SUPPLEMENT TO MULTISCALE BLIND SOURCE SEPARATION

By Merle Behr*, Chris Holmes†, and Axel Munk*,‡

*University of Goettingen *, University of Oxford †,*
and Max Planck Institute for Biophysical Chemistry ‡

S1. Additional poofs.

S1.1. *Proof of Theorem 1.4.*

Proof. As $g, \tilde{g} \in \mathcal{M}^\delta$, (10) implies that there exist $a^1, \ldots, a^m, \tilde{a}^1, \ldots, \tilde{a}^m \in \mathfrak{A}^m$ such that

$$\begin{aligned} \left|\omega^\top a^i - \tilde{\omega}^\top [A]_i\right| &< \epsilon \quad \text{for } i = 1, \ldots, m, \\ \left|\tilde{\omega}^\top \tilde{a}^i - \omega^\top [A]_i\right| &< \epsilon \quad \text{for } i = 1, \ldots, m, \end{aligned} \tag{44}$$

with A as in (8).
First, we show by induction that (44) implies 1..
W.l.o.g. let $\tilde{\omega}_1 > \omega_1$. Assume that $\tilde{\omega}^\top \tilde{a}^1 < \tilde{\omega}^\top [A]_1 = a_1 + (a_2 - a_1)\tilde{\omega}_1$, i.e.,

$$\sum_{i=2}^m \tilde{\omega}_i(\tilde{a}^1_i - a_1) < \tilde{\omega}_1(a_2 - \tilde{a}^1_1). \tag{45}$$

As $\tilde{\omega}_1$ denotes the smallest mixing weight (recall $\tilde{\omega}_1 \leq \ldots \leq \tilde{\omega}_m$ in (3)) and a_1 and a_2 denote the smallest and second smallest, respectively, alphabet values (recall $a_1 < \ldots < a_k$ in (1)), it holds for any alphabet value $e \in \mathfrak{A} \setminus \{a_1\} = \{a_2, \ldots, a_k\}$ and $i = 1, \ldots, m$ that

$$\tilde{\omega}_i(e - a_1) \geq \tilde{\omega}_1(a_2 - a_1) \geq \tilde{\omega}_1(a_2 - \tilde{a}^1_1). \tag{46}$$

(45) and (46) imply that $\tilde{a}^1 = (a_1, \ldots, a_1)^\top$, i.e., $\tilde{\omega}^\top \tilde{a}^1 = a_1$. In particular, (44) yields $\left|a_1 - \omega^\top [A]_1\right| < \epsilon < \delta$, which contradicts $ASB(\omega) \geq \delta$. Consequently,

$$\tilde{\omega}^\top \tilde{a}^1 \geq \tilde{\omega}^\top [A]_1 = a_1 + (a_2 - a_1)\tilde{\omega}_1 > a_1 + (a_2 - a_1)\omega_1 = \omega^\top [A]_1$$

and therefore, by (44)

$$(a_2 - a_1)\left|\tilde{\omega}_1 - \omega_1\right| = \left|\tilde{\omega}^\top [A]_1 - \omega^\top [A]_1\right| < \epsilon.$$

Now, assume that $(a_2 - a_1)\,|\tilde{\omega}_i - \omega_i| < \epsilon$ for $i = 1, \ldots, l-1$.
W.l.o.g., let $\tilde{\omega}_l > \omega_l$. Assume that $\tilde{\omega}^\top \tilde{a}^l < \tilde{\omega}^\top [A]_l = a_1 + (a_2 - a_1)\tilde{\omega}_l$, i.e.,

$$\sum_{i=1,\, i\neq l}^{m} \tilde{\omega}_i(\tilde{a}_i^l - a_1) < \tilde{\omega}_l(a_2 - \tilde{a}_l^l). \tag{47}$$

Again, as $\tilde{\omega}_1 \leq \ldots \leq \tilde{\omega}_m$ and $a_1 < \ldots < a_k$, it holds for any alphabet value $e \in \mathfrak{A} \setminus \{a_1\} = \{a_2, \ldots, a_k\}$ and $i \geq l$ that

$$\tilde{\omega}_i(e - a_1) \geq \tilde{\omega}_l(a_2 - a_1) \geq \tilde{\omega}_l(a_2 - \tilde{a}_l^l). \tag{48}$$

(47) and (48) imply that $\tilde{a}_l^l = \ldots = \tilde{a}_m^l = a_1$ and therefore,

$$\begin{aligned}
\left|\omega^\top [A]_l - \omega^\top \tilde{a}^l\right| &\leq \left|\omega^\top [A]_l - \tilde{\omega}^\top \tilde{a}^l\right| + \left|\tilde{\omega}^\top \tilde{a}^l - \omega^\top \tilde{a}^l\right| \\
&< \epsilon + \left|\sum_{i=1}^{l-1} (\tilde{a}_i^l - a_1)(\tilde{\omega}_i - \omega_i)\right| \\
&\leq \epsilon + (m-1)\frac{a_k - a_1}{a_2 - a_1}\epsilon \leq m\frac{a_k - a_1}{a_2 - a_1}\epsilon < \delta,
\end{aligned}$$

which contradicts $ASB(\omega) \geq \delta$. Consequently, $\tilde{\omega}^\top \tilde{a}^l \geq \tilde{\omega}^\top [A]_l > \omega^\top [A]_l$ and therefore,

$$(a_2 - a_1)\,|\tilde{\omega}_l - \omega_l| = \left|\tilde{\omega}^\top [A]_l - \omega^\top [A]_l\right| < \epsilon.$$

By induction 1. follows.
To prove 2., assume the contrary. Then there exist $a \neq \tilde{a} \in \mathfrak{A}^m$ such that

$$\epsilon > \left|\omega^\top a - \tilde{\omega}^\top \tilde{a}\right| \geq \left|\omega^\top a - \omega^\top \tilde{a}\right| - \left|\omega^\top \tilde{a} - \tilde{\omega}^\top \tilde{a}\right|$$

and by 1.

$$\left|\omega^\top \tilde{a} - \tilde{\omega}^\top \tilde{a}\right| = \left|\sum_{i=1}^{m} (\omega_i(\tilde{a}_i - a_1) - \tilde{\omega}_i(\tilde{a}_i - a_1))\right| \leq m\frac{a_k - a_1}{a_2 - a_1}\epsilon.$$

The last two inequalities give $\epsilon > \delta - m\epsilon(a_k - a_1)/(a_2 - a_1)$, which contradicts $2m(a_k - a_1)\epsilon < \delta(a_2 - a_1)$ as $m(a_k - a_1) > (a_2 - a_1)$.

□

S1.2. *Proof of Theorem 2.7.* The following Theorem is needed for the proof of Theorem 2.7 and shows that SLAM admits a solution with probability converging to one at a superpolynomial rate.
Let $N_1^\star$ be such that

$$\frac{\delta}{\sigma}\ln(N_1^\star) \geq 139\left(1+2m\frac{a_k-a_1}{a_2-a_1}\right)\sqrt{2\ln(e/\lambda^\star)}+70 \tag{49}$$

and $\lambda^\star \geq \lambda$ as in R4. Analog to $\mathcal{M}_\lambda^\delta$ in (39) define

$$\mathcal{S}(\mathfrak{A})_\lambda^m := \{f \in \mathcal{S}(\mathfrak{A})^m \text{ separable}: \min_{j\in\{0,,\dots,K(f)\}} |\tau_{j+1}-\tau_j| \geq \lambda\}, \tag{50}$$

where τ_j denote the change points of f, that is, at least one of the f^i's jumps, and $K(f)$ the number of change points of f.

THEOREM S1.1. *Consider the SBSSR-model with $g \in \mathcal{M}_\lambda^\delta$. Let α_n and β_n be as in (23). Further, let $\mathcal{C}_{1-\alpha}(Y)$ be as in (20) and let $\hat\omega$ be any weight vector in $\mathcal{C}_{1-\alpha_n}(Y)$. Then for all $n \geq N_1^\star$ in (49)*

$$\boldsymbol{P}\left(\min_{\tilde f \in \mathcal{S}(\mathfrak{A})_\lambda^m} T_n(Y,\hat\omega^\top \tilde f) \leq q_n(\beta_n)\,\middle|\, T_n(Y,g)\leq q_n(\alpha_n)\right) = 1.$$

PROOF. Let $\tilde\omega \in \Omega(m)$ and $\alpha \in (0,1)$ be fixed. Define the set

$$\mathcal{N}(\tilde\omega) := \left\{\breve\omega^\top a : a \in \mathfrak{A}^m \text{ and } \|\breve\omega - \tilde\omega\|_\infty \leq 2\sigma\frac{q_n(\alpha)+\sqrt{2\ln(e/\lambda^\star)}}{\sqrt{n\lambda^\star}(a_2-a_1)}\right\}$$

and, analog to $\mathcal{S}(\mathfrak{A})_\lambda^m$ in (50),

$$\mathcal{S}(\mathcal{N}(\tilde\omega))_\lambda :=$$
$$\left\{g \in \mathcal{S}(\mathcal{N}(\tilde\omega)): \min_{j\in\{0,,\dots,K(g)\}}|\tau_{j+1}-\tau_j| \geq \lambda \text{ and } a_1+(a_2-a_1)\tilde\omega_i \in \mathrm{Im}(g)\right\},$$

where $\mathrm{Im}(g) := \{g(x): x \in [0,1)\}$ denotes the image of g. Then it follows from R1, R3, R4, (16), and Remark 2.2 that conditioned on $\{\tilde\omega \in \mathcal{C}_{1-\alpha}(Y)\}$ and $\{T_n(Y,g) \leq q_n(\alpha)\}$

$$\inf_{\tilde g \in \mathcal{S}(\mathcal{N}(\tilde\omega))_\lambda} T_n(Y,\tilde g) \leq q_n(\alpha) \quad \text{a.s..} \tag{51}$$

