Parzen-PNN Gaussian Mixture Estimator: Results Report

Fabrizio Benvenuti

January 30, 2026

1 Results

1.1 Parzen Window

In this section are reported the estimation results obtained by varying
the input parameters of the Parzen Window estimator.
(i.e., window size h_1 and number of sampled points per gaussian in the mixture).

1.1.1 Parzen Window Errors

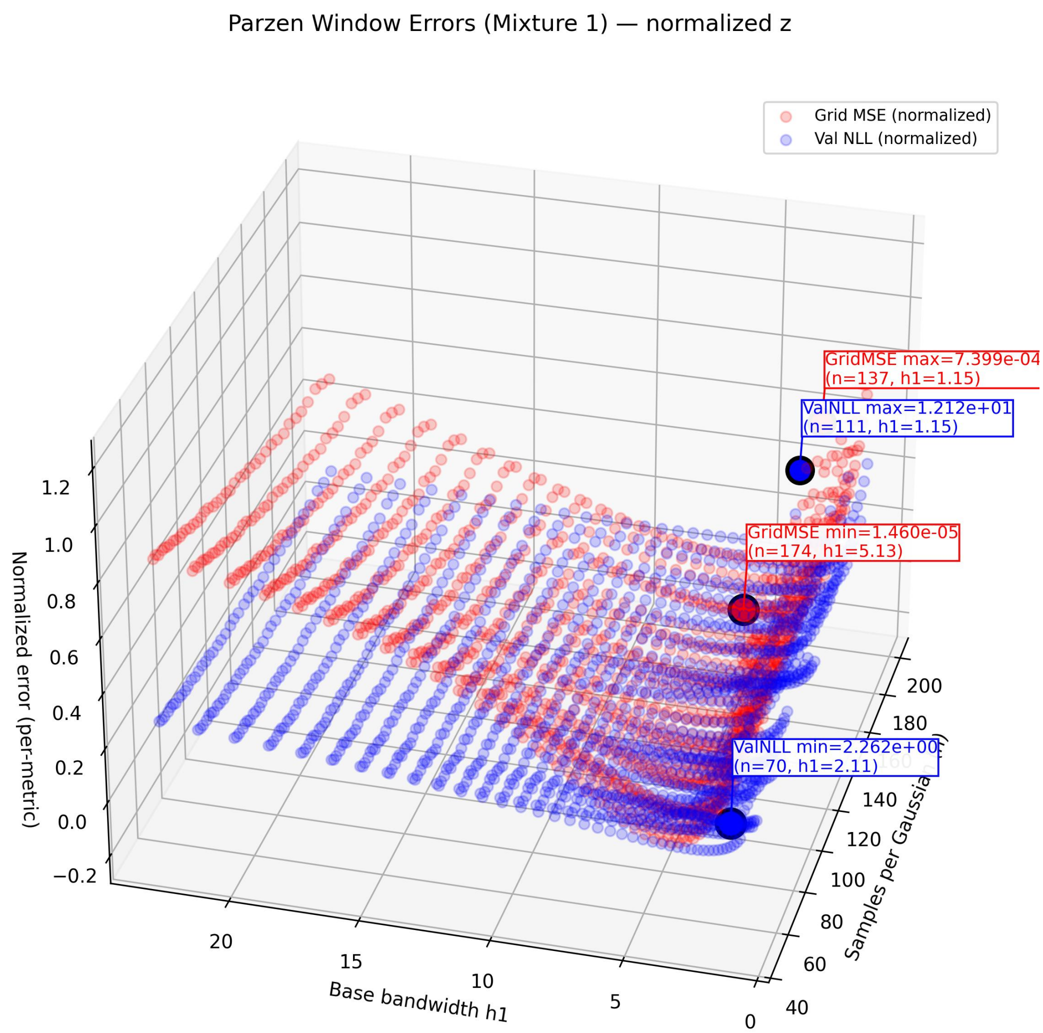

Figure 1: MSE between the mixture 1 pdf and its estimate PW estimate; while varying the window size h_1 and the sampled points for each gaussian.

In this graph it's also noticeable how the MSE does not vary linearly with window size; approaching zero at the optimal value of h_1 and increasing exponentially when undersmoothing occurs; transitioning to the oversmoothed region, the MSE still increases but less steeply. Samples per gaussian seem to have an exponential impact on reducing MSE, even if its effect is less visible than the base bandwidth one. NLL does not show the same steepness in the oversmoothing region as MSE does.

