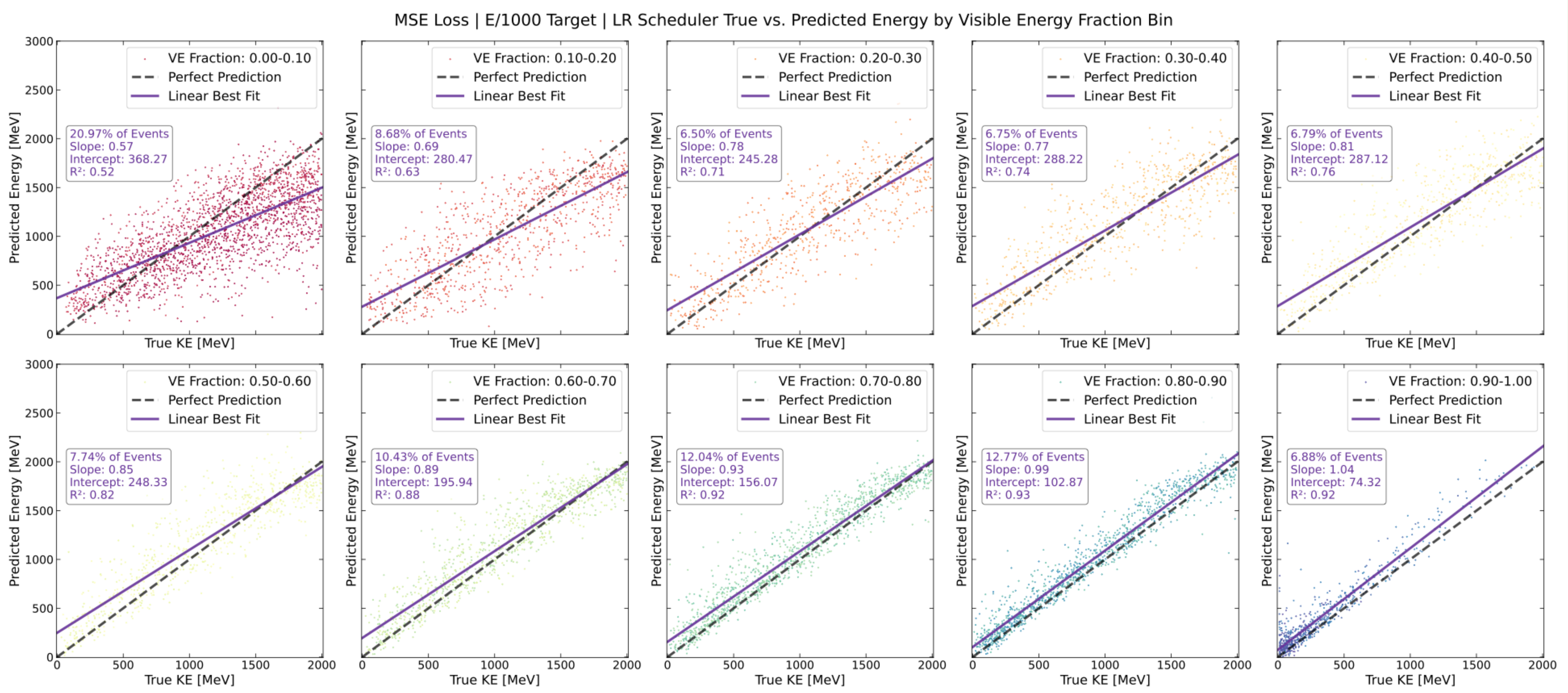


# MSE Loss, E/1000 Target w/ LR Increase >20 ep.

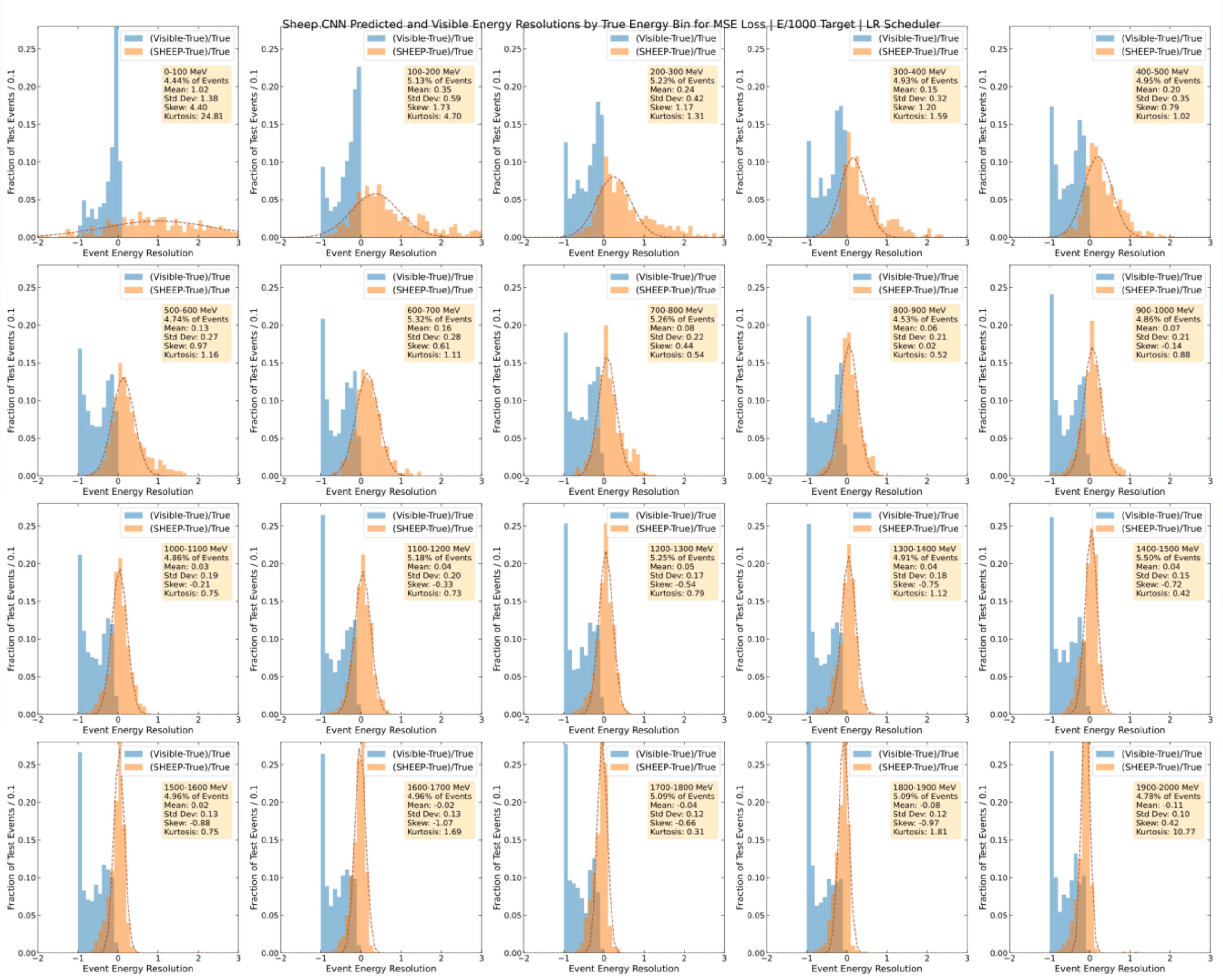


# MSE Loss, E/1000 Target w/ LR Increase >20 ep.

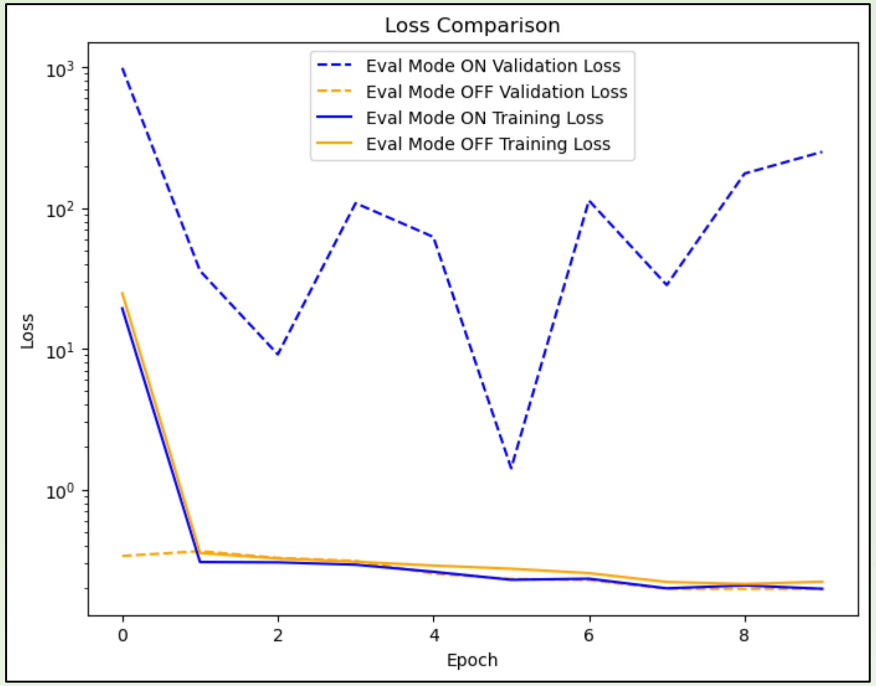


