ROBUST SAMPLING DESIGNS FOR A POSSIBLY MISSPECIFIED STOCHASTIC PROCESS

Yassir Rabhi and Douglas P. Wiens

University of Alberta

Abstract: We address the problem of finding robust sampling designs for the estimation of a discrete time second-order stationary process when its autocorrelation function is only approximately specified and has a spectral density belonging to a neighbourhood of a specified 'base' density. The value of the stochastic process is predicted by the best – for the assumed autocorrelation function – linear unbiased predictor on the basis of a finite sample of observations. Following the approach of minimax robustness, we find the least favourable – in the sense of maximizing the average mean squared error (AMSPE) of these predictions – spectral density. We then obtain, through a genetic algorithm, robust sampling designs which minimize this maximum AMSPE. Several examples are discussed and assessed, on the basis of which we conclude that the robust designs can offer substantial protection against model errors, at a minimal cost in efficiency at the base model. The techniques are illustrated in a case study, using a series of interest in statistical climatology.

Key words and phrases: Average mean squared error, best linear unbiased predictor, climatology, genetic algorithm, glacial varves, least favourable spectral density, minimax robustness.

1. Introduction

Consider the problem of interpolating or forecasting the values $X(t)$ of a stochastic process, evolving over time, at discrete time points $t = 0, \pm 1, \pm 2, \ldots$. Suppose that attention is restricted to a finite interval $[1, N]$ of time points, and that the investigator is able to sample the data at n distinct times $\{t_1 < \cdots < t_n\} \subset [1, N_0]$. Here n, N_0, and N are fixed, $n \leq N_0 \leq N$. The goal is accurate prediction of $\{X(t) | t = 1, \ldots, N\}$. A complicating factor motivating this study is that, in such a situation, the experimenter must typically operate with at best only approximate knowledge of the mechanisms generating the evolution of the process.

We assume that $\{X(t)\}$ is second-order stationary, with mean zero. The experimenter takes the 'working' autocovariance function

$$R_0(s) = \mathrm{COV}\left[X(t \pm s), X(t)\right] = \int_{-\pi}^{\pi} \cos(s\,\omega) f_0(\omega)\, d\omega,$$

for a particular 'base' spectral density $f_0(\cdot)$. If $R_0(\cdot)$ is correctly specified then, given a sample $\boldsymbol{x}_n = (X(t_1), \ldots, X(t_n))'$, the best linear unbiased predictor (BLUP), minimizing the mean squared prediction error (MSPE), is, in the notation of (1.2),

$$\widehat{X}(t) = \boldsymbol{x}_n' \, \boldsymbol{A}_n^{-1} \boldsymbol{a}_n(t). \tag{1.1}$$

Other methods of signal estimation have recently been investigated by Wein and Srinivasan (2013).

Here vectors and matrices with a single index n have dimension n, and depend on the sampling design. We take

$$\boldsymbol{A}_n = (R_0(|t_j - t_k|))_{1 \leq j,k \leq n}, \tag{1.2a}$$

$$\boldsymbol{a}_n(t) = (R_0(|t - t_j|))_{1 \leq j \leq n}, \tag{1.2b}$$

so that $\boldsymbol{A}_n = [\boldsymbol{a}_n(t_1), \cdots, \boldsymbol{a}_n(t_n)]$.

Mukherjee (2003) studied problems of a form close to ours, and details a number of possible applications in geostatistics (Matheron (1963), Journel and Huijbregts (1978)), Bayesian numerical analysis (Diaconis (1988)) and environmental monitoring (Christakos (1992)). She assumes a known covariance function $R_0(\cdot)$ of a particularly tractable 'product' form, allowing her to obtain explicit solutions and designs. Here we seek robustness against a possibly misspecified covariance function, and so entertain the possibility that the true covariances may be of the form

$$R(s) = \int_{-\pi}^{\pi} \cos(s\,\omega) \, f(\omega) \, d\omega, \tag{1.3}$$

with f different from, but 'close to', f_0.

Hosoya (1978) considered problems of a similar nature, in which optimal *one-step ahead* predictions were sought, with the MSPE first maximized over an ε-contamination neighbourhood of the spectral distribution corresponding to f_0: $F(\omega) = (1 - \varepsilon) \, F_0(\omega) + \varepsilon G(\omega)$. He found that the maximizing G has a density which is uniform on its support. Here we are concerned instead with the *average mspe* (AMSPE) – averaged over all of $[1, N]$ – and first maximize this over an L_2 neighbourhood

$$\mathcal{F}_\varepsilon = \left\{ f \; \middle| \; \begin{array}{l} f \geq 0, \; \int_{-\pi}^{\pi} (f(\omega) - f_0(\omega)) \, d\omega = 0 \text{ and} \\ \|f - f_0\|_{L_2}^2 = \int_{-\pi}^{\pi} (f(\omega) - f_0(\omega))^2 \, d\omega \leq \varepsilon^2 \end{array} \right\}$$

of f_0. The second condition in the definition of $\mathcal{F}_\varepsilon$ fixes the process variance, which we take as $\int_{-\pi}^{\pi} f_0(\omega) d\omega = 1$ without loss of generality.