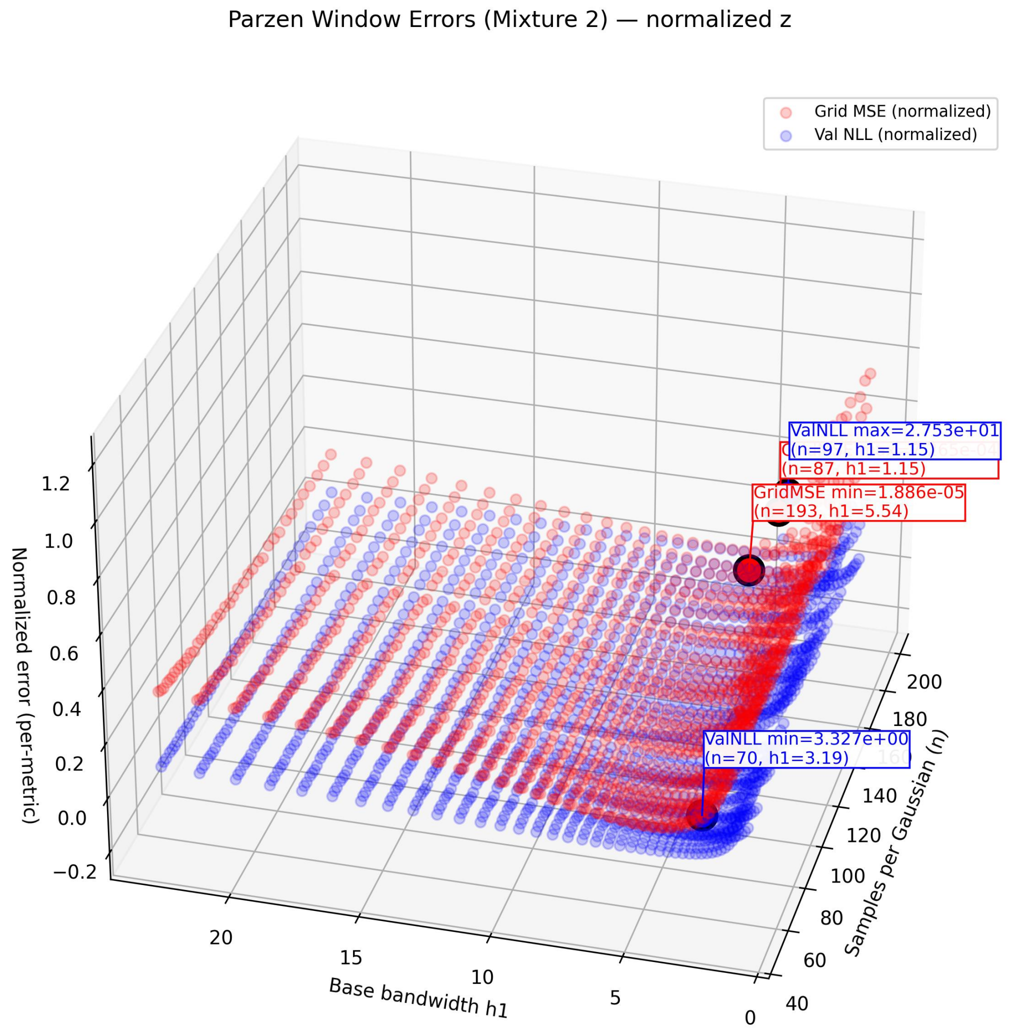

Figure 2: MSE between the mixture 2 pdf and its estimate PW estimate; while varying the window size h_1 and the sampled points for each gaussian.

In this graph it's also noticeable how the MSE steepnes in the oversmoothed region is less pronounced compared to mixture 1; this is probably due to the fact that oversmothing the probability mass causes less error when the modes are closer together; whilst, in the mixture 1, overmoothing causes the probability mass to fall on the tails; which generates more error.