# MSE Loss, E/1000 Target w/ LR Increase >20 ep.



# Backup

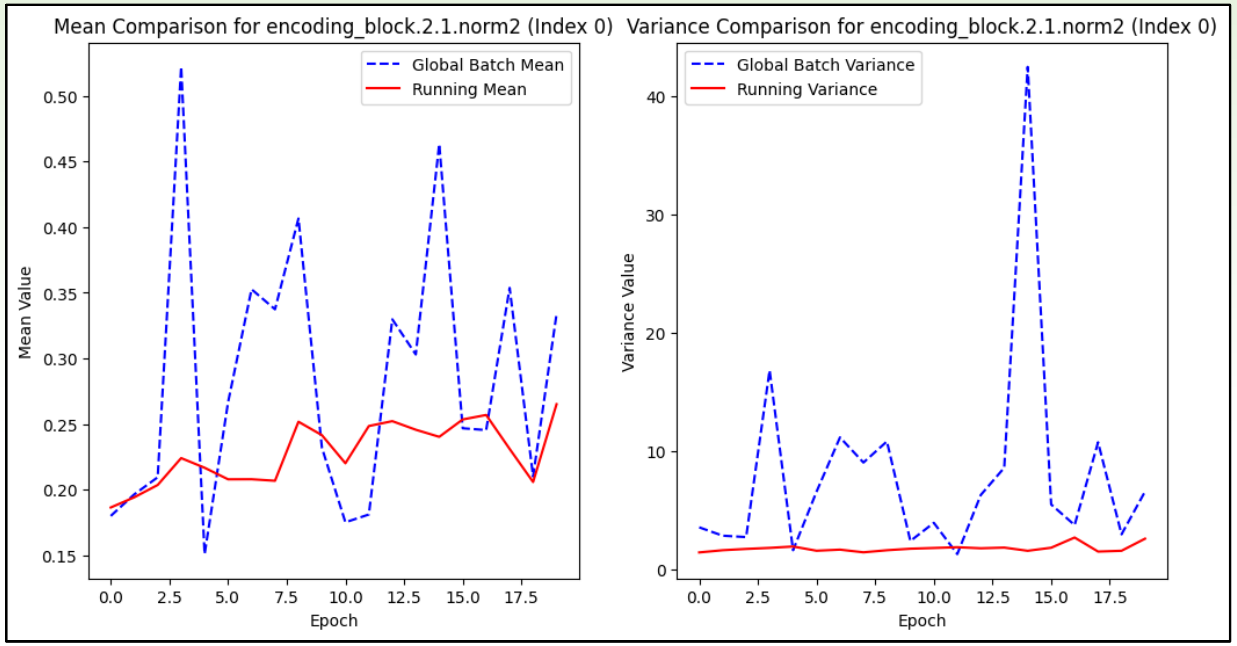
# SHEEP MODEL

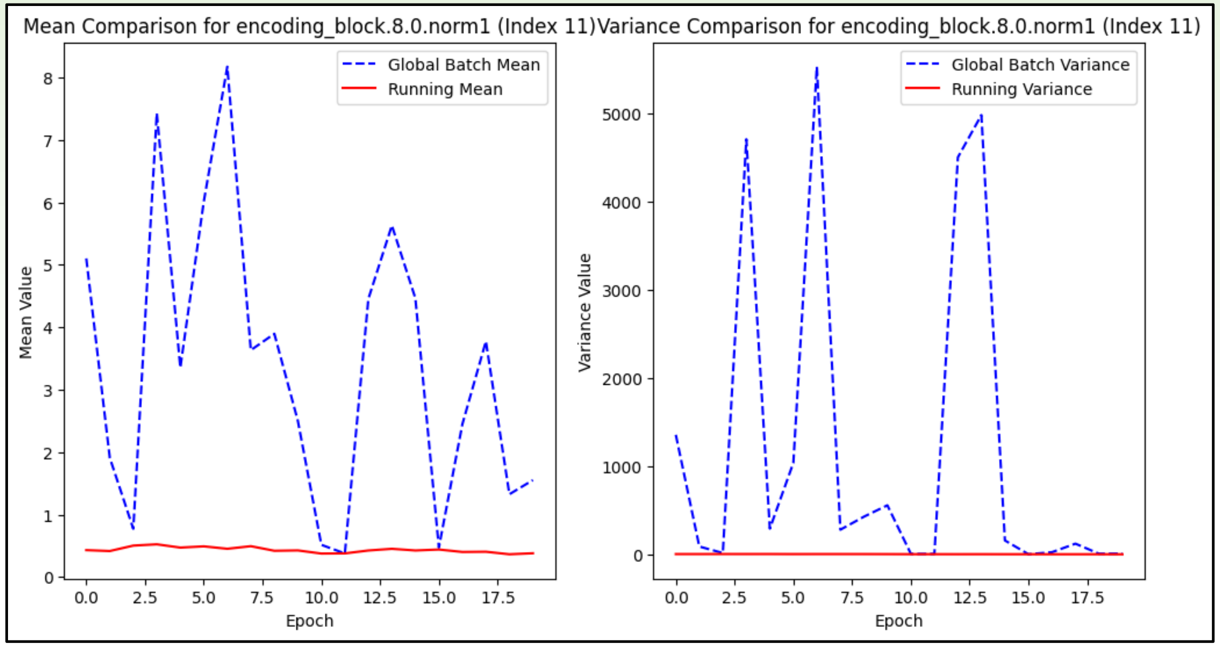



- Validation loss instability seems to be caused by difference in how **batch normalization layers** are treated in training vs. validation
  - Comes down to a normalization being done with running mean/variance vs. mean/variance calculated per batch (e.g. small sample vs. "full" sample)
- Fixed validation loss issue!