Franke (1985) worked with $\mathcal{F}_\varepsilon$ (without its second condition) and found a member f_* which satisfied an information-theoretic property of being *most*

indeterministic; he went on to study the robustness properties of the BLUP, for one-step ahead prediction, computed assuming that the autocovariances were generated by f_*. In a regression framework a somewhat analogous approach is that of Samarov (1987). Our interest is in seeking robustness through the choice of the *sampling design*, should the model be misspecified. The robustness is to be realized through an optimal bound on the increased AMSPE, in a neighbourhood of $f_0(\cdot)$. The often quoted analogy (Anscombe (1960)) of robustness methods with insurance risks and premiums is apt – the investigator pays a premium, in terms of lost efficiency at $f_0(\cdot)$, through the use of a design not quite optimal for $f_0(\cdot)$. In return he receives the protection of reduced losses when $f_0(\cdot)$ is not quite correct.

In §2 we develop the AMSPE and then exhibit the least favourable spectral density, maximizing the AMSPE, in $\mathcal{F}_\varepsilon$. This maximized AMSPE depends on the sampled points; we go on to build and implement an algorithm developed to minimize the maximum AMSPE over samples $\{t_1, \ldots, t_n\} \subset [1, N_0]$. Some particular examples are discussed in §3, where they are assessed with respect to a variety of criteria. In §4 we illustrate our methods on a climatological series of 'glacial varve' records, and consider a scenario in which the analyst seeks to predict part of a series from an incomplete sample, using the initial, fully observed portion of the series as 'training data'. Here the analyst assumes that the model fitted to the training series continues to hold, but wishes to build the protection afforded by robustness into his sampling mechanism, should this assumption fail.

2. Minimax Spectra and Designs

2.1. Least favourable spectral density

In this section we exhibit the least favourable spectral density f_* that maximizes the AMSPE for a fixed sampling design over the set $\mathcal{F}_\varepsilon$. With the true autocovariance function given by (1.3), the MSPE of the BLUP $\widehat{X}(t)$ at (1.1), after $X(t)$ has been sampled at times $\{t_1 < \cdots < t_n\}$, is

$$\text{MSPE}(t) = E\left[\left(\widehat{X}(t) - X(t)\right)^2\right] = 1 + \boldsymbol{b}_n'(t)\boldsymbol{R}_n\boldsymbol{b}_n(t) - 2\,\boldsymbol{b}_n'(t)\boldsymbol{r}_n(t).$$

Here we write $\boldsymbol{b}_n(t) = \boldsymbol{A}_n^{-1}\boldsymbol{a}_n(t)$, $\boldsymbol{R}_n = (R(|t_j - t_k|))_{1 \le j,k \le n}$, $\boldsymbol{r}_n(t) = (R(|t - t_j|))_{1 \le j \le n}$. Using $\boldsymbol{M}_n(\omega) = (\cos(\omega|t_j - t_k|))_{1 \le j,k \le n} : n \times n$ and $\boldsymbol{m}_n(\omega, t) = (\cos(\omega|t - t_j|))_{1 \le j \le n} : n \times 1$, we express the autocorrelations in terms of the spectrum, obtaining

$$\boldsymbol{R}_n = \int_{-\pi}^{\pi} \boldsymbol{M}_n(\omega)\, f(\omega) d\omega,$$

$$\boldsymbol{r}_n(t) = \int_{-\pi}^{\pi} \boldsymbol{m}_n(\omega, t)\, f(\omega)\, d\omega.$$

Then the average MSPE is

$$\text{AMSPE}\,(f) = \frac{1}{N}\sum_{t=1}^{N}\text{MSPE}(t) = \int_{-\pi}^{\pi} H_n(\omega)\,f(\omega)\,d\omega,$$

with

$$H_n(\omega) = 1 + \frac{1}{N}\sum_{t=1}^{N}\boldsymbol{b}_n'(t)\boldsymbol{M}_n(\omega)\boldsymbol{b}_n(t) - \frac{2}{N}\sum_{t=1}^{N}\boldsymbol{b}_n'(t)\boldsymbol{m}_n(\omega,t). \qquad (2.1)$$

We assume that $H_n(\omega)$ is non-constant, since otherwise AMSPE(f) does not depend upon f and no spectral density is worse – or better – with respect to AMSPE than is $f_0(\cdot)$. The following theorem is proven in the Appendix.

