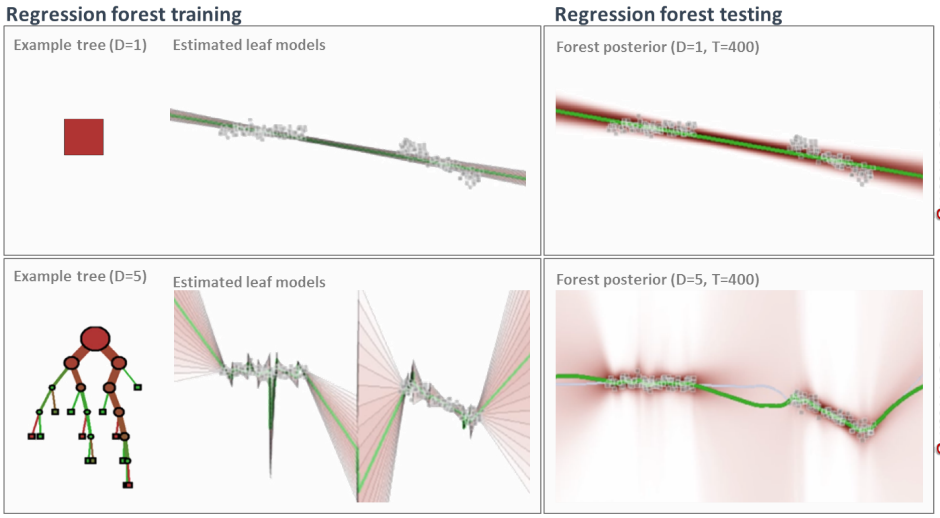

Fig. 4.7: **The effect of tree depth. (Top row)** Regression forest trained with $D = 1$. Trees are degenerate (each tree corresponds only to their root node). This corresponds to conventional linear regression. In this case the data is more complex than a single linear model and thus this forest under-fits. **(Bottom row)** Regression forest trained with $D = 5$. Much deeper trees produce the opposite effect, *i.e.* over-fitting. This is evident in the high-frequency, spiky nature of the testing posterior. In both experiments we use $T = 400$, axis-aligned weak learners and probabilistic-linear prediction models.

4.4 Comparison with alternative algorithms

The previous sections have introduced the probabilistic regression forest model and discussed some of its properties. This section shows a comparison between forests and allegedly the most common probabilistic regression technique, Gaussian processes [73].

4.4.1 Comparison with Gaussian processes

The hallmark of Gaussian processes is their ability to model uncertainty in regression problems. Here we compare regression forests with