# Pre-Fix Batch vs. Population Stats

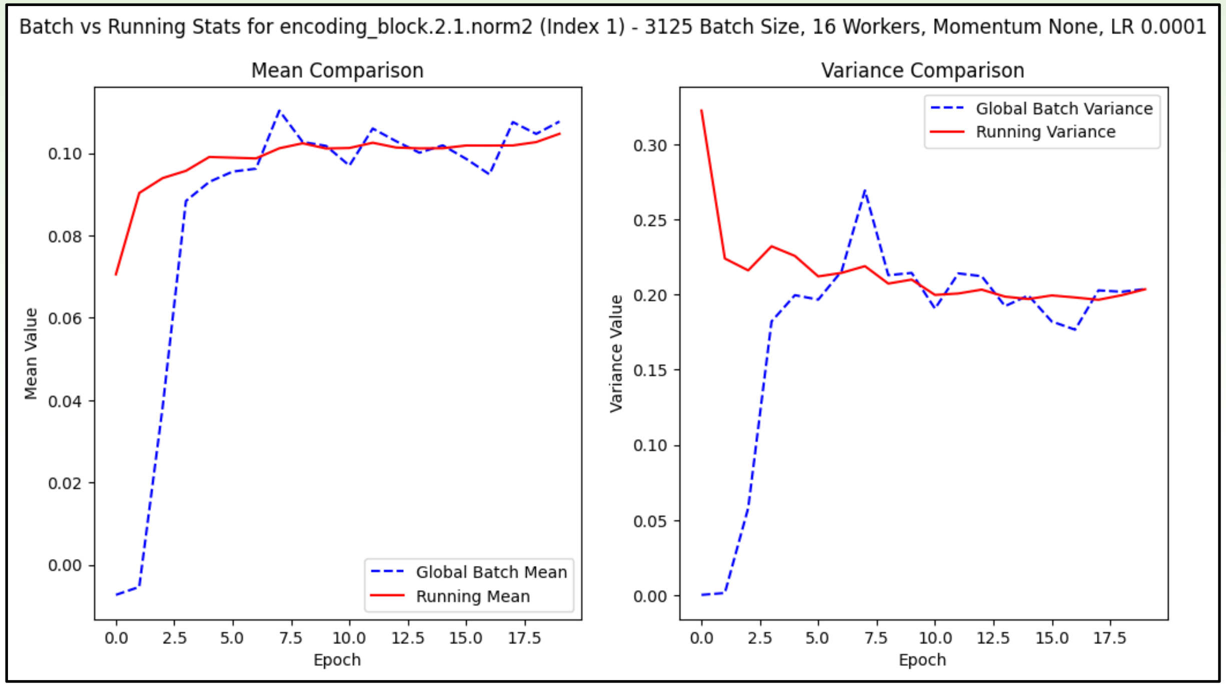
- Batch stats were quite different from population statistics (tried different variations of exponential moving average and simple moving average for population statistics)