Further, for $\epsilon_n := 2m\sigma\frac{a_k-a_1}{a_2-a_1}\left(q_n(\alpha)+\sqrt{2\ln(e/\lambda^\star)}\right)/\sqrt{n\lambda^\star}$ we have that

$$\sup_{\tilde g \in \mathcal{S}(\mathcal{N}(\tilde\omega))_\lambda}\min_{\tilde f \in \mathcal{S}(\mathfrak{A})_\lambda^m} \|\tilde g - \tilde\omega^\top \tilde f\|_\infty \leq \epsilon_n. \tag{52}$$

Let $(y_n)_{n\in\mathbb{N}}$ be a fixed sequence in $\mathbb{R}$, and denote $y^n := (y_1,\dots,y_n)$. Let $\epsilon > 0$, and $g, g' \in \mathcal{M}_\lambda$ be such that $\sup_{x\in[0,1)} |g(x) - g'(x)| \le \epsilon$. Then by the reverse triangle inequality

$$\begin{aligned}
\left|T_n(y^n, g) - T_n(y^n, g')\right| &\le \max_{\substack{1\le i\le j\le n\\ j-i+1\ge n\lambda}} \left| \frac{\left|\sum_{l=i}^j y_l - g(x_l)\right| - \left|\sum_{l=i}^j y_l - g'(x_l)\right|}{\sigma\sqrt{j-i+1}} \right| \\
&\le \max_{\substack{1\le i\le j\le n\\ j-i+1\ge n\lambda}} \frac{\left|\sum_{l=i}^j g(x_l) - g'(x_l)\right|}{\sigma\sqrt{j-i+1}} \le \frac{\sqrt{n\lambda}}{\sigma}\epsilon.
\end{aligned}$$

This, together with (51) and (52), implies that conditioned on $\{\tilde\omega \in \mathcal{C}_{1-\alpha}(Y)\}$ and $\{T_n(Y, g) \le q_n(\alpha)\}$

$$\begin{aligned}
&\inf_{\tilde\omega\in\Omega(m)} \mathbf{P}\left(\min_{\tilde f\in\mathcal{S}(\mathfrak{A})_\lambda^m} T_n(Y, \tilde\omega^\top \tilde f) \le q_n(\alpha) + \frac{\sqrt{n\lambda}}{\sigma}\epsilon_n \right) \\
&\ge \inf_{\tilde\omega\in\Omega(m)} \mathbf{P}\left(\inf_{\tilde g\in\mathcal{S}(\mathcal{N}(\tilde\omega))_\lambda} T_n(Y, \tilde g) \le q_n(\alpha) \right) = 1,
\end{aligned} \tag{53}$$

where the inequality results from

$$\begin{aligned}
&\min_{\tilde f\in\mathcal{S}(\mathfrak{A})_\lambda^m} T_n(Y, \tilde\omega^\top \tilde f) \\
&= \inf_{\tilde g\in\mathcal{S}(\mathcal{N}(\tilde\omega))_\lambda} T_n(Y, \tilde g) + \left(\min_{\tilde f\in\mathcal{S}(\mathfrak{A})_\lambda^m} T_n(Y, \tilde\omega^\top \tilde f) - \inf_{\tilde g\in\mathcal{S}(\mathcal{N}(\tilde\omega))_\lambda} T_n(Y, \tilde g) \right) \\
&\le \inf_{\tilde g\in\mathcal{S}(\mathcal{N}(\tilde\omega))_\lambda} T_n(Y, \tilde g) + \sup_{\tilde g\in\mathcal{S}(\mathcal{N}(\tilde\omega))_\lambda} \min_{\tilde f\in\mathcal{S}(\mathfrak{A})_\lambda^m} \left| T_n(Y, \tilde\omega^\top \tilde f) - T_n(Y, \tilde g) \right|.
\end{aligned}$$

It remains to show that for all $n \ge N_1^\star$

$$q_n(\alpha_n) + \frac{\sqrt{n\lambda}}{\sigma}\epsilon_n \le q_n(\beta_n). \tag{54}$$

To this end, we need some results about the quantile function of the multiscale statistic T_n from (14). Easy calculations and Mill's ratio give for all $n \in \mathbb{N}$

$$\mathbf{P}(T_n > q) \ge \sqrt{\frac{2}{\pi}} \left(\frac{1}{\tilde q} - \frac{1}{\tilde q^3} \right) \exp\left(-\tilde q^2/2\right), \quad \text{with } \tilde q := q + \sqrt{2\ln(e/\lambda^\star)},$$

which implies

$$q_n(\alpha) \ge \sqrt{\left| -\ln(\alpha\sqrt{\pi/2}) \right|} - \sqrt{2\ln(e/\lambda^\star)}. \tag{55}$$

Further, a slight modification of [2, Corollary 4] gives for all $n \in \mathbb{N}$ and $q > C$, for some constant $C < \infty$, that

$$\mathbf{P}(T_n > q) \leq \exp(-q^2/8), \tag{56}$$

which implies

$$q_n(\alpha) \leq \sqrt{-8\ln(\alpha)}. \tag{57}$$

From (57) and (23) we follow that

$$\begin{aligned} q_n(\alpha_n) + \frac{\sqrt{n\lambda}}{\sigma}\epsilon_n &= q_n(\alpha_n) + 2m\frac{a_k - a_1}{a_2 - a_1}\left(q_n(\alpha_n) + \sqrt{2\ln(e/\lambda^\star}\right) \\ \leq \left(\sqrt{8c_1} + 2m\frac{a_k - a_1}{a_2 - a_1}\sqrt{8c_1}\right)\ln(n) &+ 2m\frac{a_k - a_1}{a_2 - a_1}\sqrt{2\ln(e/\lambda^\star)} \end{aligned} \tag{58}$$

and from (55) and (23) that

$$q_n(\beta_n) \geq \sqrt{75m^2\left(\frac{a_k - a_1}{a_2 - a_1}\right)^2 c_1\ln(n)} - \sqrt{\ln(\sqrt{\pi/2})} - \sqrt{2\ln(e/\lambda^\star)}. \tag{59}$$

(49) yields that the right hand side of (58) is smaller than the right hand side of (59) for all $n \geq N_1^\star$, which yields (54) and, thus, together with (53), that conditioned on $\{\tilde{\omega} \in \mathcal{C}_{1-\alpha_n}(Y)\}$ and $\{T_n(Y, g) \leq q_n(\alpha_n)\}$

$$\inf_{\tilde{\omega}\in\Omega(m)} \mathbf{P}\left(\min_{\tilde{f}\in\mathcal{S}(\mathfrak{A})_\lambda^m} T_n(Y, \tilde{\omega}^\top\tilde{f}) \leq q_n(\beta_n)\right) = 1.$$

As $\hat{\omega} \in \mathcal{C}_{1-\alpha_n}$ a.s., this yields the assertion. □

The following theorem is a slight variation of Theorem 2.7, from which, together with Theorem S1.1, Theorem 2.7 will follow easily.

THEOREM S1.2. *Consider the SBSSR-model with $g \in \mathcal{M}_\lambda^\delta$. Let $q_n(\alpha)$ be as in (17), α_n as in (23), and β_n such that*

$$q_n(\alpha_n) < q_n(\beta_n) < \frac{\delta}{9\sigma}\ln(n). \tag{60}$$

Let $\hat{g} = \hat{\omega}^\top\hat{f} \in \mathcal{M}$ be the SLAM estimator of g with $\alpha = \alpha_n$, $\beta = \beta_n$, and $T_n(Y, \hat{g}) \leq q_n(\beta_n)$. Further, let $\hat{\tau}$ and τ be the vectors of all change points of

$\hat{g}$ and g, respectively. Define

$$\begin{aligned} A_n :=& \Big\{ \max_j |\hat{\tau}_j - \tau_j| \leq 2\frac{\ln(n)^2}{n} \Big\} \cap \Big\{ K(\hat{g}) = K(g) \Big\} \\ &\cap \Big\{ \max_j \max_i \Big| \hat{f}^i|_{[\hat{\tau}_j, \hat{\tau}_{j+1})} - f^i|_{[\tau_j,\tau_{j+1})} \Big| = 0 \Big\} \\ &\cap \Big\{ \max_i |\hat{\omega}_i - \omega_i| < \frac{\delta + \sqrt{2\sigma^2 \ln(e/\lambda)}}{\sqrt{\lambda}(a_2 - a_1)} \frac{\ln(n)}{\sqrt{n}} \Big\}. \end{aligned}$$

Then for all $n > N^\star$ in (41) and (42) $\boldsymbol{P}(A_n|\, T_n(Y, g) \leq q_n(\alpha_n)) = 1$.