Theorem 1. *For $0 \leq \varepsilon \leq \|f_0\|_{L_2} \leq \infty$ let*

$$f_*\,(\omega;\varepsilon) = [f_0(\omega) + a\,(H_n(\omega) - b)]^{+}\,,$$

where $f^{+} = \max\,(f,0)$. Then:

(i) *there exist constants $a > 0$ and $b \in \left[\min_{\omega\in[-\pi,\pi]} H_n(\omega), \max_{\omega\in[-\pi,\pi]} H_n(\omega)\right]$ satisfying*

$$\int_{-\pi}^{\pi} f_*\,(\omega;\varepsilon)\,d\omega = 1, \qquad (2.2\text{a})$$

$$\|f_*\,(\cdot;\varepsilon) - f_0\,(\cdot)\|_{L_2} = \varepsilon; \qquad (2.2\text{b})$$

(ii) *with (a,b) determined by (2.2), the function $f_*\,(\cdot;\varepsilon)$ maximizes AMSPE(f) in $\mathcal{F}_\varepsilon$:*

$$\max_{f\in\mathcal{F}_\varepsilon} \text{AMSPE}\,(f) = \text{AMSPE}\,(f_*) = \int_{-\pi}^{\pi} H_n(\omega)\,f_*\,(\omega;\varepsilon)\,d\omega.$$

Our algorithm, described in §2.2, to compute the minimax designs requires a huge number of evaluations of the loss, hence of f_*, and so is very slow if each evaluation requires several separate numerical integrations and optimizations as called for by Theorem 1. Under a condition on the designs – one which is satisfied by all designs we have encountered – a much more explicit solution is possible. For this, a central role is played by the function

$$K_n\,(\omega) = H_n(\omega) - \frac{1}{2\pi}\int_{-\pi}^{\pi} H_n(\omega)d\omega,\ \omega \in [-\pi,\pi]\,, \qquad (2.3)$$

which has $\int_{-\pi}^{\pi} K_n(\omega)d\omega = 0$ and so assumes both positive and negative values. Define subsets of $[-\pi,\pi]$ by

$$I_+ = \{\omega \in [-\pi,\pi]\,|\,K_n\,(\omega) \geq 0\}\,,$$
$$I_- = \{\omega \in [-\pi,\pi]\,|\,K_n\,(\omega) < 0\}\,,$$

and make the restriction to sampling designs for which

$$\inf_{\omega \in I_-} \frac{f_0(\omega)}{|K_n(\omega)|} > \frac{\varepsilon}{\|K_n\|_{L_2}}. \tag{2.4}$$

Then for $\varepsilon \leq \|f_0\|_{L_2}$ the density in Theorem 1 is given by

$$f_{**}(\omega; \varepsilon) = f_0(\omega) + \varepsilon \frac{K_n(\omega)}{\|K_n\|_{L_2}}, \ \omega \in [-\pi, \pi]. \tag{2.5}$$

This function $f_{**}(\omega; \varepsilon)$ is the $f_*(\omega; \varepsilon)$ of Theorem 1 with $a = \varepsilon / \|K_n\|_{L_2}$ and $b = \int_{-\pi}^{\pi} H_n(\omega) d\omega / (2\pi)$, and is nonnegative on $[-\pi, \pi]$ by virtue of (2.4). The AMSPE arising from $f_{**}(\cdot; \varepsilon)$ is

$$\int_{-\pi}^{\pi} H_n(\omega) f_0(\omega) d\omega + \varepsilon \frac{\int_{-\pi}^{\pi} H_n(\omega) K_n d\omega}{\|K_n\|_{L_2}} = \text{AMSPE}(f_0) + \varepsilon \|K_n\|_{L_2}.$$

Theorem 2. *In the class of designs satisfying (2.4), the function at (2.5) maximizes* AMSPE(f) *in* $\mathcal{F}_\varepsilon$. *The maximum* AMSPE *is*

$$\text{AMSPE}(f_{**}(\cdot; \varepsilon)) = \text{AMSPE}(f_0) + \varepsilon \|K_n\|_{L_2}.$$

Our procedure, implemented numerically, is as follows. Each time we evaluate a candidate design, we first compute $f_{**}(\cdot; \varepsilon)$ and check (2.4). If the condition holds, then Theorem 2 applies and we move on. Should the condition fail, we compute instead $f_*(\cdot; \varepsilon)$ of Theorem 1, and the resulting loss. In this manner we find the maximum loss for *any* design, not merely one satisfying (2.4). We have yet to find a design, and $\varepsilon \leq \|f_0\|_{L_2}$, for which (2.4) fails.