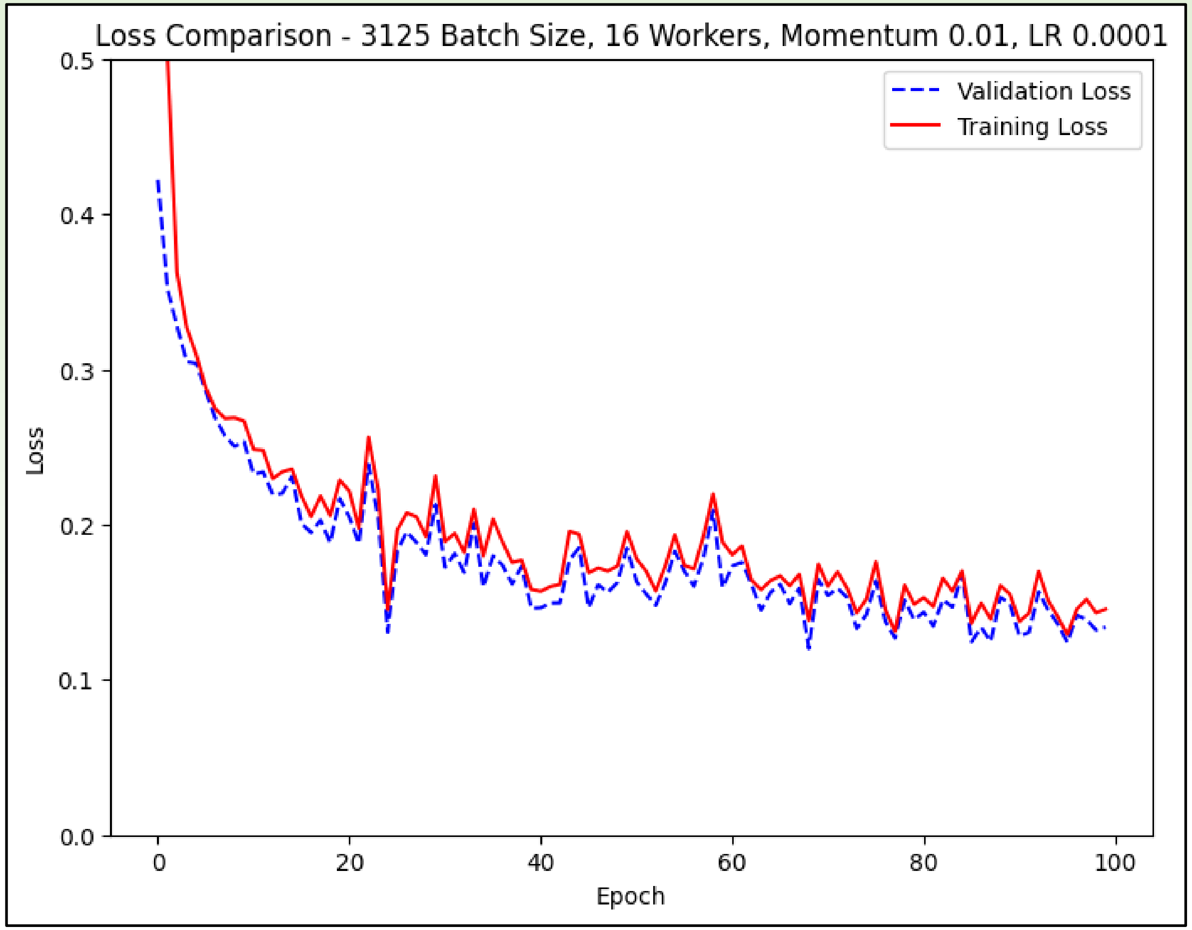
# Post-Fix Batch vs. Population Stats

- In addition to exp. moving avg momentum = 0.01, fix involves:
  - Batch size 200 → ~3000
  - Learning rate 0.001 → 0.0001