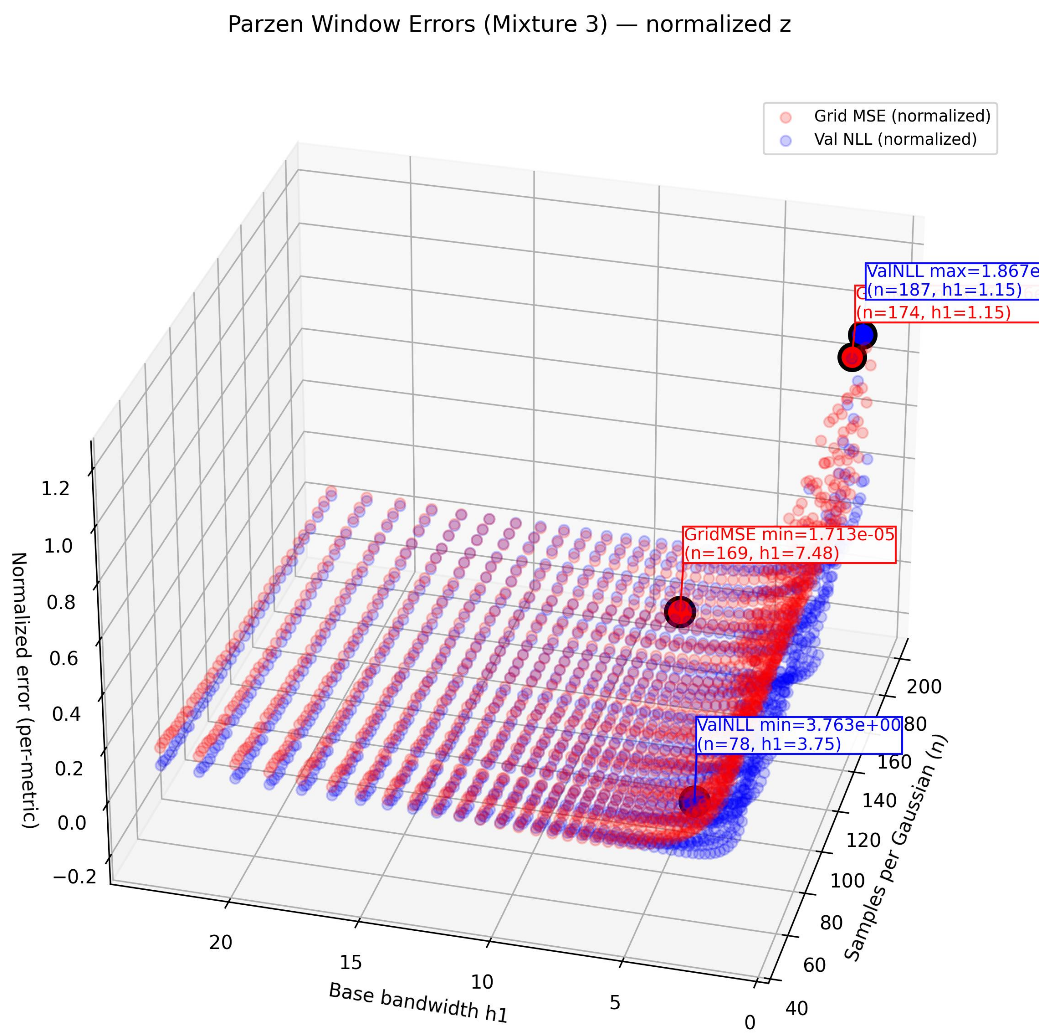

Figure 3: MSE between the mixture 3 pdf and its estimate PW estimate; while varying the window size h_1 and the sampled points for each gaussian.

In this graph it's see how increasing sampled points seems to have less impact on reducing MSE compared to mixture 1 and 2. NLL does also not seem to change neither the shape in the undersmoothing region nor the optimal h_1 value. whilst the real optimal h_1 for the MSE seems to change with the number of gaussians in the mixture. The NLL does not, causing the selected mixtures to be more oversmoothed, the more gaussians there are in the mixture.

$$h^{\text{opt}}_{1,\text{MSE}}(\text{mix}_1, \text{mix}_2, \text{mix}_3) = (5.13,\ 5.54,\ 7,48), \qquad h^{\text{opt}}_{1,\text{NLL}}(\text{mix}_1, \text{mix}_2, \text{mix}_3) = (2.11,\ 3.19,\ 3.75).$$

