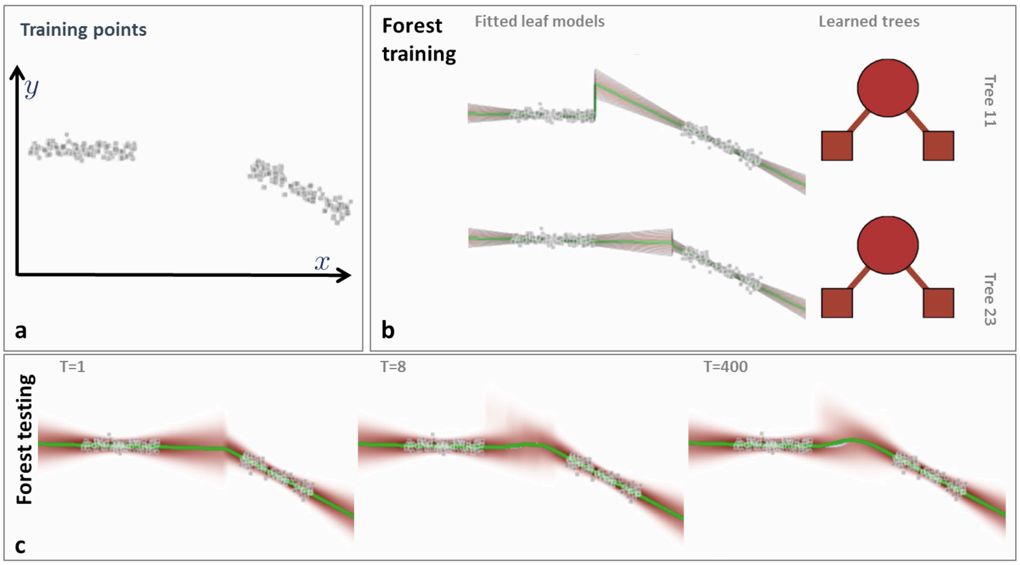

Fig. 4.6: **A first regression forest and the effect of its size T.** (**a**) Training points. (**b**) Two different shallow trained trees ($D = 2$) split the data into two portions and produce different piece-wise probabilistic-linear predictions. (**c**) Testing posteriors evaluated for all values of x and increasing number of trees. The green curve denotes the conditional mean $\mathcal{E}[y|x] = \int y \cdot p(y|x)\ dy$. The mean curve corresponding to a single tree ($T = 1$) shows a sharp change of direction in the gap. Increasing the forest size produces smoother class posteriors $p(y|x)$ and smoother mean curves in the interpolated region. All examples have been run with $D = 2$, axis-aligned weak learners and probabilistic-linear prediction models.

the training points and larger away from them. In the gap the actual split happens in different places along the x axis for different trees.

The bottom row (fig. 4.6c) shows the regression posteriors evaluated for *all* positions along the x axis. For each x position we plot the entire distribution $p(y|x)$, where darker red indicates larger values of the posterior. Thus, very compact, dark pixels correspond to high prediction confidence.