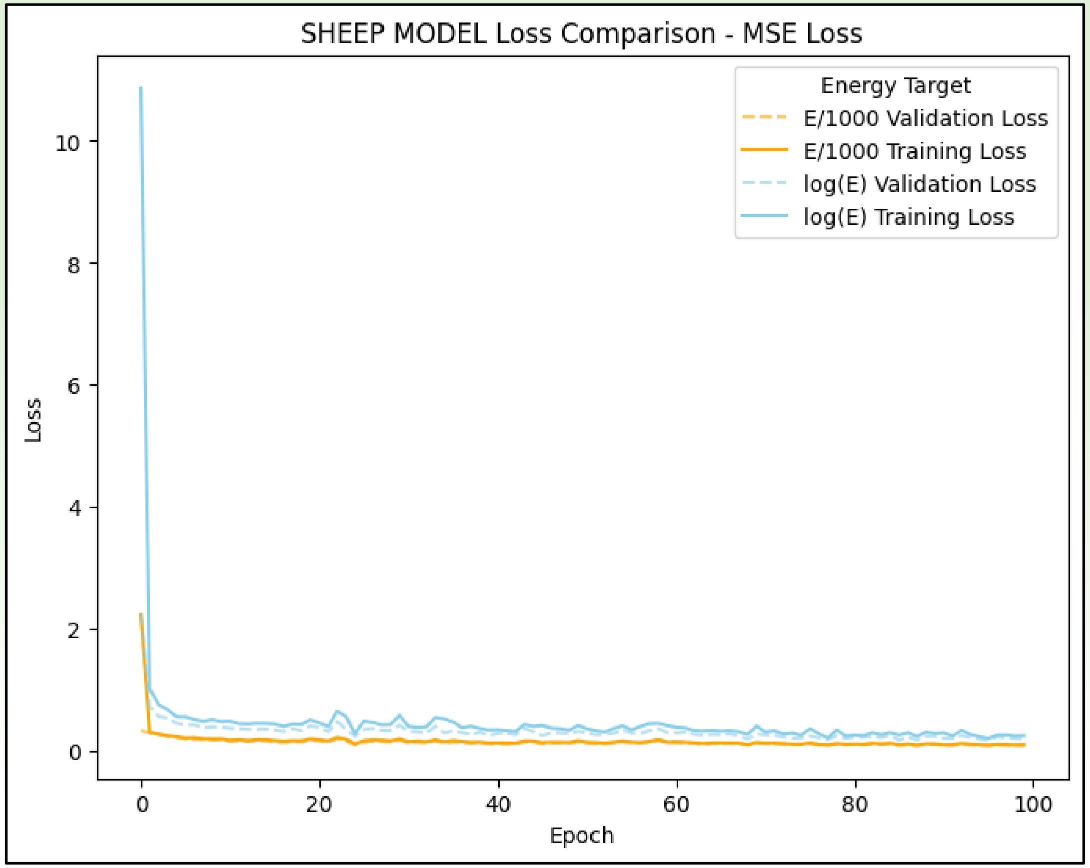


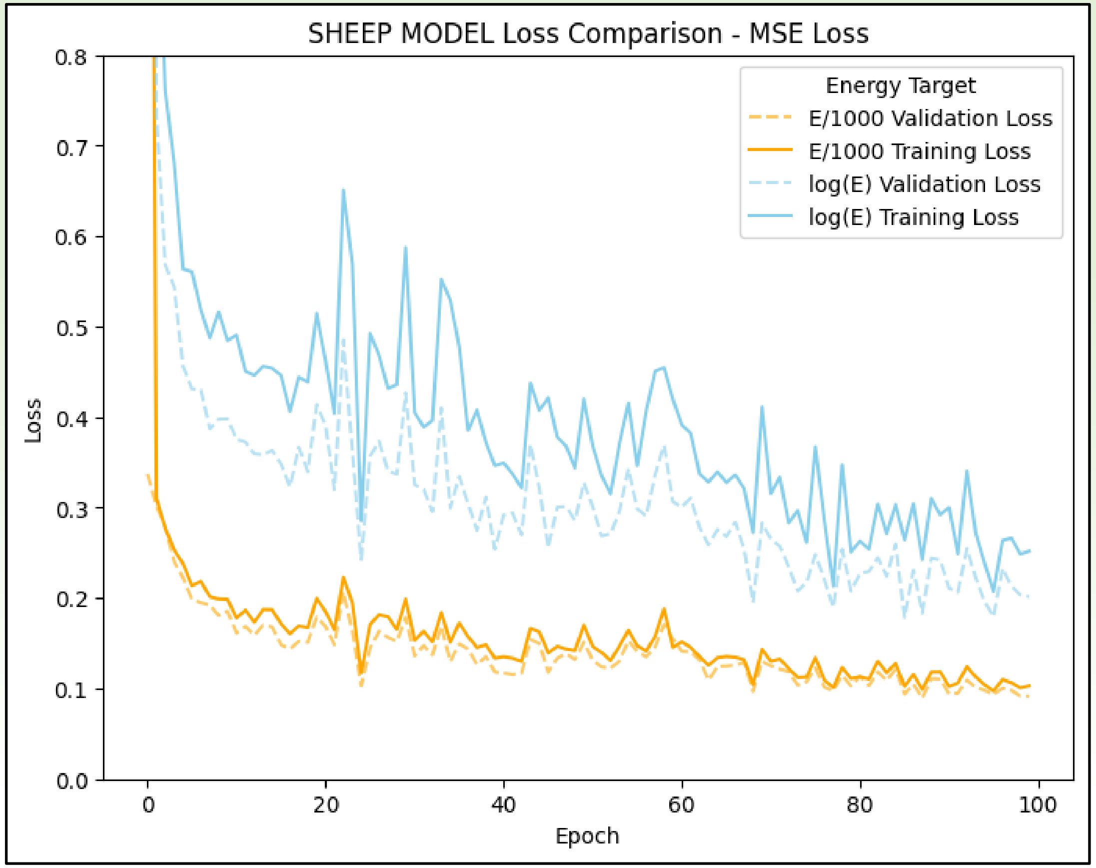

# Post-Fix Loss Curves – Initial Test

- Testing on smaller dataset
  - 50k training events (6125 each for validation and testing)
- Loss function = Mean Squared Error (MSE)
- Target value = True Kinetic Energy / 1000 (i.e. convert to GeV)



# Post-Fix Loss Curves – MSE Loss



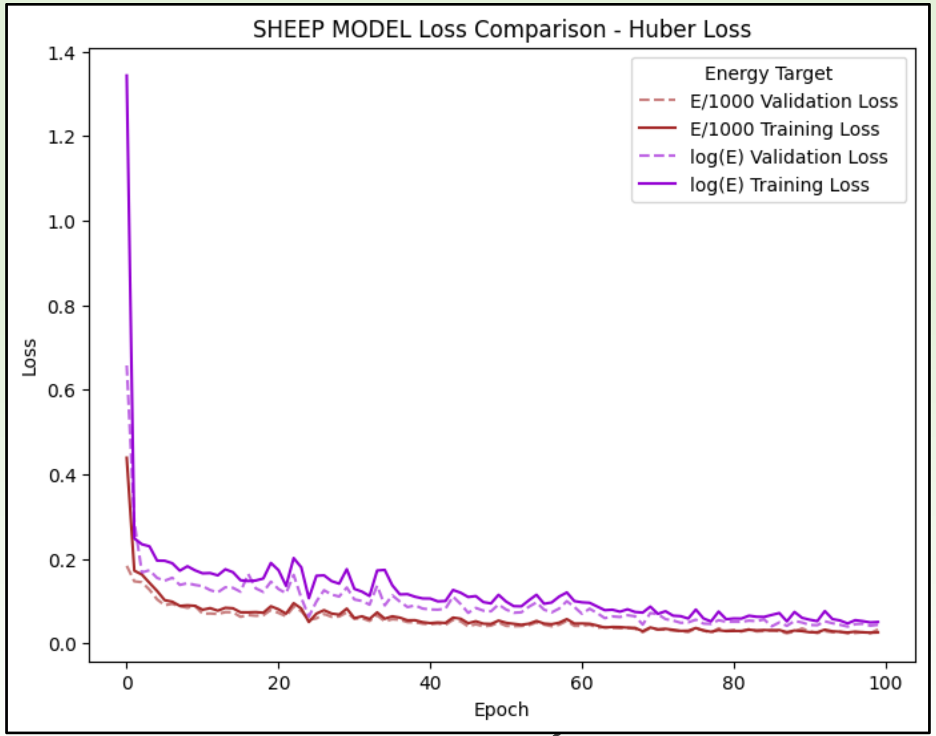

- **MSE Loss:** $\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2$ ; $\boldsymbol{y_i}$= true KE, $\boldsymbol{\hat{y}_i}$ = SHEEP prediction