Proof. Let $d_n := \ln^2(n)/n$ and

$$\mathcal{I} := \{[x_i, x_j] : 1 \leq i \leq j \leq n \text{ and } j - i + 1 \geq n\lambda\}.$$

We define a partition $\mathcal{I} = \mathcal{I}_1 \cup \mathcal{I}_2 \cup \mathcal{I}_3$ as follows.

$$\begin{aligned} &\mathcal{I}_1 := \{I \in \mathcal{I} :\ I \text{ contains more than two change points of } g\}, \\ &\mathcal{I}_2 := \{I \in \mathcal{I} : g|_I = g_1^I \mathbb{1}_{I_1} + g_2^I \mathbb{1}_{I_2} + g_3^I \mathbb{1}_{I_3},\ \text{with } |I_1| \geq |I_2| \geq |I_3|\,, \\ &|I_2| \leq d_n,\ \text{and } g_1^I, g_2^I, g_3^I \in \mathrm{Im}(g) \text{ pairwise different}\} \\ &\mathcal{I}_3 := \{I \in \mathcal{I} : g|_I = g_1^I \mathbb{1}_{I_1} + g_2^I \mathbb{1}_{I_2} + g_3^I \mathbb{1}_{I_3},\ \text{with } |I_1| \geq |I_2| \geq |I_3|\,, \\ &|I_2| > d_n,\ \text{and } g_1^I, g_2^I, g_3^I \in \mathrm{Im}(g) \text{ pairwise different}\}. \end{aligned}$$

Moreover, let $\mathfrak{B} := \{B(I) = B(i,j) \ :\ I = [x_i, x_j] \in \mathcal{I}\}$ be as in (16) with $q = q_n(\beta_n)$ and define $\|B(I)\| := \overline{b} - \underline{b}$ with $B(I) = [\underline{b}, \overline{b}]$. Furthermore, let $\mathfrak{B}_{\mathrm{nc}}$ be as in (32) and define

$$\epsilon_n := \frac{\delta + \sqrt{2\sigma^2 \ln(e/\lambda)}}{\sqrt{\lambda}} \frac{\ln(n)}{\sqrt{n}} \tag{61}$$

and

$$\begin{aligned} E_1 &:= \bigcap_{I \in \mathcal{I}_1 \cup \mathcal{I}_3} \{B(I) \in \mathfrak{B}_{\mathrm{nc}}\}, \\ E_2 &:= \bigcap_{I \in \mathcal{I}_2} \{B(I) \subset [g_1^I - \epsilon_n, g_1^I + \epsilon_n]\}, \\ E_3 &:= \{K(\hat{g}) = K(g)\} \cap \{\max_j |\hat{\tau}_j - \tau_j| \leq 2d_n\} \cap \{\max_j |\hat{g}(\hat{\tau}_j) - g(\tau_j)| < \epsilon_n\}. \end{aligned}$$

First, we show that

$$E_1 \cap E_2 \subset E_3. \tag{62}$$

To this end, consider Figure S1.1 and note that (conditioned on $\{T_n(Y,g) \leq q_n(\alpha_n)\}$) by Theorem S1.1 and (24) $\hat{g}$ has minimal scale λ for all $n > N^\star$.
If $B(I) \in \mathfrak{B}_{\rm nc}$, then $\hat{g}$ is not constant on I. Therefore, it follows from E_1 that $\hat{g}$ is constant only on intervals $I \in \mathcal{I}_2$.
Conversely, if $\hat{g}$ is constant on $I \in \mathcal{I}_2$ then $\hat{g}|_I \in B(I)$ (see orange bars in Figure S1.1) as $T_n(Y,\hat{g}) \leq q_n(\beta_n)$ by assumption.
Now, consider a change point of $\hat{g}$. Let $I, I' \in \mathcal{I}_2$ be the constant parts of $\hat{g}$ left and right of this change point and I_1, I_1' be those sub-intervals which include the largest constant piece of g (see green lines in Figure S1.1), with $g|_{I_1} \equiv g_1^I$ and $g|_{I_1'} \equiv g_1^{I'}$.
As $\epsilon_n < \delta/2$ for all $n > N^\star$ (see (42)) $\left|g_1^I - g_1^{I'}\right| > 0$ (see the vertical distance between the left and the right green line in Figure S1.1), such that g has at least one jump in a $2d_n$-neighborhood of a jump of $\hat{g}$. Conversely, as $2d_n < \lambda$ for all $n > N^\star$ (see (42)) g has at most one jump in a $2d_n$-neighborhood of a jump of $\hat{g}$. Consequently, (62) follows.

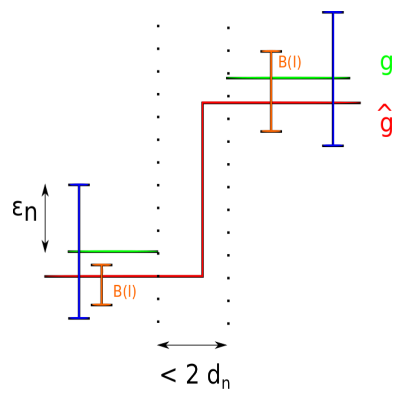

Fig S1.1. *The key argument underlying $E_1 \cap E_2 \subset E_3$.*

Furthermore, as $\epsilon_n < \delta(a_2 - a_1)/(2m(a_k - a_1))$ for all $n > N^\star$ (see (42)), Theorem 1.4 implies that

$$E_3 \subset A_n. \tag{63}$$

In the following we write $q_n \coloneqq q_n(\beta_n)$.
(62) and (63) implies that for all $n > N^\star$

$$\mathbf{P}\left(A_n \middle|\, T_n(Y,g) \leq q_n(\alpha_n)\right) \geq \mathbf{P}\left(E_1 \cap E_2 \middle|\, T_n(Y,g) \leq q_n(\alpha_n)\right).$$

First, consider E_1 conditioned on $\{T_n(Y,g) \leq q_n(\alpha_n)\}$:

Every interval $I \in \mathcal{I}_1$ includes a sub-interval I', which is the union of two constant pieces of g and, as $2d_n < \lambda$ for all $n > N^\star$ (see (42)), $I' \in \mathcal{I}_3$.
Consequently, conditioned on $\{T_n(Y,g) \leq q_n(\alpha_n)\}$ we have that for all $n > N^\star$

$$E_1 \supseteq \bigcap_{I \in \mathcal{I}_3} \{B(I) \in \mathfrak{B}_{\text{nc}}\} \supseteq \bigcap_{I \in \mathcal{I}_3} \{\delta > \|B(I_1)\| + \|B(I_2)\|\},$$

where I_1 and I_2 are the sub-intervals of $I \in \mathcal{I}_3$ such that $g|_{I_i} \equiv g_i^I$ for $i = 1, 2$ (as in the definition of $\mathcal{I}_3$).
By the definition of $\mathcal{I}_3$ it follows that $|I_1| \geq \lambda - 2d_n \geq \lambda/3$ for all $n > N^\star$ and $|I_2| > d_n$ and hence, (16) implies

$$\begin{aligned}\|B(I_1)\| + \|B(I_2)\| &\leq 2\left(\frac{q_n + \sqrt{2\ln(3e/\lambda)}}{\sqrt{n\lambda/3}/\sigma} + \frac{q_n + \sqrt{2\ln(e/d_n)}}{\sqrt{nd_n}\sigma}\right)\\ &= \frac{2\sigma}{\sqrt{n}}\left(\sqrt{\frac{3}{\lambda}}\left(q_n + \sqrt{2\ln(3e/\lambda)}\right) + \sqrt{\frac{1}{d_n}}\left(q_n + \sqrt{2\ln(e/d_n)}\right)\right).\end{aligned}$$

In summary we obtain that conditioned on $\{T_n(Y,g) \leq q_n(\alpha_n)\}$ for all $n > N^\star$

$$\begin{aligned}E_1 &\supseteq \left\{\delta > \frac{2\sigma}{\sqrt{n}}\left(\sqrt{\frac{3}{\lambda}}\left(q_n + \sqrt{2\ln(3e/\lambda)}\right) + \sqrt{\frac{1}{d_n}}\left(q_n + \sqrt{2\ln(e/(d_n))}\right)\right)\right\}\\ &= \left\{q_n < \left(\frac{\sqrt{n}\delta}{2\sigma} - \sqrt{\frac{6\ln(3e/\lambda)}{\lambda}} - \sqrt{\frac{2\ln(e/(d_n))}{d_n}}\right)\left(\sqrt{\frac{3}{\lambda}} + \sqrt{\frac{1}{d_n}}\right)^{-1}\right\}\\ &\supseteq \left\{q_n < \frac{\sqrt{n}\delta}{4\sigma}\left(\sqrt{\frac{3}{\lambda}} + \frac{\sqrt{n}}{\ln(n)}\right)^{-1}\right\}\\ &\supseteq \left\{q_n < \frac{\delta}{9\sigma}\ln(n)\right\},\end{aligned} \tag{64}$$

where the second inclusion results from (41) and the last inclusion from $2d_n < \lambda$ for all $n > N^\star$ (see (42)).
In particular, (64) and (60) yield $\mathbf{P}(E_1 | T_n(Y,g) \leq q_n(\alpha_n)) = 1$ for all $n > N^\star$.
Second, consider E_2 conditioned on $\{T_n(Y,g) \leq q_n(\alpha_n)\}$:
By (60), (61), and (40) it holds for all $I = [x_i, x_j] \in \mathcal{I}$ that

$$\|B(I)\| = 2\sigma\frac{q_n + \sqrt{2\ln(\frac{en}{j-i+1})}}{\sqrt{j-i+1}} \leq 2\sigma\frac{\frac{\delta}{9\sigma}\ln(n) + \sqrt{2\ln(\frac{e}{\lambda})}}{\sqrt{n\lambda}} < \epsilon_n/2$$

and as $\bar{g}_I := \sum_{l\in I} g(x_l)/(n\,|I|) \in B(I)$,

$$\begin{aligned} E_2 &\supseteq \bigcap_{I\in\mathcal{I}_2} \{|\bar{g}_I - g_1^I| \le \epsilon_n - \|B(I)\|\} \\ &\supseteq \bigcap_{I\in\mathcal{I}_2} \{|\bar{g}_I - g_1^I| \le \epsilon_n/2\}. \end{aligned}$$

Moreover, for $I \in \mathcal{I}_2$

$$\begin{aligned} \left|\bar{g}_I - g_1^I\right| &= \left|(g_2^I - g_1^I)\frac{|I_2|}{|I|} + (g_3^I - g_1^I)\frac{|I_3|}{|I|}\right| \\ &\le \frac{|I_2| + |I_3|}{|I|}(a_k - a_1) \le \frac{2d_n}{\lambda}(a_k - a_1). \end{aligned} \tag{65}$$

Summarizing, conditioned on $\{T_n(Y,g) \le q_n(\alpha_n)\}$

$$E_2 \supseteq \left\{\frac{2d_n}{\lambda}(a_k - a_1) \le \frac{\delta + \sqrt{2\sigma^2\ln(e/\lambda)}}{2\sqrt{\lambda}}\frac{\ln(n)}{\sqrt{n}}\right\} \tag{66}$$

$$= \left\{\frac{\ln(n)}{\sqrt{n}} \le \sqrt{\lambda}\,\frac{\delta + \sqrt{2\sigma^2\ln(e/\lambda)}}{4(a_k - a_1)}\right\}. \tag{67}$$

(42) implies that the right hand side of (66) holds for all $n \ge N^\star$ and in particular, $\mathbf{P}(E_2|T_n(Y,g) \le q_n(\alpha_n)) = 1$ for all $n \ge N^\star$.
Together with (64) this gives $\mathbf{P}(E_1 \cap E_2|T_n(Y,g) \le q_n(\alpha_n)) = 1$ for all $n > N^\star$. This proves the assertion. □

With Theorem S1.2 and Theorem S1.1 the proof of Theorem 2.7 is straight forward.