Remark. Franke (1985) also notes the difficulties involved in addressing the positivity requirement of spectral densities. He deals with it by allowing the consideration of negative 'densities', and then checks that the maximizer is indeed non-negative.

2.2. Computational issues

The rather lengthy calculations leading to the results of this section are detailed in an online supplement. For the purposes of the computations, an expression more convenient than (2.1) is

$$H_n(\omega) = 1 + \frac{1}{N} tr \mathbf{M}_n(\omega) \mathbf{B}_{n,N} \mathbf{B}'_{n,N} - \frac{2}{N} tr \mathbf{P}_{n,N}(\omega) \mathbf{B}'_{n,N},$$

where $\mathbf{B}_{n,N} = (\mathbf{b}_n(1) \cdots \mathbf{b}_n(N)) : n \times N$ and $\mathbf{P}_{n,N}(\omega) = (\mathbf{m}_n(\omega, 1) \cdots \mathbf{m}_n(\omega, N))$: $n \times N$. In this notation the predictions (1.1) are

$$\left(\hat{X}(1), \ldots, \hat{X}(N)\right) = \mathbf{x}'_n \mathbf{B}_{n,N}. \tag{2.6}$$

We find that

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} H_n\left(\omega\right)d\omega = 1 + \frac{1}{N}\left(\left\|\boldsymbol{B}_{n,N}\right\|_E^2 - 2n\right),$$

where we use the Euclidean matrix norm $\left\|\boldsymbol{P}\right\|_E = \sqrt{tr\boldsymbol{P}'\boldsymbol{P}} = \sqrt{\sum_{j,k} p_{jk}^2}$. Then in implementing Theorem 2, from (2.3) we obtain

$$K_n(\omega) = \frac{1}{N}\left[\left(tr\boldsymbol{M}_n(\omega)\boldsymbol{B}_{n,N}\boldsymbol{B}_{n,N}' - \left\|\boldsymbol{B}_{n,N}\right\|_E^2\right) - 2\left(tr\boldsymbol{P}_{n,N}(\omega)\boldsymbol{B}_{n,N}' - n\right)\right].$$

In terms of the incidence matrices $\boldsymbol{E}_n\left(s\right) = \left(I\left(\left|t_j - t_k\right| = s\right)\right)_{j,k} : n \times n$ and $\boldsymbol{E}_{n,N}\left(s\right) = \left(I\left(\left|k - t_j\right| = s\right)\right)_{j,k} : n \times N$, and using (2.5), the least favourable autocorrelations

$$R_{**}(s) = \int_{-\pi}^{\pi}\cos\left(s\omega\right)f_{**}\left(\omega;\varepsilon\right)d\omega = R_0\left(s\right) + \frac{\varepsilon}{\left\|K_n\right\|_{L_2}}\int_{-\pi}^{\pi}K_n\left(\omega\right)\cos\left(s\omega\right)d\omega$$

are given by

$$R_{**}(s) = R_0\left(s\right) + \frac{\varepsilon\pi I\left(s \neq 0\right)}{N\left\|K_n\right\|_{L_2}}\left[tr\boldsymbol{E}_n\left(s\right)\boldsymbol{B}_{n,N}\boldsymbol{B}_{n,N}' - 2tr\boldsymbol{E}_{n,N}\left(s\right)\boldsymbol{B}_{n,N}'\right].$$

Further calculations yield

$$N\left\|K_n\right\|_{L_2} = \sqrt{\pi\sum_{t=1}^{N-1}\left[tr\boldsymbol{E}_n\left(t\right)\boldsymbol{B}_{n,N}\boldsymbol{B}_{n,N}' - 2\left(\boldsymbol{1}_n'\boldsymbol{E}_{n,N}(t)\boldsymbol{B}_{n,N}'\boldsymbol{1}_n\right)\right]^2}.$$

The next step is to find the sampling design $T = (t_1,\ldots,t_n)$ that minimizes the maximum loss

$$\text{AMSPE}\left(f_{**}\left(\cdot;\varepsilon\right)\right) = \text{AMSPE}\left(f_0\right) + \varepsilon\left\|K_n\right\|_{L_2}$$
$$= 1 - \frac{1}{N}tr\boldsymbol{B}_{n,N}\boldsymbol{B}_{n,N}'\boldsymbol{A}_n + \varepsilon\left\|K_n\right\|_{L_2}.$$

For this we use a genetic algorithm similar to that in Welsh and Wiens (2013), adapted to our case. The optimal design is approached over multiple generations of designs as explained in the outline below. The MATLAB code is available from the authors.

