3.2 Nonparametric estimation of the variance function

Model description An assumption underlying the ordinary regression

$$y_i = a + bx_i + \varepsilon_i'$$

is that all observations have the same variance, i.e. Var(ε_i') = σ^2. This assumption does not always hold, e.g. in Figure 3.2 a). It is clear that the variance increases to the right (for large values of x). It is also clear that the mean of y is not a linear function of x. We thus t the model

$$y_i = f(x_i) + \sigma(x_i)\varepsilon_i,$$

where $\varepsilon_i \sim N(0,1)$, and $f(x)$ and $\sigma(x)$ are modelled nonparametrically. As in Example 3.1 we take f to be a penalized spline. To ensure that $\sigma(x) > 0$ we model $\log[\sigma(x)]$, rather than $\sigma(x)$, as a spline function. For f we use a cubic spline (20 knots) with a 2nd order di erence penalty

$$-\lambda^2 \sum_{k=3}^{20} (u_j - 2u_{j-1} + u_{j-2})^2,$$

while we take $\log[\sigma(x)]$ to be a linear spline (20 knots) with the 1st order di erence penalty (3.2).

Implementation details Details on how to implement spline components are given Example 3.1.

- In order to estimate the variation function, one rst needs to have tted the mean part. Parameter associated with f should thus be given `phase 1' in ADMB, while those associated with $\sigma$ should be given `phase 2'.

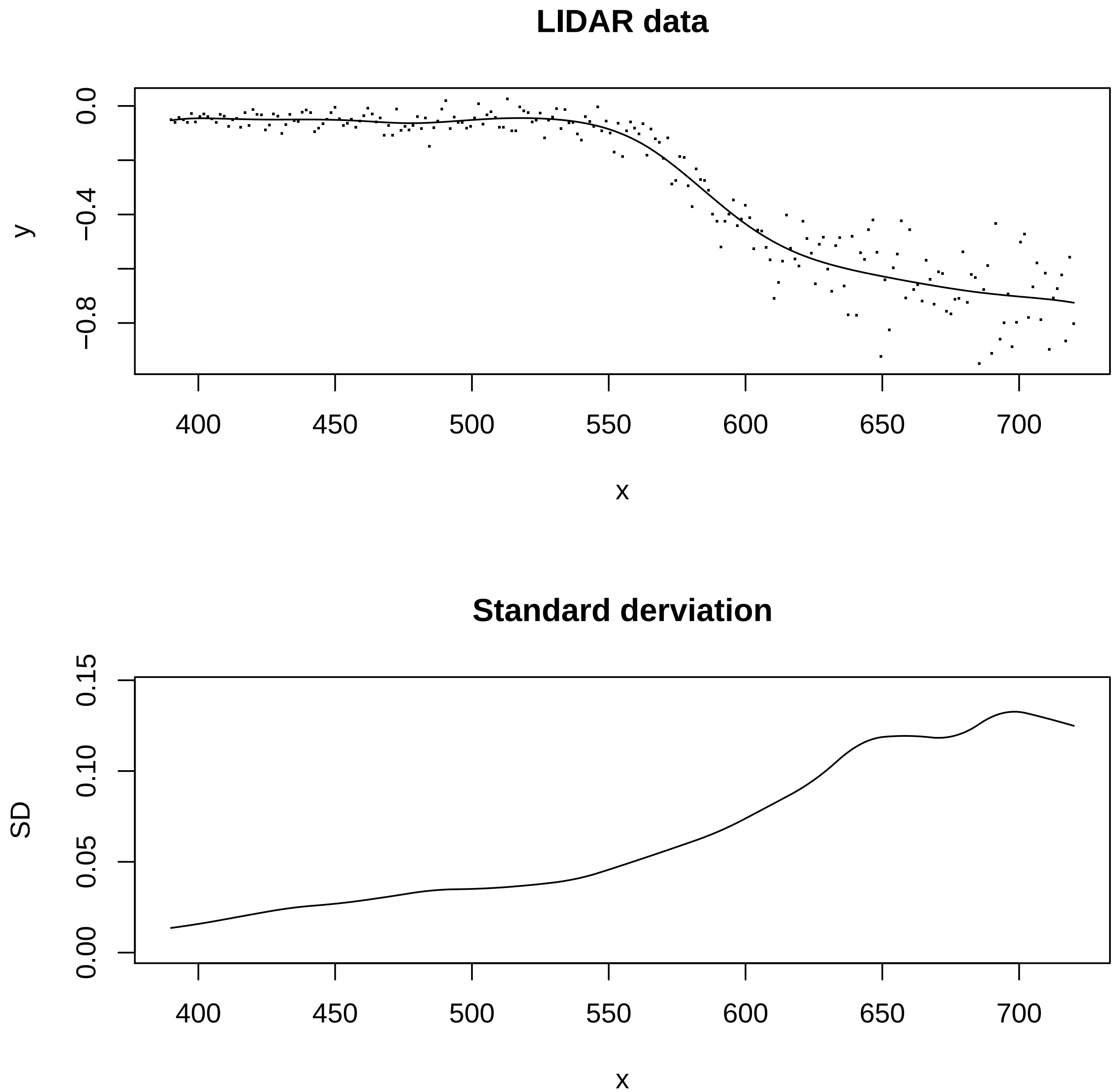

Figure 3.2: LIDAR data (upper panel) used by Ruppert et al. (2003) with tted mean. Fitted standard deviation is shown in the lower panel.