Proof of Theorem 2.7. Let A_n be as in Theorem S1.2,

$$\mathfrak{T}_\alpha := \{T_n(Y,g) \le q_n(\alpha)\}, \quad \text{and} \quad \hat{\mathfrak{T}}_\alpha := \{T_n(Y,\hat{g}) \le q_n(\alpha)\}.$$

Theorem S1.1 implies that

$$\mathbf{P}\left(\hat{\mathfrak{T}}_{\beta_n}\middle|\,\mathfrak{T}_{\alpha_n}\right) = 1. \tag{68}$$

From (56) we deduce that for β_n as in (23) $q_n(\beta_n) < \delta/(9\sigma)\ln(n)$. Thus, Theorem S1.2 yields

$$\mathbf{P}\left(A_n\middle|\,\mathfrak{T}_{\alpha_n} \cap \hat{\mathfrak{T}}_{\beta_n}\right) = 1. \tag{69}$$

(68) and (69) give

$$\begin{aligned}\mathbf{P}\left(A_n\right) &\geq \mathbf{P}\left(A_n \middle| \mathfrak{T}_{\alpha_n} \cap \hat{\mathfrak{T}}_{\beta_n}\right) \mathbf{P}\left(\mathfrak{T}_{\alpha_n} \cap \hat{\mathfrak{T}}_{\beta_n}\right) \\ &\geq \mathbf{P}\left(\mathfrak{T}_{\alpha_n} \cap \hat{\mathfrak{T}}_{\beta_n}\right) = \mathbf{P}(\mathfrak{T}_{\alpha_n}) \geq 1-\alpha_n.\end{aligned}$$

Finally, remember that the identfiability condition $ASB(\omega) \geq \delta > 0$ implies that g jumps if and only if f jumps. Hence, when f^i and $\hat{f}^i$ take the same function values on constant pieces, results about change points of g directly translate to results about change points of $f^1, \ldots, f^m$. □

S1.3. *Proof of Theorem 2.5.*

Proof. It follows from the proof of Theorem S1.2 that conditioned on $\{T_n(Y,g) \leq q_n(\alpha_n)\}$

$$\max_{a \in \mathrm{Im}(f)} \left|\omega^\top a - \hat{\omega}^\top a\right| \leq \left(\delta \frac{\ln(n)}{\sqrt{n}} + \sqrt{\frac{8\sigma^2 \ln(e/\lambda)}{n\lambda}}\right) \tag{70}$$

and

$$K(\hat{\omega}^\top f) = K(\hat{g}). \tag{71}$$

Let $B(i,j) = [\underline{b}_{ij}, \overline{b}_{ij}]$ be as in (16) and

$$\tilde{B}(i,j) := \left[\underline{b}_{ij} - \left(\delta \frac{\ln(n)}{\sqrt{n}} + \sqrt{\frac{8\sigma^2 \ln(e/\lambda)}{n\lambda}}\right), \overline{b}_{ij} + \left(\delta \frac{\ln(n)}{\sqrt{n}} + \sqrt{\frac{8\sigma^2 \ln(e/\lambda)}{n\lambda}}\right)\right],$$

with $q = q_n(\beta)$ as in (23), then

$$\begin{aligned}&\mathbf{P}\left(f = (f^1, ..., f^m)^\top \in \tilde{\mathcal{H}}(\beta)\right) \\ =&\mathbf{P}\left(\bigcap_{\substack{1 \leq i \leq j \leq n \\ (\hat{\omega}^\top f)|_{[i,j]} \equiv (\hat{\omega}^\top f)_{ij}}} (\hat{\omega}^\top f)_{ij} \in \tilde{B}(i,j) \text{ and } K\left(\hat{\omega}^\top f\right) = K(\hat{g})\right) \\ \geq&\mathbf{P}\left(\bigcap_{\substack{1 \leq i \leq j \leq n \\ g|_{[ij]} \equiv g_{ij}}} g_{ij} \in B(i,j) \text{ and } T_n(Y,g) \leq q_n(\alpha_n)\right) \\ =&\mathbf{P}\left(T_n(Y,g) \leq q_n(\beta)\right) + \mathcal{O}(1),\end{aligned}$$

where the inequality in the third line follows from (70) and (71). Finally, the assertion follows from the fact that $\delta \leq (a_2 - a_1)/m$. □

S2. Algorithms.

S2.1. *Pseudocode for Algorithm CRW.*

Algorithm CRW (Confidence region for weights)

Input: Y, m, $\mathfrak{A}$, α, λ, $\lambda^\star$ ▷ see the SBSSR-model and Remark 2.1
1: $\overline{\mathfrak{B}} \leftarrow \{B(i,j) \in \mathfrak{B} \setminus \mathfrak{B}_{\mathrm{nc}} : j - i + 1 \geq \lambda^\star n\}$ ▷ see R1 and R4
2: $\mathfrak{B}^\star \leftarrow \{[\underline{b}, \overline{b}] \in \overline{\mathfrak{B}} : \overline{b} \geq a_1$ and $\underline{b} \leq a_1 + \frac{a_2 - a_1}{m}\}$ ▷ see R2
3: **for** i=2...m **do**
4: $\mathfrak{B}^\star \leftarrow$

$$\Big\{[\underline{b}_1, \overline{b}_1] \times ... \times [\underline{b}_i, \overline{b}_i] \in \mathfrak{B}^\star \times \overline{\mathfrak{B}} :$$
$$\frac{a_2 + (m-1)a_1 - \sum_{k=1}^{r-1} \underline{b}_k}{m - r + 1} \geq \underline{b}_r \text{ and } \underline{b}_{r-1} \leq \overline{b}_r\Big\}$$

▷ see R2
5: **end for**
6: $\mathfrak{B}^\star \leftarrow \Big\{[\underline{b}_1, \overline{b}_1] \times ... \times [\underline{b}_m, \overline{b}_m] \in \mathfrak{B}^\star : \sum_{j=1}^m \overline{b}_r \geq a_2 + (m-1)a_1\Big\}$ ▷ see R2
7: $\mathfrak{B}^\star \leftarrow$ R3 applied to $\mathfrak{B}^\star$ ▷ see Remark 2.1
8: **return** $\bigcup_{B \in \mathfrak{B}^\star} A^{-1}B$

S2.2. *Computation of* $(\hat{f}^1, \ldots, \hat{f}^m)$. For a given $\beta \in (0,1)$ SLAM solves the constrained optimization problem (25).
Note that $\hat{f}^1, \ldots, \hat{f}^m$ are the unique source functions such that $\sum_{i=1}^m \hat{\omega}_i \hat{f}^i = \hat{g}$ for

$$\hat{g} \coloneqq \mathrm{argmax}_{\tilde{g} \in \mathfrak{H}(\beta)} \sum_{i=1}^n \phi_{\tilde{g}(x_i)}(Y_i), \tag{72}$$

with

$$\mathfrak{H}(\beta) \coloneqq \{\tilde{g} \in \mathcal{S}(\{\hat{\omega}^\top a : a \in \mathfrak{A}^m\}) : T_n\left(Y, \tilde{g}\right) \leq q_n(\beta) \text{ and } K\left(\tilde{g}\right) = \hat{K}\} \tag{73}$$

and $\hat{K}$ as in (24). Frick et al. [1] provide a pruned dynamic programming algorithm how to efficently solve (72) without the restriction that $\hat{g}$ can only attain values in $\{\hat{\omega}^\top a : a \in \mathfrak{A}^m\}$ as it is the case here, see (73). As this restriction is crucial for SLAM we outline the details of the necessary modifications below.
To this end, it is necessary for a finite set $\mathfrak{L} = \{l_1, \ldots, l_k\}$ of possible function values to check finiteness of their *minimal cost* $d^\star_{[i,j]} = \min_{\theta \in \mathbb{R}} d_{[i,j]}$ (see [1, eq. 30]) with $\mathbb{R}$ replaced by $\mathfrak{L}$.