The genetic algorithm proceeds as follows.

Step 1 Generate independently a first generation of N_g designs. Each design can be identified with a vector of n ones and $N_0 - n$ zeros, where $t_1,\ldots,t_n$ are the locations of the ones in this vector.

Step 2 Compute the AMSPE $\mathcal{L}_k$, $k = 1, \ldots, N_g$, for each design of the current generation, and evaluate the corresponding "fitness levels"

$$fitness_k = \frac{1}{\sqrt{rank\,(\mathcal{L}_k)}}, \; k = 1, \ldots, N_g,$$

where $rank\,(\cdot)$ refers to the ranking of the values of $\mathcal{L}$ from 1 to N_g, with the smallest receiving rank 1. Then transform the fitness levels to a probability distribution

$$p_k = \frac{fitness_k}{\sum_{j=1}^{Ng} fitness_j}.$$

Step 3 Replace the current generation by a new generation of N_g designs in the following manner:

(1) Include the fittest $n_{elite} = N_g \times P_{elite}$ members of the current generation. This group survives, intact, through to the next generation. The remaining $N_g - n_{elite}$ designs are formed by crossover and mutation as follows.

(2) Crossover:

(a) Select two members T_{i_1} and T_{i_2} from the current generation to be parents with probability proportional to their fitness levels:

$$i_1 = \min\left\{i : \sum_{j=1}^{i} p_j \geq u_1\right\} \quad \text{and} \quad i_2 = \min\left\{i : \sum_{j=1}^{i} p_j \geq u_2\right\},$$

where u_1, u_2 are independent Uniform$(0, 1)$ random variables. This method, the so-called roulette-wheel selection method, ensures that one is most likely to choose as parents the fittest designs of the current generation.

(b) Add the child of the two members to the new generation. This child is, with probability $1 - P_{crossover}$, identical to the fittest parents. With probability $P_{crossover}$ it is the combined contribution of both parents, formed in the following way. Sum up the two vectors of ones and zeros that identify the design-parents. The resulting vector, with elements in $\{0, 1, 2\}$, is adjusted to have n ones and $N_0 - n$ zeros: First, choose randomly an appropriate number of ones and replace them by zeros. Second, convert the twos to ones. A consequence is that points common to both parents are preserved by the child.
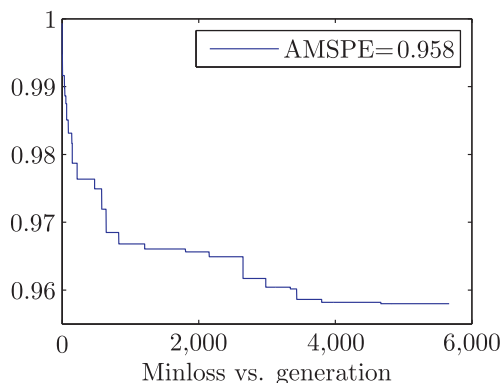
Figure 1. Progress of the genetic algorithm – the minimax design for an approximate AR(1) model, as presented in §3, was found after about 6,000 generations.

(3) Possible mutation is applied to the child in the following manner. Select randomly a one and a zero. With probability $P_{mutation}$ interchange them, and with probability $1 - P_{mutation}$ do nothing. This is repeated n times, for each child.

The procedure is repeated until N_g members of the new generation are formed.

Step 4 Return to Step 2, replace the old generation by the new one and repeat the process. The inclusion of the elite group ensures that the minimum loss per generation is nonincreasing. The algorithm is carried out until the loss of the optimal design has not changed in l_g consecutive generations.

See Figure 1 for an illustration of the progress of the algorithm, as it found the minimax design for the approximate AR(1) model as discussed in the next section. We typically use the tuning parameters $N_g = 40$, $P_{elite} = 0.1$, $P_{crossover} = 0.9$, and decrease the mutation probability linearly, from 0.2 to 0, as the number of generations, without an improved design being found, runs from 0 to l_g. The value of l_g seems to have a much greater influence on the performance of the algorithm than do any of the other parameters; we use $l_g = 1,000$.