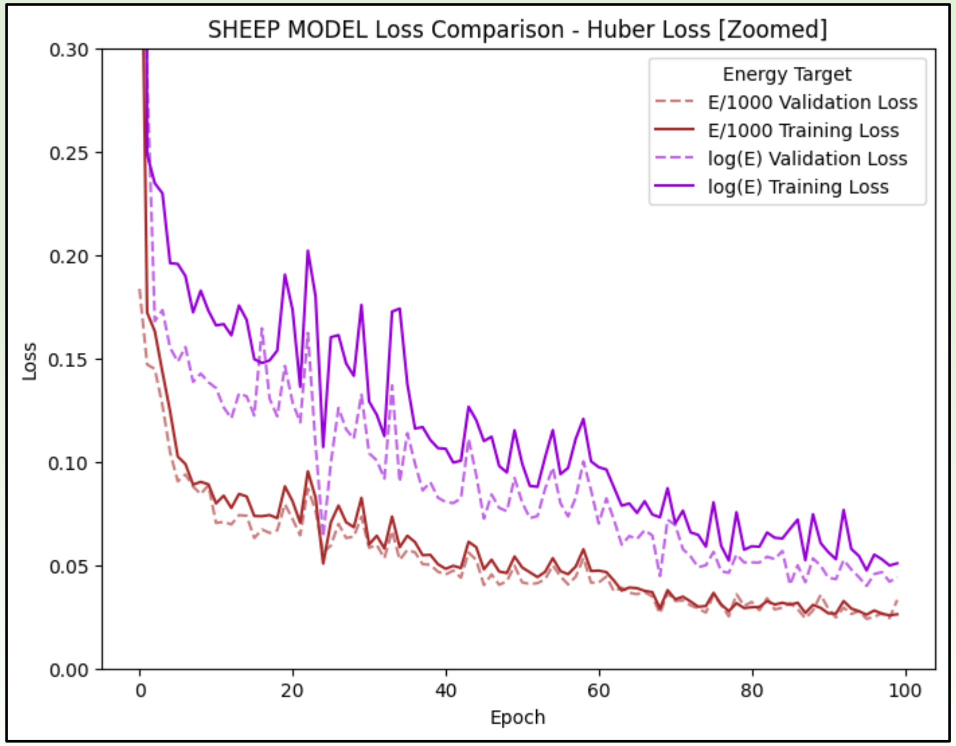
# Post-Fix Loss Curves – L1 Loss



- **L1 Loss:** $\frac{1}{N}\sum_{i=1}^{N} |\,\hat{y}_i - y_i\,|$ ; $\boldsymbol{y_i}$= true KE, $\boldsymbol{\hat{y}_i}$ = SHEEP prediction