In [1] finiteness of $d^\star_{[i,j]} = \min_{\theta\in\mathbb{R}} d_{[i,j]}$ is examined by the relation

$$\min_{\theta\in\mathbb{R}} d_{[i,j]} = \infty \quad \Leftrightarrow \quad \max_{i\le u\le v\le j} \underline{b}_{uv} > \min_{i\le u\le v\le j} \overline{b}_{uv}, \tag{74}$$

with $\{B(i,j) = [\underline{b}_{ij}, \overline{b}_{ij}] : 1 \le i \le j \le n\}$ as in (16).
Let L be any number such that $L > \max(\mathfrak{L})$ and define $Q(i,j) =$

$$[\underline{q}_{ij}, \overline{q}_{ij}] \coloneqq \begin{cases} [\max(\mathfrak{L}\cap B(i,j)), \min(\mathfrak{L}\cap B(i,j))] & \text{if } \mathfrak{L}\cap B(i,j) \neq \emptyset \\ [L, L] & \text{else} \end{cases}. \tag{75}$$

Then we observe, as in (74), that

$$\min_{\theta\in\mathfrak{L}} d_{[i,j]} = \infty \quad \Leftrightarrow \quad \max_{i\le u\le v\le j} \underline{q}_{uv} > \min_{i\le u\le v\le j} \overline{q}_{uv}. \tag{76}$$

This allows to adapt the dynamic program from [1].
Again, in order to reduce computation time, one can only consider subintervals, e.g., of dyadic length, possibly at the expense of deletion power.

S3. Additional figures and tables.

S3.1. *Additional tables and figure from Section 4.*

TABLE S3.1
Weight vector ω for $m = 2, 3, 4, 5$ such that the $ASB(\omega) = 0.02$.

	$m=2$	$m=3$	$m=4$	$m=5$
ω	$(0.02, 0.98)$	$(0.02, 0.04, 0.94)$	$(0.04, 0.06, 0.12, 0.78)$	$(0.06, 0.08, 0.12, 0.16, 0.58)$

TABLE S3.2
Influence of the number of source functions m for $m = 2, 3, 4, 5$.

	$m=2$	$m=3$	$m=4$	$m=5$
$\mathrm{MAE}(\hat{\omega})$ $[10^{-4}]$	$(1, 1)$	$(11, 18, 24)$	$(90, 154, 62, 69)$	$(91, 68, 81, 196, 54)$
$\overline{\mathrm{dist}}(\omega, \mathcal{C}_{1-\alpha})$ $[10^{-3}]$	11	23	63	54
$\mathrm{Mean}(\omega \in \mathcal{C}_{1-\alpha})$ [%]	100	99.99	99.96	100
$\overline{\omega}_i - \underline{\omega}_i$ $[10^{-3}]$	$(21, 21)$	$(37, 33, 23)$	$(68, 93, 35, 23)$	$(40, 55, 84, 63, 23)$
$\mathrm{MIAE}(\hat{f}^i)$ $[10^{-3}]$	$(0.2, 0.0)$	$(26, 9, 0.0)$	$(115, 103, 67, 0.0)$	$(315, 317, 49, 183, 0.0)$
$\mathrm{Mean}(\hat{K}) - K$	$(0, 0)$	$(0.22, -0.03, 0)$	$(3.7, 2.6, -0.6, 0)$	$(2.75, 2.28, 0.75, -1.61, 0)$
$\mathrm{Med}(\hat{K}) - K$	$(0, 0)$	$(0, 0, 0)$	$(4, 2, 0, 0)$	$(2, 2, 0, -2, 0)$
$\mathrm{Mean}(\hat{K} = K)_i$ [%]	$(99.8, 99.8)$	$(88.5, 98, 100)$	$(15.9, 31, 69.4, 100)$	$(7.1, 30.4, 63.8, 12, 99.9)$
$\mathrm{Mean}(\hat{K} = K)$ [%]	99.8	87.2	15.8	1
$\max_i \min_j \lvert\tau_i - \hat{\tau}_j\rvert$	$(0.37, 0.02)$	$(33.82, 4.77, 0.00)$	$(245.49, 95.75, 2.52, 0.00)$	$(374.38, 208.32, 40.12, 7.41, 0.02)$
$\max_j \min_i \lvert\tau_i - \hat{\tau}_j\rvert$	$(0.03, 0.00)$	$(18.59, 12.53, 0.000)$	$(9.61, 18.66, 126.33, 0.00)$	$(83.09, 117.17, 61.13, 348.89, 0.00)$
V_1 [%]	$(99.9, 100)$	$(88.3, 96.2, 100)$	$(60.9, 83.4, 68.6, 100)$	$(37.5, 54.1, 82.8, 12.6, 100)$
FPSLE	$(0.07, 0.00)$	$(8.98, 6.05, 0.00)$	$(51.52, 21.36, 78.23, 0.00)$	$(110.3, 92.21, 34.98, 216.82, 0.00)$
FNSLE	$(0.3, 0.02)$	$(24.04, 3.22, 0.00)$	$(168.04, 45.09, 62.15, 0.00)$	$(205.75, 137.64, 41.29, 90.02, 0.02)$
$\mathrm{Mean}(f \in \tilde{\mathcal{H}}(\beta))$ [%]	99.93	99.49	98.77	91.08

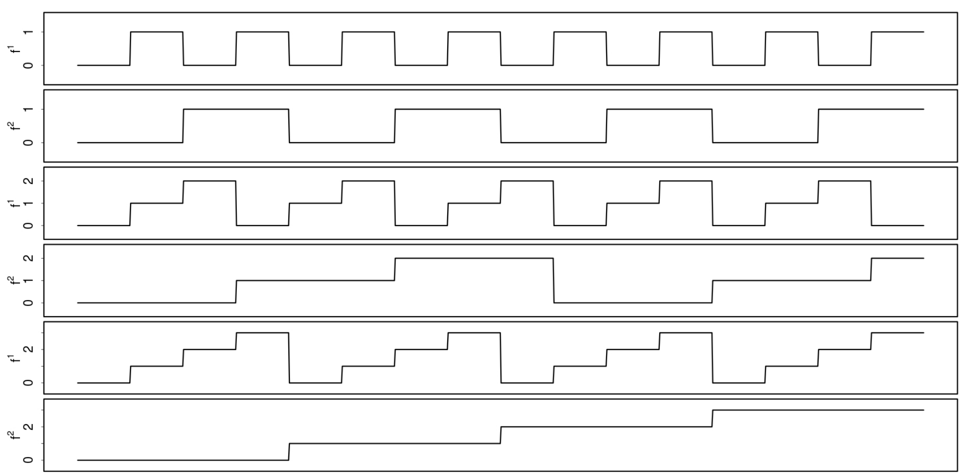

FIG S3.1. *f^1 and f^2 from (43) in Section 4.2 for $\mathfrak{A} = \{0, 1\}, \{0, 1, 2\}$, and $\{0, 1, 2, 3\}$ (from top to bottom).*

Table S3.3
Influence of the number of alphabet values k for $k = 2, 3, 4$.

	$k = 2$	$k = 3$	$k = 4$
$\mathrm{MAE}(\hat{\omega})$ $[10^{-3}]$	$(19, 12)$	$(18, 12)$	$(15, 11)$
$\overline{\mathrm{dist}}(\omega, \mathcal{C}_{1-\alpha})$ $[10^{-3}]$	51	51	47
$\mathrm{Mean}(\omega \in \mathcal{C}_{1-\alpha})$ [%]	100	100	100
$\overline{\omega}_i - \underline{\omega}_i$ $[10^{-3}]$	$(71, 71)$	$(71, 71)$	$(67, 67)$
$\mathrm{MIAE}(\hat{f}^i)$ $[10^{-2}]$	$(29, 0)$	$(49, 0)$	$(60, 0)$
$\mathrm{Mean}(\hat{K}) - K$	$(-6.65, 0)$	$(-7.42, 0)$	$(-7.04, 0)$
$\mathrm{Med}(\hat{K}) - K$	$(-6, 0)$	$(-7, 0)$	$(-7, 0)$
$\mathrm{Mean}(\hat{K} = K)_i$ [%]	$(0.39, 99.99)$	$(0, 100)$	$(0, 100)$
$\mathrm{Mean}(\hat{K} = K)$ [%]	0.39	0	0
$\max_i \min_j \lvert\tau_i - \hat{\tau}_j\rvert$	$(17.5, 0.0)$	$(22.0, 0.0)$	$(23.31, 0.00)$
$\max_j \min_i \lvert\tau_i - \hat{\tau}_j\rvert$	$(96.0, 0.0)$	$(134.4, 0.0)$	$(79.8, 0.0)$
V_1 [%]	$(81.7, 100)$	$(78, 100)$	$(81.5, 100)$
FPSLE	$(0.4, 0.0)$	$(58.3, 0.0)$	$(37.2, 0.0)$
FNSLE	$(25.7, 0.0)$	$(29.3, 0.0)$	$(25.2, 0.0)$
$\mathrm{Mean}(f \in \tilde{\mathcal{H}}(\beta))$ [%]	94.60	98.49	98.60

Table S3.4
Influence of the confidence level α on $\hat{\omega}$ and $\mathcal{C}_{1-\alpha}$ for $\alpha = 0.01, 0.05, 0.1$.