3. Examples and Comparisons

In this section we illustrate the methodology outlined in §2, and find robust designs $T^* = (t_1, \ldots, t_n)$ for several base models. We considered four base spectral densities, each normalized by $\int_{-\pi}^{\pi} f_0(\omega)d\omega = 1$; these densities, together with the parameter values used here and the resulting L_2-norms of f_0, are

AR(1): $f_0(\omega) = \dfrac{1 - \phi^2}{2\pi \left[1 + \phi^2 - 2\phi \, \cos(\omega)\right]}$, $\phi = -0.5$, $\|f_0\|_{L_2} = 0.515$;

AR(2): $f_0(\omega) = \dfrac{(1 + \phi_2)\left[(1 - \phi_2)^2 - \phi_1^2\right]}{2\pi(1 - \phi_2)\left[(1 + \phi_1^2 + \phi_2^2) - 2\phi_1(1 - \phi_2)\cos(\omega) - 2\phi_2 \, \cos(2\omega)\right]}$,

$(\phi_1, \phi_2) = (1, -0.5)$, $\|f_0\|_{L_2} = 0.592$;

MA(1): $f_0(\omega) = \dfrac{1 + \theta^2 + 2\theta \cos(\omega)}{2\pi \left(1 + \theta^2\right)}$, $\theta = -0.9$, $\|f_0\|_{L_2} = 0.488$;

MA(2): $f_0(\omega) = \dfrac{1 + \theta_1^2 + \theta_2^2 + 2\theta_1 \left(1 + \theta_2\right) \cos(\omega) + 2\theta_2 \cos(2\omega)}{2\pi \left(1 + \theta_1^2 + \theta_2^2\right)}$,

$(\theta_1, \theta_2) = (-1, 0.6)$, $\|f_0\|_{L_2} = 0.571$.

In practice of course these parameter values would be furnished by some prior knowledge of the process, perhaps garnered from a preliminary study as envisioned in our case study of §4. Similarly, our assumption of a mean of zero would in practice be accommodated by the subtraction of any mean or covariate effects which might be observed or estimated along with $X(t)$.

Figure 2 shows results for an approximate AR(1) model and $\varepsilon = 0.25$. We used $N = 200$ and $n = 30$, on the principle that small sample simulations tend to be more revealing than their large sample counterparts. Figure 1 illustrates the progress of the algorithm. Decreases in the AMSPE are sometimes realized only after extremely long stretches without any improvement, hence our choice of $l_g = 1,000$. As with all such evolutionary algorithms there is still no guarantee that we have found a global minimum, but our experience has been that any further improvements are slight. In this example we used $N_0 = 200$; Figure 3 gives similar output for the MA(1) base model, with the change that there we used $N_0 = 150$.

Figure 4 shows minimax designs for the four base models detailed above, as ε varies and with $N = 100$, $n = 10$. In (a) and (c) we used $N_0 = 75$, whereas (b) and (d) employed $N_0 = 100$.

In addition to those shown in Figures 2 and 3, we have made several other comparisons between our minimax designs and others, with respect to their performance in (neighbourhoods of) a variety of base models. The results are given in Tables 1 and 2, where we consider as well the ARMA(1,1) and independence models.

For each base model detailed in these tables we assess:

(i) our robust design, minimax in a neighbourhood $\mathcal{F}_{0.25}$ of the given base model;

(ii) the 'optimal' design computed under the assumption that the base model is exactly correct, in our notation $\varepsilon = 0$;

Table 1. Comparative losses (AMSPE) for various designs; $n = 10$, $N_0 = 100$, $N = 100$.

Base model	Loss in $\mathcal{F}_\varepsilon$ of robust (for $\mathcal{F}_{0.25}$) design[1]		Loss in $\mathcal{F}_\varepsilon$ of optimal (for $\mathcal{F}_0$) design[2]		Loss in $\mathcal{F}_\varepsilon$ of uniform design[3]	
	$\varepsilon = 0$	$\varepsilon = 0.25$	$\varepsilon = 0$	$\varepsilon = 0.25$	$\varepsilon = 0$	$\varepsilon = 0.25$
AR(1)[4]	0.856	1.084	0.833	1.164	0.833	1.235
MA(1)[4]	0.859	1.044	0.851	1.110	0.851	1.227
ARMA(1,1)[4]	0.799	1.087	0.750	1.225	0.750	1.250
White noise[4]	0.900	1.019	0.900	1.047	0.900	1.199

[1]Designer computed minimax design using $\varepsilon = 0.25$ and the given base model.
[2]Designer assumed $\varepsilon = 0$, the given base model is correct, optimized for it alone.
[3]Equally spaced: $\{5, 15, \ldots, 95\}$.
[4]Parameters of the base models were AR(1): $\phi = 0.5$, MA(1): $\theta = 0.9$, ARMA(1,1): $(\phi, \theta) = (0.5, 0.9)$; for white noise $f_0(\omega) \equiv 1/(2\pi)$.