1.1.2 Parzen Window Overlays

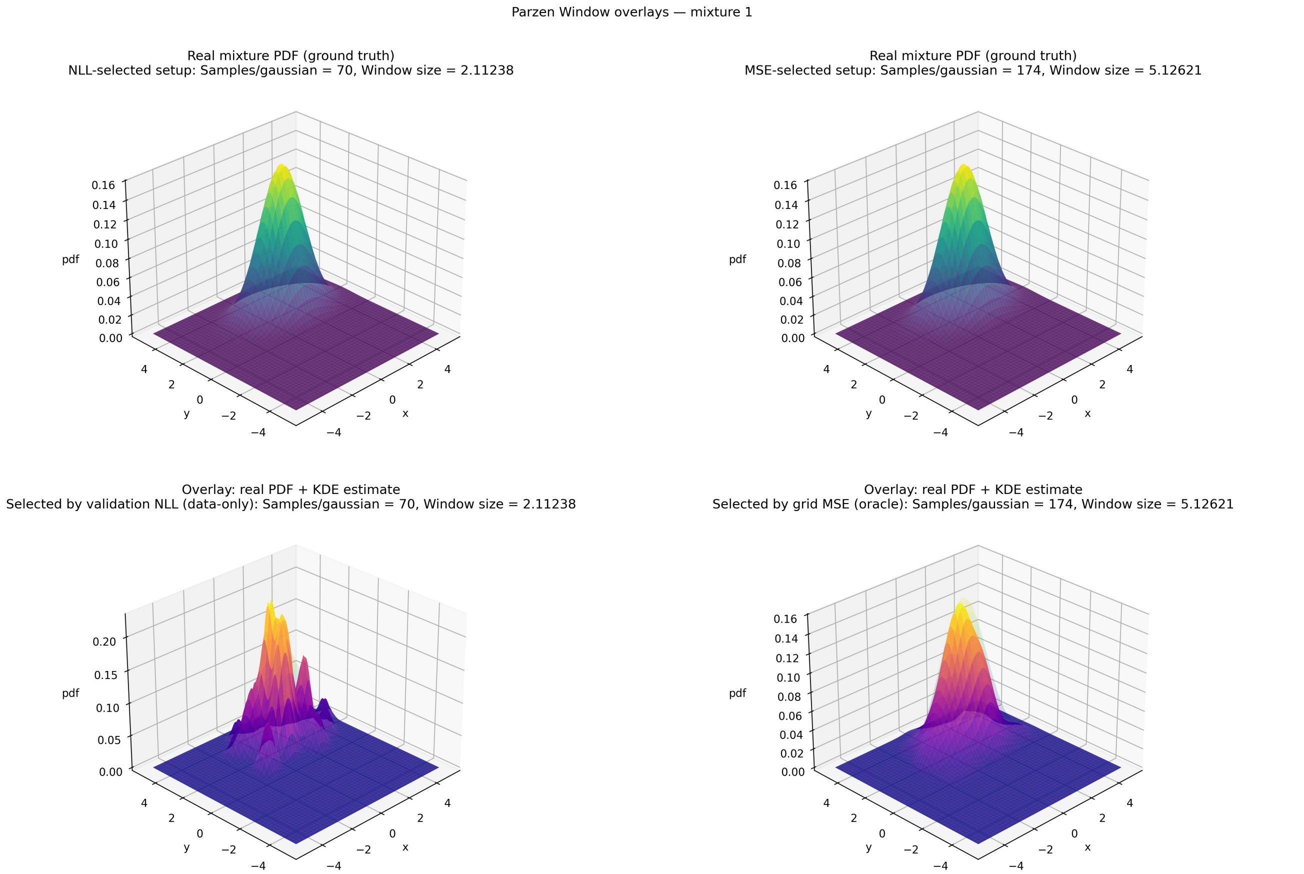

Figure 4: Figure displaying the overlay between the real pdf of mixture 1 and its PW estimate; with the optimal parameters selected by MSE and NLL respectively.

In this graph it's noticeable how increasing the sampled points per gaussian does help in reducing the MSE; but the effect on the NLL is less visible. Therefore the selected parameters by NLL are undersmoothed compared to the MSE ones.

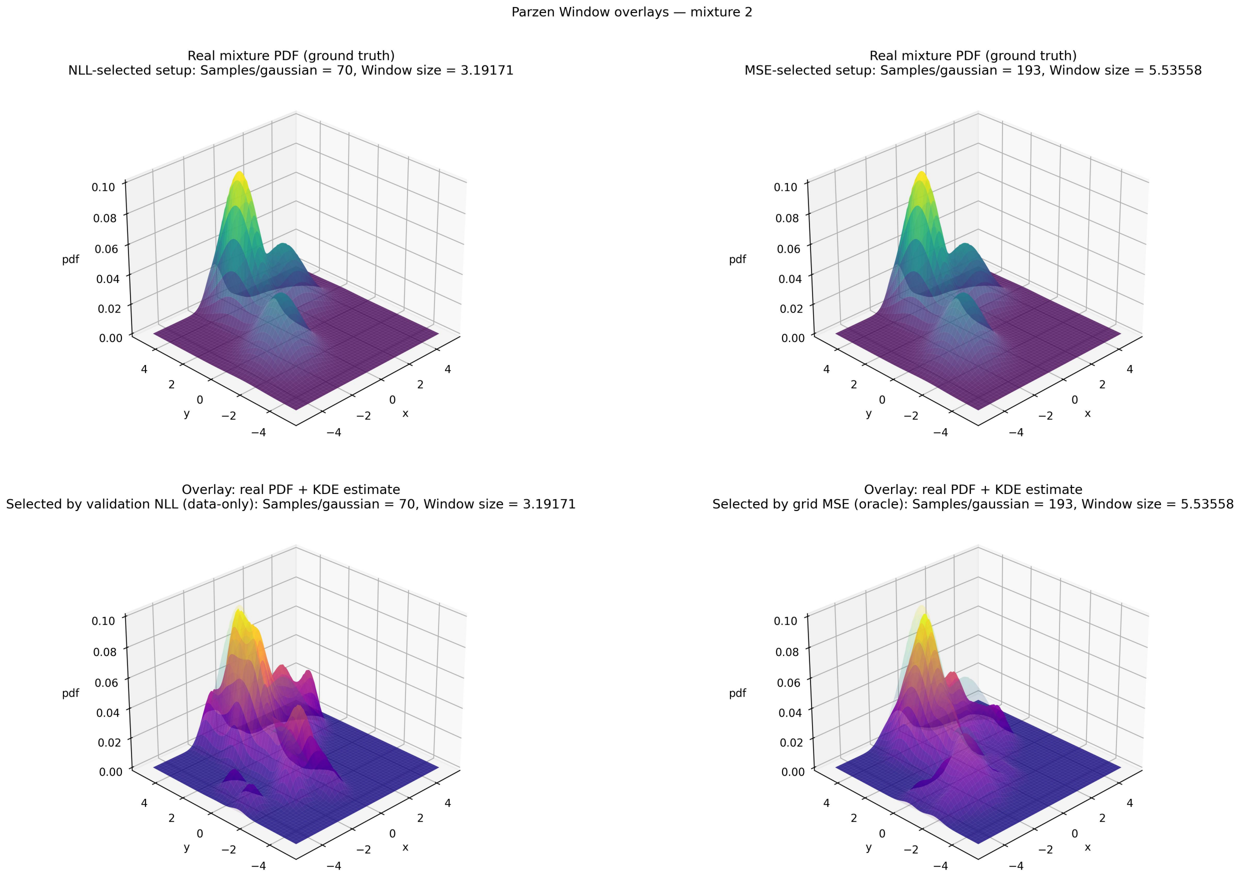

Figure 5: Figure displaying the overlay between the real pdf of mixture 2 and its PW estimate; with the optimal parameters selected by MSE and NLL respectively.