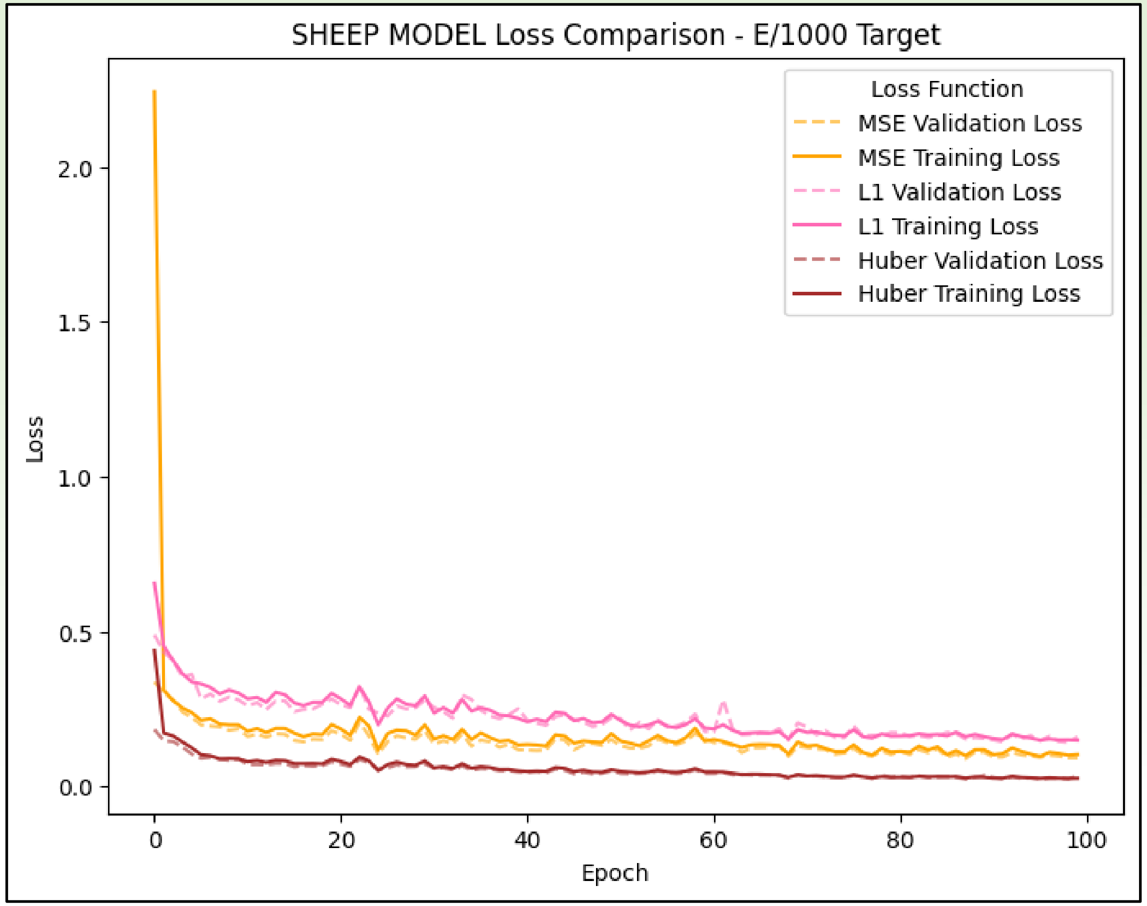
# Post-Fix Loss Curves – Huber Loss

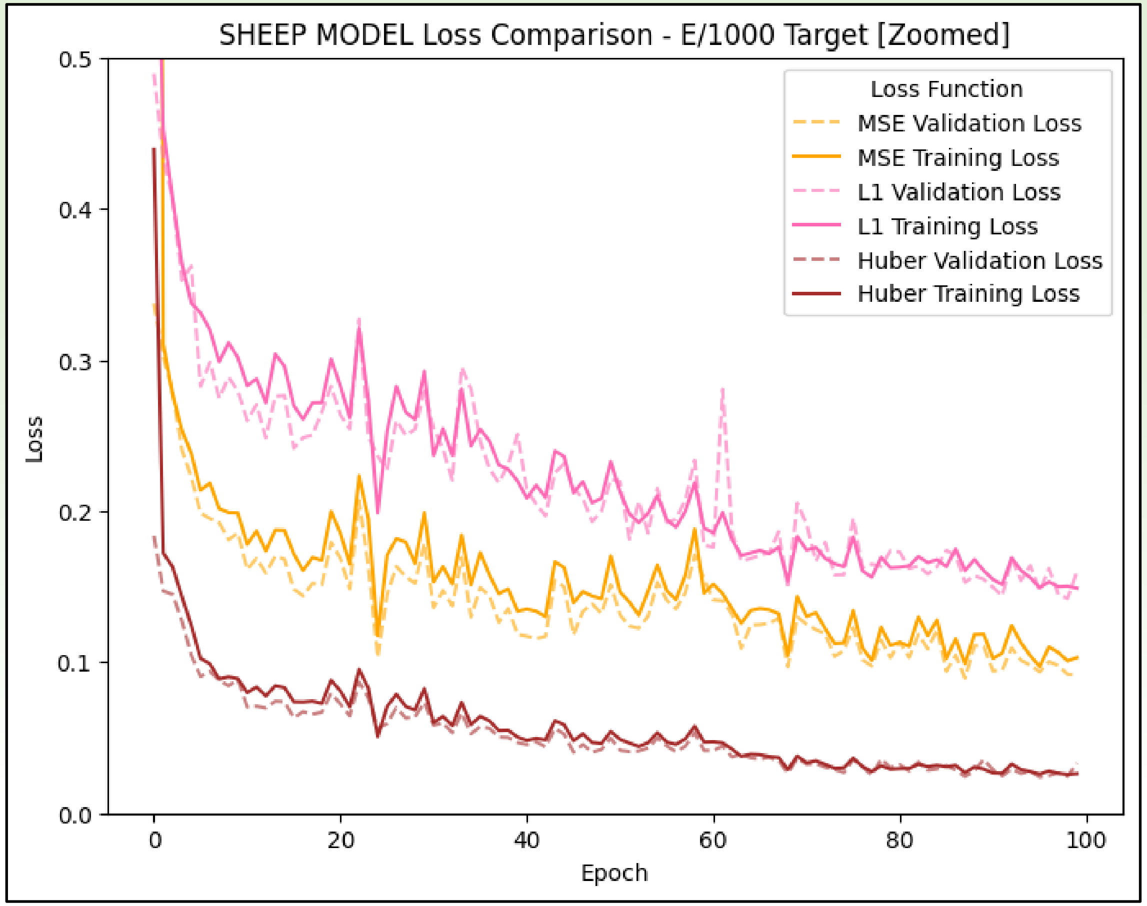



- **Huber Loss:** $\frac{1}{N}\sum_{i=1}^{N}\begin{cases} 0.5(\hat{y}_i - y_i)^2, & \text{if } |\hat{y}_i - y_i| < \delta \\ \delta \cdot (|\hat{y}_i - y_i| - 0.5 * \delta), & \text{otherwise} \end{cases}$
- $\boldsymbol{y_i}$= true KE, $\boldsymbol{\hat{y}_i}$ = SHEEP prediction, $\boldsymbol{\delta}$= 1.0 (default value)

# Post-Fix Loss Curves – All E/1000 Target Variations



# Post-Fix Loss Curves – All log(E) Target Variations