Table 2. Comparative losses (AMSPE) for various designs; $n = 10$, $N_0 = 75$, $N = 100$.

Base model	Loss in $\mathcal{F}_\varepsilon$ of robust (for $\mathcal{F}_{0.25}$) design[1]		Loss in $\mathcal{F}_\varepsilon$ of optimal (for $\mathcal{F}_0$) design[2]		Loss in $\mathcal{F}_\varepsilon$ of uniform design[3]	
	$\varepsilon = 0$	$\varepsilon = 0.25$	$\varepsilon = 0$	$\varepsilon = 0.25$	$\varepsilon = 0$	$\varepsilon = 0.25$
AR(1)[4]	0.864	1.106	0.833	1.198	0.833	1.230
MA(1)[4]	0.863	1.060	0.851	1.103	0.851	1.210
ARMA(1,1)[4]	0.805	1.110	0.750	1.248	0.751	1.253
White noise[4]	0.900	1.022	0.900	1.049	0.900	1.154

[1,2,4]As in Table 3. Equally spaced: $\{4, 11, 19, 26, 34, 41, 49, 56, 64, 71\}$.

(iii) the uniform (equally-spaced, modulo rounding) design, that requires no model assumptions.

For each we computed the maximum loss AMSPE in a neighbourhood $\mathcal{F}_\varepsilon$ of the given base model, for $\varepsilon = 0$ and for $\varepsilon = 0.25$.

The designs obtained for use in the preparation of Table 1 were:

Base model	Minimax	Optimal
AR(1)	$\{1, 2, 13, 15, 52, 69, 77, 95, 99, 100\}$,	$\{9, 21, 32, 42, 51, 60, 69, 78, 86, 94\}$,
MA(1)	$\{1, 4, 9, 26, 42, 55, 69, 89, 90, 100\}$,	$\{8, 26, 31, 34, 43, 57, 62, 66, 69, 71\}$,
ARMA(1,1)	$\{1, 3, 28, 30, 49, 84, 86, 91, 99, 100\}$,	$\{8, 18, 28, 38, 48, 57, 67, 76, 85, 94\}$,
White noise	$\{1, 13, 34, 50, 64, 72, 91, 95, 98, 100\}$,	$\{21, 52, 65, 66, 72, 73, 83, 84, 87, 100\}$.

The designs used in the preparation of Table 2 were:

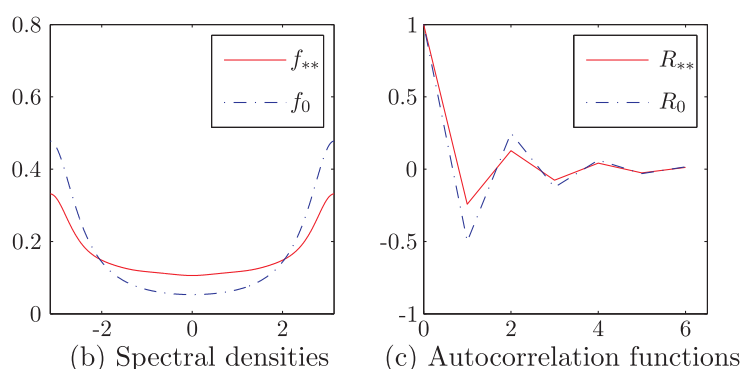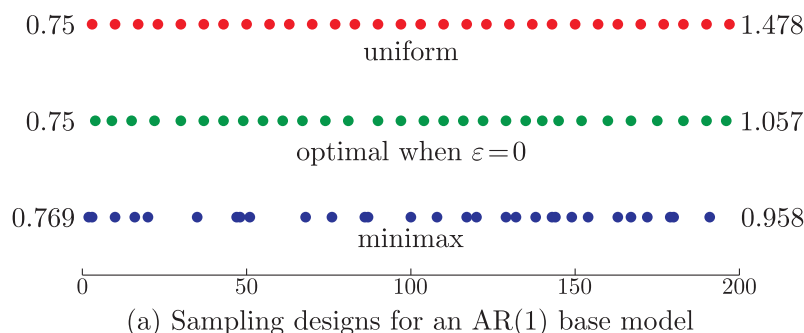| Base model	Minimax	Optimal
AR(1)	$\{1, 2, 5, 15, 16, 47, 53, 54, 74, 7\}$,	$\{5, 12, 19, 26, 34, 43, 51, 59, 67, 75\}$,
MA(1)	$\{1, 7, 8, 20, 21, 37, 47, 70, 71, 74\}$,	$\{5, 8, 11, 15, 18, 21, 33, 50, 63, 75\}$,
ARMA(1,1)	$\{1, 2, 13, 15, 46, 48, 53, 54, 72, 74\}$,	$\{6, 14, 22, 30, 37, 45, 53, 60, 68, 75\}$,
White noise	$\{2, 5, 6, 24, 34, 55, 61, 63, 70, 75\}$,	$\{10, 21, 31, 33, 38, 40, 43, 64, 65, 70\}$.