It's also noticeable how the MSE selected parameters still cannot provide high accuracy in estimating the pdf's gaussians with very different variances, this is due to the fach that h_1 is constant across the whole input space. therefore high peaks result oversmoothed to accommodate for the wider modes and viceversa.

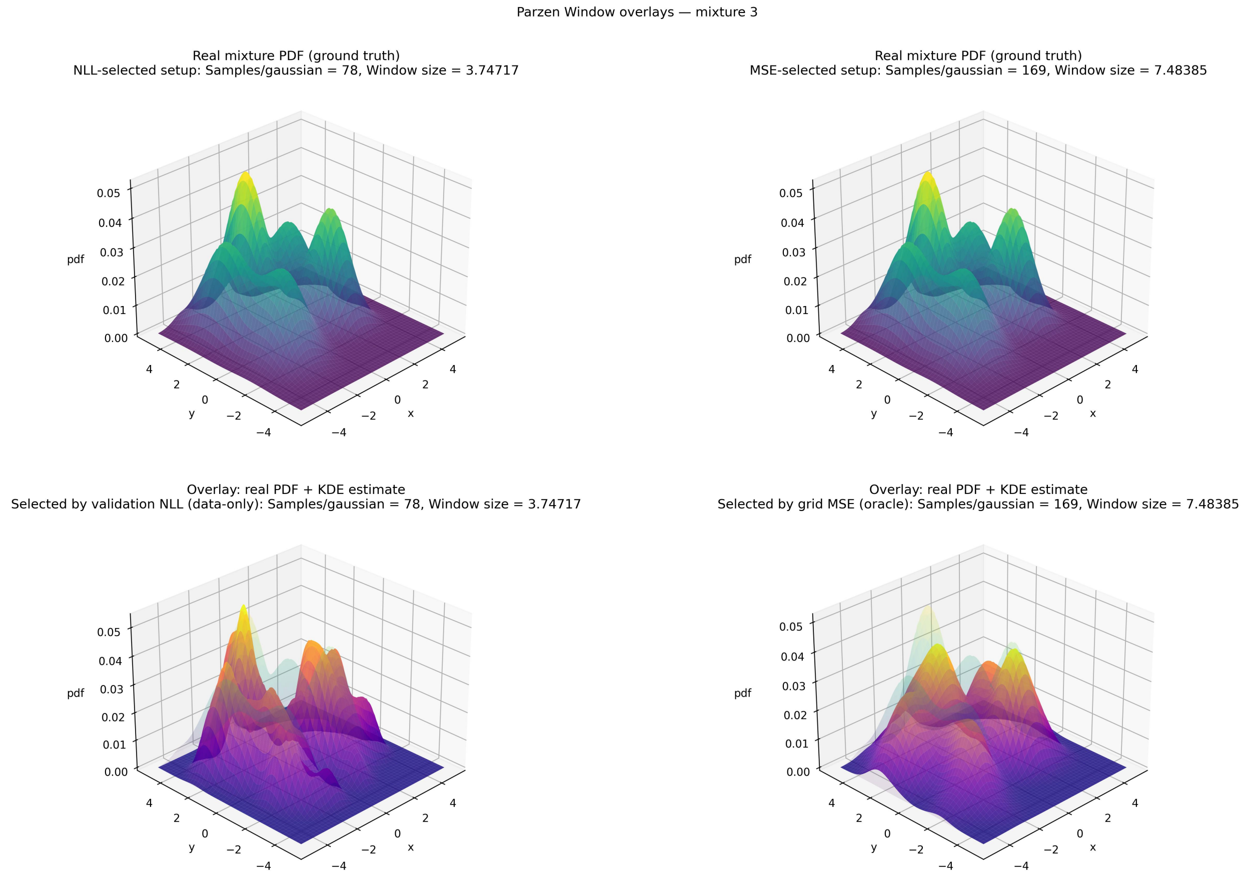

Figure 6: Figure displaying the overlay between the real pdf of mixture 2 and its PW estimate; with the optimal parameters selected by MSE and NLL respectively.

In this graph it's instead noticeable how NLL selected parameters
may sometime lead the predicted pdf with undersmoothed regions;
that approximates single peaks as two distinct modes.
to still have a low enough NLL value; this is because data points drawn
from those peaks will still have high likelihood even if there is a valley in between them.
This effect is less visible in MSE selected parameters since the overall shape of the pdf
is more important than the likelihood of single data points.