$\sigma = 0.02$			
	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$
$\mathrm{MAE}(\hat{\omega})$ $[10^{-3}]$	$(2, 2, 2)$	$(1, 1, 1)$	$(1, 1, 1)$
$\overline{\mathrm{dist}}(\omega, \mathcal{C}_{1-\alpha})$ $[10^{-3}]$	29	25	24
$\mathrm{Mean}(\omega \in \mathcal{C}_{1-\alpha})$ [%]	100	100	100
$\overline{\omega}_i - \underline{\omega}_i$ $[10^{-3}]$	$(48, 46, 44)$	$(43, 42, 42)$	$(42, 42, 42)$
$\sigma = 0.05$			
	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$
$\mathrm{MAE}(\hat{\omega})$ $[10^{-3}]$	$(22, 7, 16)$	$(23, 7, 16)$	$(22, 7, 16)$
$\overline{\mathrm{dist}}(\omega, \mathcal{C}_{1-\alpha})$ $[10^{-3}]$	109	105	102
$\mathrm{Mean}(\omega \in \mathcal{C}_{1-\alpha})$ [%]	100	100	99
$\overline{\omega}_i - \underline{\omega}_i$ $[10^{-3}]$	$(168, 123, 115)$	$(160, 112, 106)$	$(155, 107, 102)$
$\sigma = 0.1$			
	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$
$\mathrm{MAE}(\hat{\omega})$ $[10^{-3}]$	$(59, 51, 13)$	$(45, 48, 13)$	$(32, 43, 18)$
$\overline{\mathrm{dist}}(\omega, \mathcal{C}_{1-\alpha})$ $[10^{-3}]$	231	218	210
$\mathrm{Mean}(\omega \in \mathcal{C}_{1-\alpha})$ [%]	100	100	100
$\overline{\omega}_i - \underline{\omega}_i$ $[10^{-3}]$	$(329, 344, 282)$	$(305, 323, 226)$	$(276, 312, 212)$

TABLE S3.5
Influence of the confidence levels α and β on $\hat{f}$ and $\tilde{\mathcal{H}}(\beta)$ for each $(\alpha,\beta) \in \{0.01, 0.05, 0.1\}^2$, for $\sigma = 0.02, 0.05, 0.1$. In the displayed matrices α increases within a column and β increases within a row.

$\sigma = 0.02$			
	f^1	f^2	f^3
MIAE$(\hat{f}^i)$ $[10^{-4}]$	$\begin{pmatrix} 0 & 2 & 10 \\ 0 & 2 & 10 \\ 0 & 2 & 10 \end{pmatrix}$	$\begin{pmatrix} 6 & 3 & 11 \\ 9 & 5 & 12 \\ 11 & 7 & 13 \end{pmatrix}$	$\begin{pmatrix} 3 & 1 & 4 \\ 5 & 2 & 4 \\ 6 & 3 & 5 \end{pmatrix}$
Med$(\hat{K}) - K$	$\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$
Mean$(\hat{K} = K)_i$ [%]	$\begin{pmatrix} 100 & 100 & 100 \\ 100 & 100 & 100 \\ 100 & 100 & 100 \end{pmatrix}$	$\begin{pmatrix} 98 & 100 & 100 \\ 97 & 99 & 99 \\ 96 & 98 & 99 \end{pmatrix}$	$\begin{pmatrix} 99 & 100 & 100 \\ 98 & 99 & 99 \\ 97 & 99 & 99 \end{pmatrix}$
Mean$(\hat{K} = K)$ [%]		$\begin{pmatrix} 98 & 99 & 100 \\ 97 & 99 & 99 \\ 96 & 98 & 99 \end{pmatrix}$	
V_1 [%]	$\begin{pmatrix} 100 & 100 & 100 \\ 100 & 100 & 100 \\ 100 & 100 & 100 \end{pmatrix}$	$\begin{pmatrix} 100 & 100 & 100 \\ 100 & 100 & 100 \\ 100 & 100 & 100 \end{pmatrix}$	$\begin{pmatrix} 100 & 100 & 100 \\ 100 & 100 & 100 \\ 100 & 100 & 100 \end{pmatrix}$
Mean$(f \in \tilde{\mathcal{H}}(\beta))$ [%]		$\begin{pmatrix} 95.8 & 93.3 & 92.3 \\ 99.0 & 97.7 & 97.0 \\ 99.2 & 98.6 & 98.1 \end{pmatrix}$	
Mean$(f^i \in \tilde{\mathcal{H}}(\beta)_i)$ [%]	$\begin{pmatrix} 99.90 & 99.74 & 99.34 \\ 99.94 & 99.78 & 99.64 \\ 99.90 & 99.70 & 99.68 \end{pmatrix}$	$\begin{pmatrix} 99.84 & 99.60 & 99.38 \\ 99.92 & 99.84 & 99.74 \\ 99.90 & 99.82 & 99.74 \end{pmatrix}$	$\begin{pmatrix} 96.68 & 95.46 & 94.92 \\ 99.18 & 98.34 & 98.10 \\ 99.42 & 99.02 & 98.64 \end{pmatrix}$
$\sigma = 0.05$			
	f^1	f^2	f^3
MIAE$(\hat{f}^i)$ $[10^{-3}]$	$\begin{pmatrix} 6 & 7 & 8 \\ 6 & 8 & 9 \\ 6 & 8 & 9 \end{pmatrix}$	$\begin{pmatrix} 160 & 161 & 160 \\ 164 & 165 & 164 \\ 160 & 161 & 161 \end{pmatrix}$	$\begin{pmatrix} 80 & 80 & 80 \\ 82 & 83 & 82 \\ 80 & 80 & 80 \end{pmatrix}$
Med$(\hat{K}) - K$	$\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 2 & 2 & 2 \\ 2 & 2 & 2 \\ 2 & 2 & 2 \end{pmatrix}$	$\begin{pmatrix} -2 & -2 & -2 \\ -2 & -2 & -2 \\ -2 & -2 & -2 \end{pmatrix}$
Mean$(\hat{K} = K)_i$ [%]	$\begin{pmatrix} 96 & 90 & 85 \\ 93 & 86 & 80 \\ 93 & 85 & 80 \end{pmatrix}$	$\begin{pmatrix} 21 & 19 & 17 \\ 19 & 16 & 15 \\ 21 & 19 & 17 \end{pmatrix}$	$\begin{pmatrix} 24 & 25 & 27 \\ 21 & 23 & 24 \\ 24 & 25 & 26 \end{pmatrix}$
Mean$(\hat{K} = K)$ [%]		$\begin{pmatrix} 19 & 16 & 14 \\ 17 & 14 & 12 \\ 19 & 16 & 14 \end{pmatrix}$	
V_1 [%]	$\begin{pmatrix} 99 & 99 & 99 \\ 99 & 99 & 99 \\ 99 & 99 & 99 \end{pmatrix}$	$\begin{pmatrix} 92 & 92 & 92 \\ 92 & 92 & 92 \\ 92 & 92 & 92 \end{pmatrix}$	$\begin{pmatrix} 91 & 91 & 91 \\ 91 & 91 & 91 \\ 91 & 91 & 91 \end{pmatrix}$
Mean$(f \in \tilde{\mathcal{H}}(\beta))$ [%]		$\begin{pmatrix} 83.1 & 76.7 & 74.0 \\ 81.3 & 75.6 & 73.4 \\ 81.7 & 76.4 & 74.5 \end{pmatrix}$	
Mean$(f^i \in \tilde{\mathcal{H}}(\beta)_i)$ [%]	$\begin{pmatrix} 100 & 100 & 100 \\ 100 & 100 & 99.98 \\ 100 & 100 & 99.98 \end{pmatrix}$	$\begin{pmatrix} 89.34 & 84.78 & 82.82 \\ 86.60 & 83.04 & 83.18 \\ 87.24 & 84.16 & 83.18 \end{pmatrix}$	$\begin{pmatrix} 85.80 & 80.56 & 78.34 \\ 83.14 & 78.48 & 77.14 \\ 83.58 & 79.48 & 78.16 \end{pmatrix}$
$\sigma = 0.1$			
	f^1	f^2	f^3
MIAE$(\hat{f}^i)$ $[10^{-3}]$	$\begin{pmatrix} 327 & 327 & 327 \\ 297 & 296 & 296 \\ 255 & 254 & 253 \end{pmatrix}$	$\begin{pmatrix} 245 & 246 & 246 \\ 233 & 234 & 234 \\ 231 & 232 & 232 \end{pmatrix}$	$\begin{pmatrix} 90 & 91 & 91 \\ 67 & 68 & 68 \\ 75 & 76 & 76 \end{pmatrix}$
Med$(\hat{K}) - K$	$\begin{pmatrix} 2 & 3 & 3 \\ 1 & 2 & 2 \\ 1 & 1 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$
Mean$(\hat{K} = K)_i$ [%]	$\begin{pmatrix} 12 & 9 & 7 \\ 22 & 19 & 17 \\ 36 & 32 & 29 \end{pmatrix}$	$\begin{pmatrix} 15 & 12 & 11 \\ 24 & 22 & 21 \\ 35 & 33 & 32 \end{pmatrix}$	$\begin{pmatrix} 44 & 37 & 34 \\ 62 & 53 & 49 \\ 59 & 52 & 48 \end{pmatrix}$
Mean$(\hat{K} = K)$ [%]		$\begin{pmatrix} 4 & 2 & 1 \\ 7 & 5 & 4 \\ 8 & 7 & 6 \end{pmatrix}$	
V_1 [%]	$\begin{pmatrix} 85 & 85 & 85 \\ 86 & 86 & 86 \\ 88 & 87 & 87 \end{pmatrix}$	$\begin{pmatrix} 74 & 74 & 75 \\ 73 & 74 & 74 \\ 75 & 76 & 76 \end{pmatrix}$	$\begin{pmatrix} 95 & 95 & 95 \\ 97 & 97 & 97 \\ 96 & 96 & 96 \end{pmatrix}$
Mean$(f \in \tilde{\mathcal{H}}(\beta))$ [%]		$\begin{pmatrix} 60.7 & 58.6 & 55.7 \\ 71.0 & 63.5 & 63.2 \\ 80.2 & 71.0 & 66.9 \end{pmatrix}$	
Mean$(f^i \in \tilde{\mathcal{H}}(\beta)_i)$ [%]	$\begin{pmatrix} 90.4 & 89.6 & 89.3 \\ 99.0 & 98.8 & 98.8 \\ 99.7 & 99.6 & 99.6 \end{pmatrix}$	$\begin{pmatrix} 96.7 & 91.5 & 86.0 \\ 97.8 & 95.0 & 94.3 \\ 97.9 & 95.2 & 92.9 \end{pmatrix}$	$\begin{pmatrix} 72.8 & 74.6 & 77.0 \\ 83.5 & 80.2 & 79.4 \\ 90.1 & 86.2 & 85.6 \end{pmatrix}$