1.2 Parzen Neural Network

1.2.1 Parzen Neural Network Errors

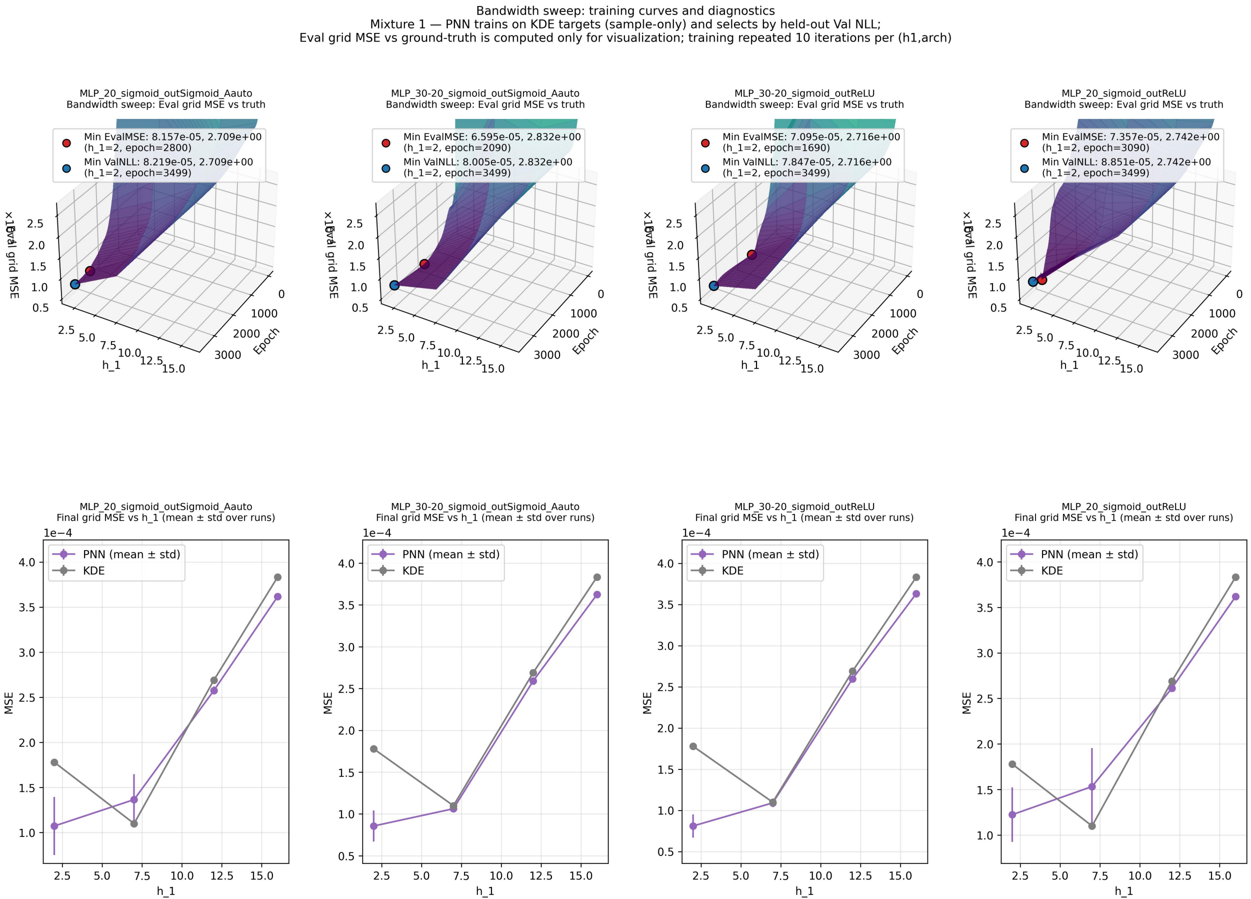

Figure 7: PNN bandwidth sweep (10 runs): top — representative iteration (closest to iteration mean) EvalMSE surface over epoch $\times$ h_1; bottom — final grid MSE vs h_1 shown as mean $\pm$ std across 10 runs (PNN) and KDE for reference.

In this graphs it's noticeable how the Min-EvalMSE tends to have a lower optimal bandwidth with respect to the Val-NLL selected one.
This is opposite to what was observed in the Parzen Window estimator;
which could also be seen in the error subgraphs with MSE vs h_1,
where it is obvious that the PNN seems to work better with undersmoothed KDEs.

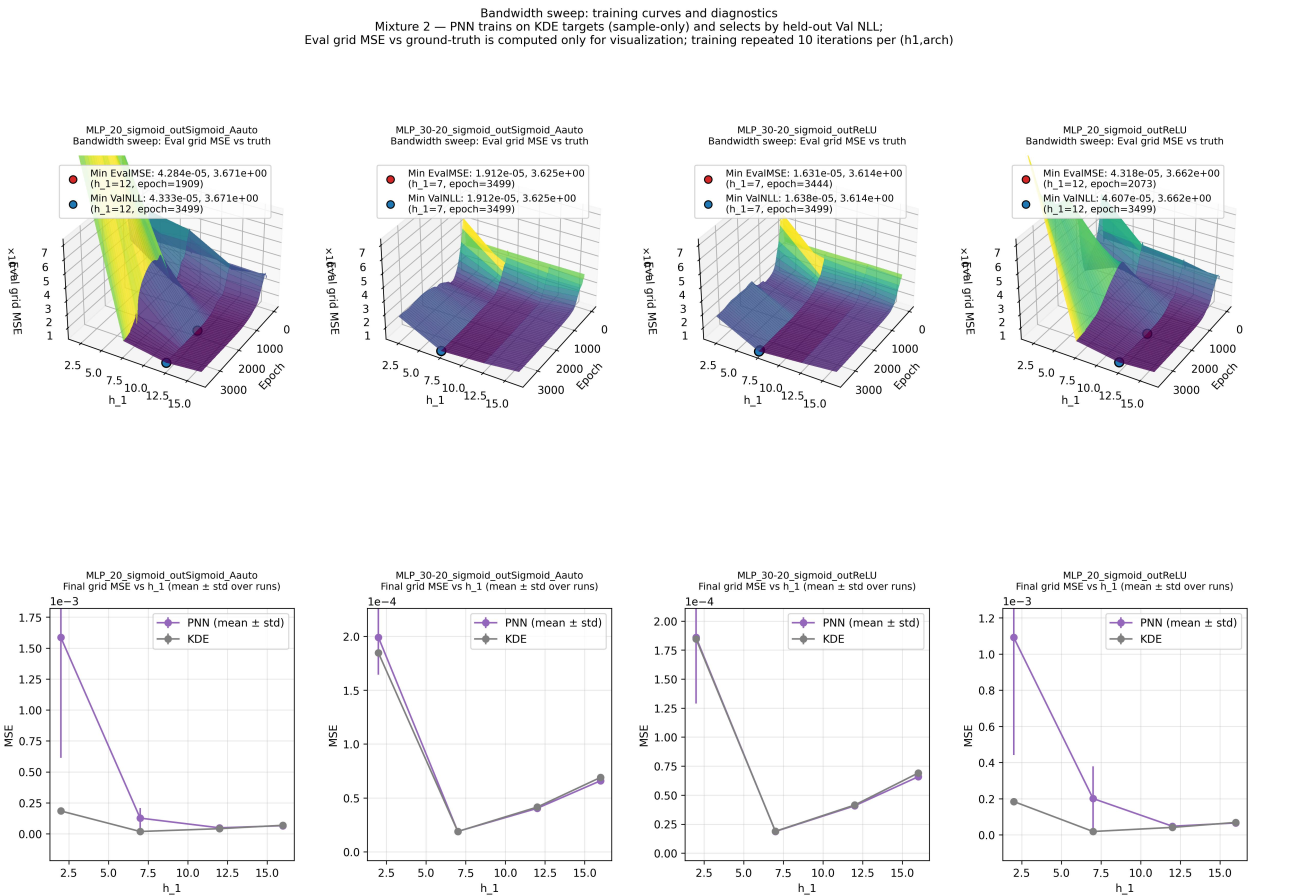

Figure 8: PNN bandwidth sweep (10 runs): top — representative iteration (closest to iteration mean) EvalMSE surface over epoch $\times$ h_1; bottom — final grid MSE vs h_1 shown as mean $\pm$ std across 10 runs (PNN) and KDE for reference.

In this graph it's noticable how increasing the number of gaussians in the mixture causes:

- Single hidden layer PNNs to perform worse than KDEs and double hidden layered PNNs, whilst beeing extremely sensitive to h_1 variations, especially in the undersmoothed region.
- The MSE mesh seems to be flattened, especially in the oversmoothed region, with respect to the mixture1.
- In this graph it's marked how PNNs with deeper architectures, have a high resiliency to having undersmoothed KDEs, whilst it still causes overfitting if the stopping rule is not set.