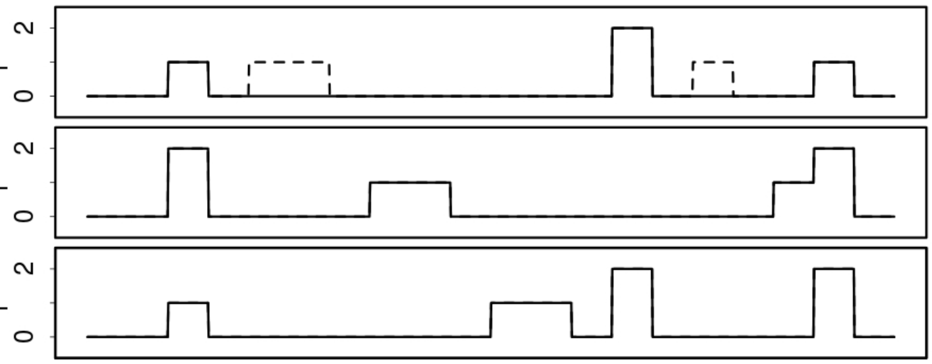

FIG S3.2. *Source functions f from Example 1.1 modified such they violate the separability condition in (7) for $r = 1$ (solid line). The dotted lines indicate the removed jumps.*

TABLE S3.6
Result illustrating robustness. (1): Setting as in Example 1.1 but with f modified such it violates the separability condition in (7) (see Figure S3.2). (2): Setting as in Example 1.1, but with t-distributed errors with 3 degrees of freedom. (3): Setting as in Example 1.1, but with χ^2-distributed errors with 3 degrees of freedom.

	(1)	(2)	(3)
MAE($\hat{\omega}$) [10^{-3}]	$(73, 36, 39)$	$(43, 58, 16)$	$(42, 59, 17)$
MIAE($\hat{f}^i$) [10^{-3}]	$(123, 181, 84)$	$(447, 435, 137)$	$(563, 279, 99)$
Med($\hat{K}$) $- K$	$(-4, 2, 0)$	$(4, 1, -2)$	$(11, 4, -2)$
Mean($\hat{K} = K$)$_i$ [%]	$(10, 10, 19)$	$(5, 0, 33)$	$(2, 1, 4)$
V_1 [%]	$(71, 85, 96)$	$(84, 72, 88)$	$(78, 82, 89)$

Table S3.7

Results illustrating the influence of the alphabet separation boundary $ASB = ASB(\omega)$ on $\hat{\omega}$ with $\omega \sim \mathcal{U}(\Omega(m))$.

	MAE($\hat{\omega}$) $[10^{-3}]$	$\overline{\text{dist}}(\omega, \mathcal{C}_{1-\alpha})$ $[10^{-3}]$
$0 \leq ASB \leq 0.0001$	$(6, 4, 5)$	29
$0.0001 \leq ASB \leq 0.01$	$(7, 4, 7)$	34
$0.01 \leq ASB \leq 0.02$	$(4, 4, 4)$	30
$0.02 \leq ASB \leq 0.03$	$(4, 4, 4)$	29
$0.03 \leq ASB \leq 0.04$	$(4, 3, 4)$	31
$0.04 \leq ASB \leq 0.05$	$(4, 3, 4)$	31
$0.05 \leq ASB \leq 0.06$	$(4, 3, 5)$	31
$0.06 \leq ASB \leq 0.07$	$(3, 3, 4)$	31

Table S3.8

Influence of the alphabet separation boundary $ASB = ASB(\omega)$ on $\hat{f}$ with $\omega \sim \mathcal{U}(\Omega(m))$.

	MIAE($\hat{f}^i$) $[10^{-4}]$	$\lvert\tilde{\mathcal{H}}_x(0.1)\rvert$		
		mean	median	
$0 \leq ASB \leq 0.0001$	$(1916, 1067, 483)$	2.71	3	$0 \leq ASB_x \leq 0.001$
$0.0001 \leq ASB \leq 0.01$	$(1536, 923, 354)$	2.68	3	$0.001 \leq ASB_x \leq 0.01$
$0.01 \leq ASB \leq 0.02$	$(671, 474, 147)$	2.67	3	$0.01 \leq ASB_x \leq 0.02$
$0.02 \leq ASB \leq 0.03$	$(236, 164, 40)$	2.66	3	$0.02 \leq ASB_x \leq 0.03$
$0.03 \leq ASB \leq 0.04$	$(96, 37, 7)$	2.53	2	$0.03 \leq ASB_x \leq 0.04$
$0.04 \leq ASB \leq 0.05$	$(100, 7, 2)$	2.49	2	$0.04 \leq ASB_x \leq 0.05$
$0.05 \leq ASB \leq 0.06$	$(42, 1, 0)$	2.36	2	$0.05 \leq ASB_x \leq 0.1$
$0.06 \leq ASB \leq 0.07$	$(16, 4, 0)$	1.97	1	$0.1 \leq ASB_x$

Table S3.9

Influence of prior information on λ for prior knowledge $\lambda \geq 0.05, 0.04, 0.025, 0.015, 0.005$.

Prior knowledge $\lambda \geq$	0.05	0.04	0.025	0.015	0.005
MAE($\hat{\omega}$) $[10^{-3}]$	$(6, 5, 3)$	$(2, 2, 1)$	$(2, 2, 1)$	$(5, 5, 6)$	$(159, 126, 186)$
$\overline{\text{dist}}(\omega, \mathcal{C}_{1-\alpha})$ $[10^{-3}]$	17	23	23	37	123
Mean($\omega \in \mathcal{C}_{1-\alpha}$) [%]	100	100	100	100	100
$\overline{\omega}_i - \underline{\omega}_i$ $[10^{-3}]$	$(24, 25, 25)$	$(42, 42, 42)$	$(42, 42, 42)$	$(65, 64, 63)$	$(183, 171, 144)$
MIAE($\hat{f}^i$) $[10^{-3}]$	$(3, 13, 6)$	$(1, 4, 2)$	$(1, 4, 2)$	$(1, 23, 11)$	$(40, 175, 88)$
Mean($\hat{K}$) $- K$	$(0.1, 0.2, 0.0)$	$(0.1, 0.1, 0.0)$	$(0.1, 0.1, 0.0)$	$(0.0, 0.3, -0.1)$	$(2.4, 2.5, -0.2)$
Med($\hat{K}$) $- K$	$(0, 0, 0)$	$(0, 0, 0)$	$(0, 0, 0)$	$(0, 0, 0)$	$(0, -2, -2)$
Mean($\hat{K} = K)_i$ [%]	$(99, 93, 97)$	$(100, 98, 99)$	$(100, 98, 99)$	$(99, 87, 93)$	$(54, 24, 16)$
Mean($\hat{K} = K$) [%]	93	98	98	86	6
$\max_i \min_j \lvert\tau_i - \hat{\tau}_j\rvert$ $[10^{-1}]$	$(13, 148, 4)$	$(6, 40, 2)$	$(6, 40, 2)$	$(7, 299, 9)$	$(508, 1794, 122)$
$\max_j \min_i \lvert\tau_i - \hat{\tau}_j\rvert$ $[10^{-1}]$	$(2, 41, 50)$	$(1, 11, 15)$	$(1, 11, 15)$	$(1, 45, 91)$	$(223, 331, 1343)$
V_1 [%]	$(100, 99, 100)$	$(100, 100, 100)$	$(100, 100, 100)$	$(100, 98, 99)$	$(96, 89, 91)$
FPSLE $[10^{-2}]$	$(16, 246, 167)$	$(8, 67, 51)$	$(8, 67, 51)$	$(5, 398, 304)$	$(708, 1994, 4491)$
FNSLE $[10^{-2}]$	$(34, 407, 41)$	$(17, 113, 14)$	$(17, 113, 14)$	$(16, 785, 71)$	$(1610, 5786, 1168)$
Mean($f \in \tilde{\mathcal{H}}(\beta)$) [%]	96.01	98.96	98.95	94.78	56.65

S3.2. *Additional figures from Section 5.*

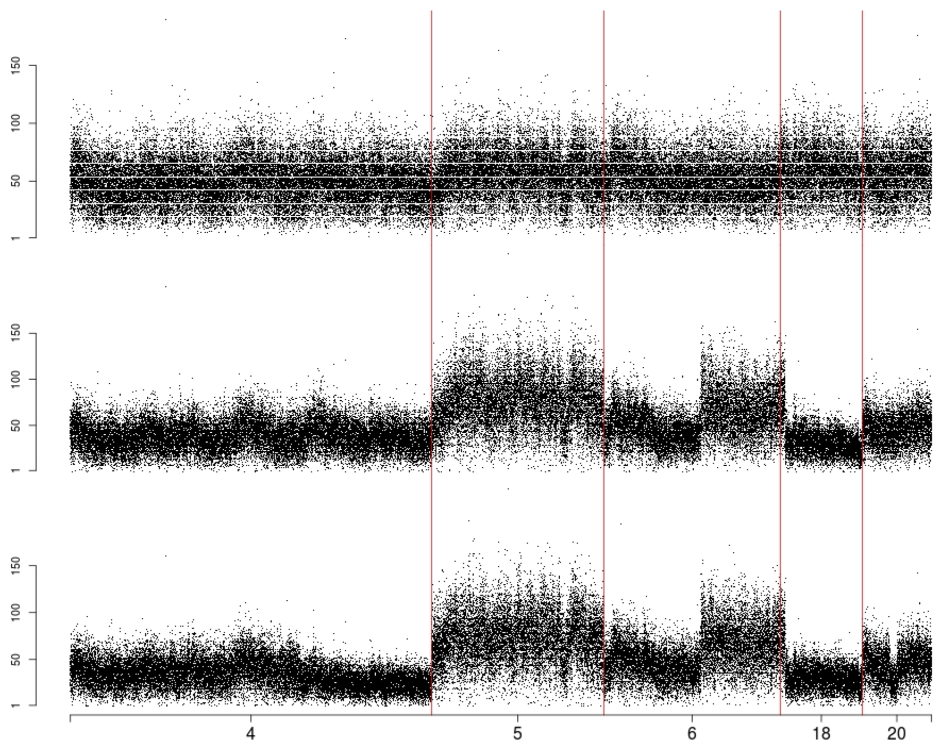

Fig S3.3. *Raw whole genome sequencing data from cell line LS411*

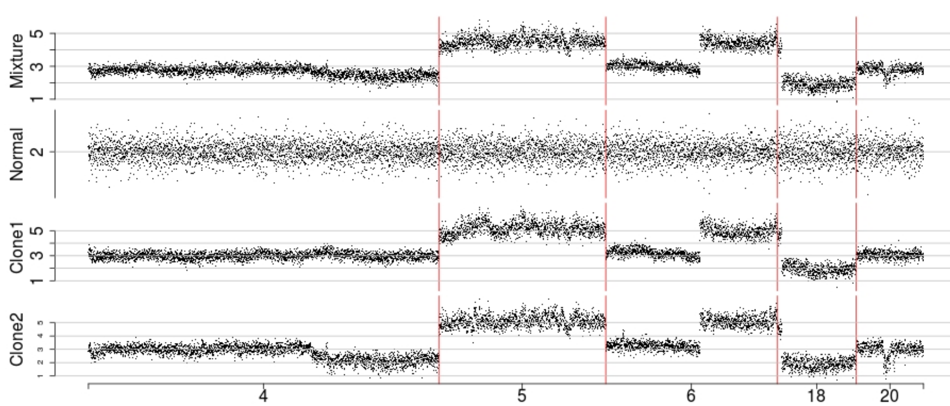

Fig S3.4. *Preprocessed whole genome sequencing data from cell line LS411*

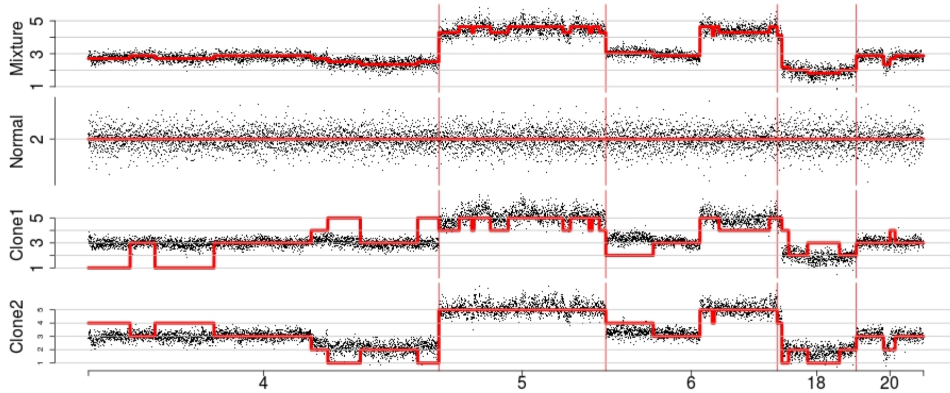

FIG S3.5. *SLAM's estimates (red lines) for $q_n(\alpha) = -0.15$ (selected with MVT-method from Section 4.6) and $q_n(\beta) = 2.2$. Top row: total copy-number estimates across the genome. Rows 2-4: estimates of the CN profiles of the germline and clones.*

S4. Data driven selection of $q_n(\alpha)$. In the following we give further details on the SST-method for selection of $q_n(\alpha)$ introduced in Section 4.6. To simplify notation let n be even. Then $Y^1 := (Y_1, Y_3, \ldots, Y_{n-1})$ and $Y^2 := (Y_2, Y_4, \ldots, Y_n)$ are both samples of size $n/2$ from the same underlying mixture g, with corresponding estimates $\hat{\omega}_q^1 := \hat{\omega}(Y^1, q)$ and $\hat{\omega}_q^2 := \hat{\omega}(Y^2, q)$, respectively. Let L be a loss function and $h(q) := \mathrm{E}[L(\hat{\omega}_q - \omega)]$ its corresponding performance measure for estimating ω, e.g., the MSE with $L = \|\cdot\|_2^2$, which is to be minimized. As ω is unknown, $h(q)$ has to be estimated. This is done by

$$\hat{h}(q) := \frac{1}{2}\left(L(\hat{\omega}_q - \hat{\omega}_q^1) + L(\hat{\omega}_q - \hat{\omega}_q^2)\right)$$

and we estimate the minimizing q of h as

$$\hat{q} := \mathrm{argmin}_{q \leq q_0} \hat{h}(q). \tag{77}$$

Bounding q from above by q_0 is necessary as for $q \to \infty$, i.e. $\alpha \to 0$, the corresponding confidence region $\mathcal{C}_{1-\alpha}$ converges to the entire domain $\Omega(m)$, hence $h(q) \to 0$ as $q \to \infty$. We found empirically that $q_0 := q_n(0.01)$ serves as a good bound (as statements with higher confidence as 0.99 are rarely demanded), also to reduce computation time for the optimization of (77).

The performance of the selector in (77) is illustrated for the setting of Example 1.1 (with $n = 1280$ and $\sigma = 0.05$) in Figure S4.1 for the MSE(q) ($L = \|\cdot\|_2^2$) and the MAE(q) ($L = \|\cdot\|_1$), respectively. From this we find that the optimal q (the minimizer of the black line) is quite well approximated by its estimate $\hat{q} \approx 0.5$ (the minimizer of the red line). Simulations

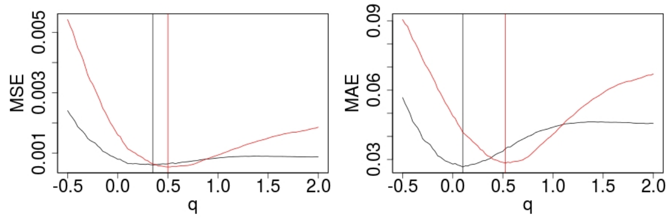

FIG S4.1. *Pointwise mean of 1,000 replications of* $\|\hat{\omega}_q - \omega\|_2$, $\|\hat{\omega}_q - \omega\|_1$, *respectively (black) and of* $\left(\|\hat{\omega}_q^1 - \hat{\omega}_q\|_2 + \|\hat{\omega}_q^2 - \hat{\omega}_q\|_2\right)$, $\left(\|\hat{\omega}_q^1 - \hat{\omega}_q\|_1 + \|\hat{\omega}_q^2 - \hat{\omega}_q\|_1\right)$, *respectively (red) (from left to right), for the setting as in Example 1.1 with* $n = 1280$ *and* $\sigma = 0.05$. *The vertical lines indicate the corresponding minima.*

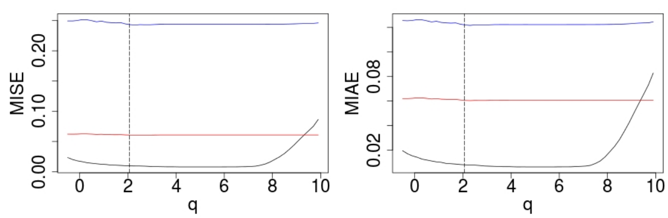

FIG S4.2. *MISE and MIAE of* $\hat{f}_q^1$ *(black),* $\hat{f}_q^2$ *(blue), and* $\hat{f}_q^3$ *(red) observed from* $1,000$ *realizations for the setting as in Example 1.1 with* $n = 1280$ *and* $\sigma = 0.05$. *The vertical dotted line indicates* $q_n(0.01) = 2.07$.

for different n and σ with $\sigma/\sqrt{n}$ in the order of our application example (see Section 4) show the same. Recall from the previous Section 4.3 that ω is estimated quite stable for a range of q. In Figure S4.1 $q \approx 0.5$ corresponds to $\alpha \approx 0.69$. The optimal q for the MSE is $q \approx 0.35$, corresponding to $\alpha \approx 0.81$ and for the MAE $q \approx 0.1$, corresponding to $\alpha \approx 0.95$.

For large noise levels, however, we found that the SST-selection method is outperformed by the MVT-method from Section 4.6 illustrated for the setting of Example 1.1 with $n = 1280$ and $\sigma = 0.05, 0.08, 0.1, 0.2$ in Table S4.1.

TABLE S4.1
MSE and MAE for the SST-method and the MVT-method for the setting of Example 1.1 with $n = 1280$ *and* $\sigma = 0.05, 0.08, 0.1, 0.2$ *obtained from* $2,000$ *replications.*

	MSE $[10^{-4}]$		MAE $[10^{-3}]$	
	SST	MVT	SST	MVT
$\sigma = 0.05$	4	4	27	18
$\sigma = 0.08$	26	34	73	81
$\sigma = 0.1$	56	30	110	78
$\sigma = 0.2$	166	44	206	95

References.

[1] Frick, K., Munk, A. and Sieling, H. (2014). Multiscale change point inference. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **76** 495–580. With discussion and rejoinder.

[2] Sieling, H. (2013). Statistical multiscale segmentation: inference, algorithms and applications PhD thesis, Universität Göttingen.

University of Goettingen
Institute for Mathematical Stochastics
Goldschmidtstr. 7
37077 Göttingen
Germany
E-mail: behr@math.uni-goettingen.de
E-mail: munk@math.uni-goettingen.de

University of Oxford
Department of Statistics
24-29 St Giles'
Oxford. OX1 3LB
United Kingdom
E-mail: cholmes@stats.ox.ac.uk