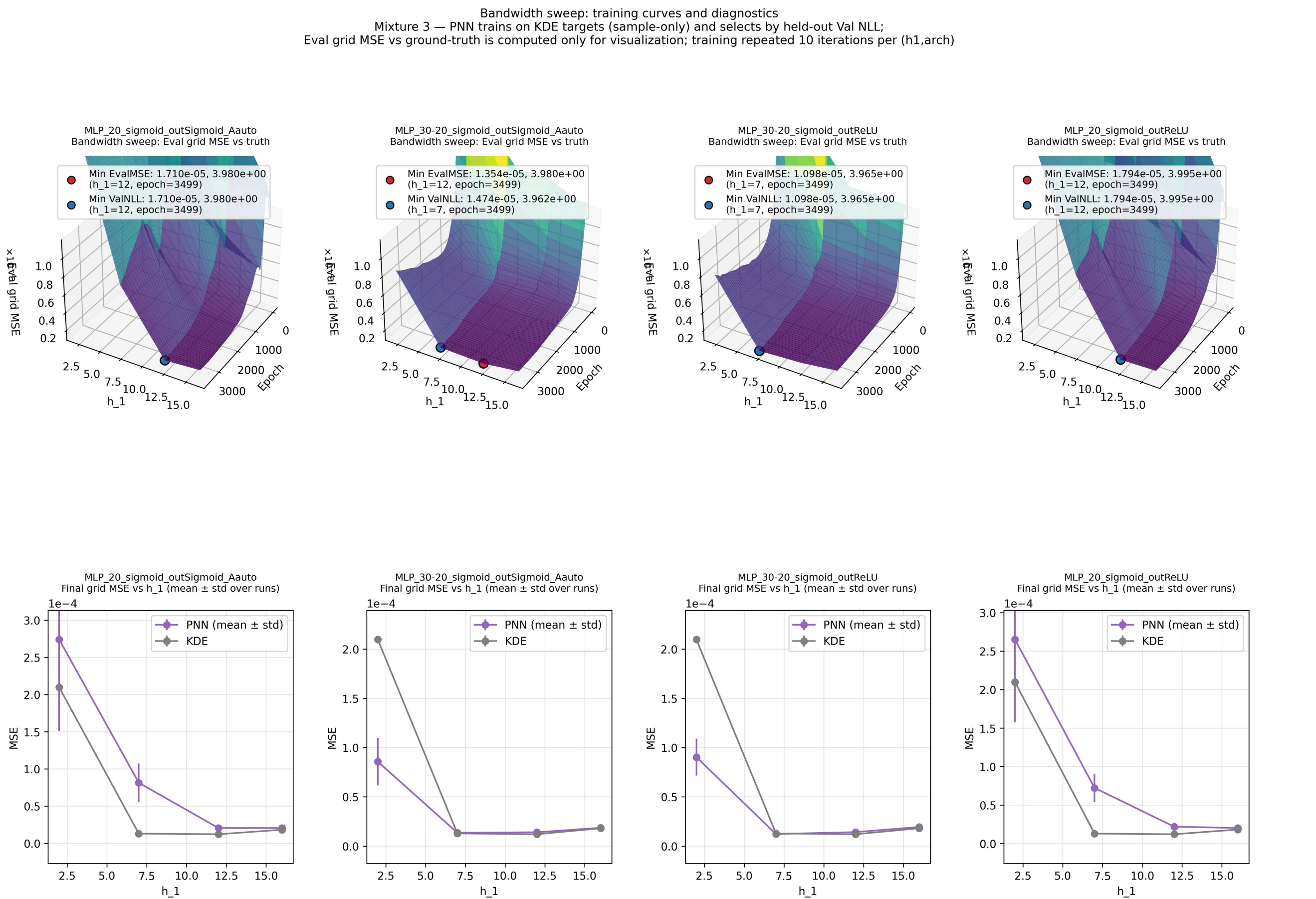

Figure 9: PNN bandwidth sweep (10 runs): top — representative iteration (closest to iteration mean) EvalMSE surface over epoch $\times$ h_1; bottom — final grid MSE vs h_1 shown as mean $\pm$ std across 10 runs (PNN) and KDE for reference.

It's marked in this graphs how increasing the number of points inside the PDF, by increasing the number of gaussians inside the mixture; causes the Val-NLL (based onto held-out data points) to be a good metric to determine the best combination of architecture, stopping epoch and KDE bandwidth. This could be seen because the mixture 3 is the one where Val-NLL points have the lowest EvalMSE, across all mixtures

1.2.2 Parzen Neural Network Overlays

In this subsection will be graphed all the overlays at the Min Val-NLL, for each mixture; to appreciate the differences between architectures.

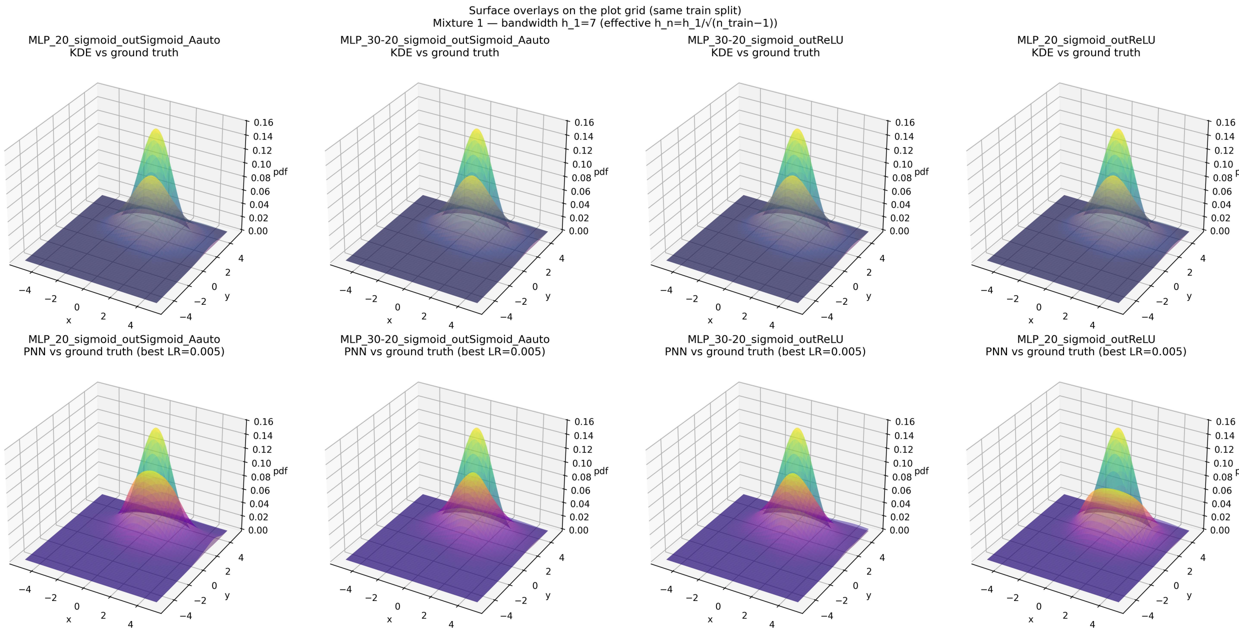

Figure 10: MSE between the mixture 3 pdf and its PNN estimate shown two ways:
top — EvalMSE surface over training epoch and Parzen bandwidth h_1;
bottom — MSE values (MSE vs h_1) for both KDE and PNN at ValNLL selection (epoch and h_1).