Figure 2. (a) Sampling designs for an approximate AR(1) model; $\varepsilon = 0.25$, $n = 30$, $N = N_0 = 200$. Values of $\max_{\mathcal{F}_\varepsilon}$ AMSPE are shown for $\varepsilon = 0$ (left) and $\varepsilon = 0.25$ (right). (b) Base spectral density f_0 and least favourable spectrum $f_{**.}$; (c) correponding autocorrelation functions.

From these displays some general features emerge. The robust designs are only slightly less efficient than the 'optimal' designs when the base model holds. However, the maximum loss in $\mathcal{F}_{.25}$ is much smaller. This insurance against model misspecification, in return for a small premium at the base model is exactly what a robust procedure is meant to achieve: when the base model holds, the design optimal for this model, and the uniform design, while not equal, have almost identical losses. The optimal designs are typically more robust than the uniform however, as evidenced by their smaller losses in $\mathcal{F}_{0.25}$.

4. Case Study

Shumway and Verosub (1992) discuss a series of interest in statistical climatology. The data represent the thicknesses of sedimentary deposits ('varves') left behind during spring melts of glaciers, and were reconstructed so as to give an annual record spanning 634 years, beginning about $12,000$ years ago, from a region in what is now Massachusetts. The rather involved reconstruction process was
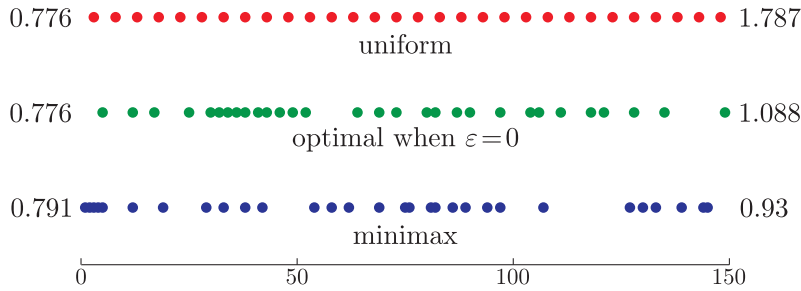
(a) Sampling designs for an MA(1) base model

(b) Spectral densities   (c) Autocorrelation functions
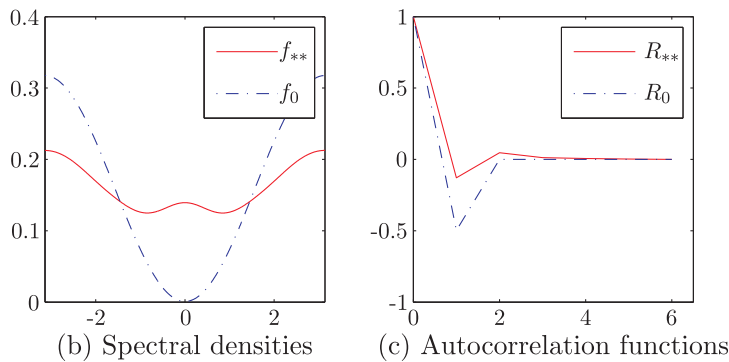
Figure 3. (a) Sampling designs for an approximate MA(1) model; $\varepsilon = 0.25$, $n = 30$, $N = 200$, $N_0 = 150$. Values of $\max_{\mathcal{F}_\varepsilon}$ AMSPE are shown for $\varepsilon = 0$ (left) and $\varepsilon = 0.25$ (right). (b) Base spectral density f_0 and least favourable spectrum f_{**}; (c) correponding autocorrelation functions.

described by De Geer (1912, 1940) and summarized by Shumway and Verosub (1992). These varves may be used as proxies for paleoclimatic parameters such as temperature, since they tend to increase in warm years; hence the prediction of such a series is of current climatological interest.

The transformed series of interest, referred to here as $\{X(t)|t = 1, \ldots, 633\}$, is the first differences of the logarithms of the original data; to first order these represent the relative annual changes in the thicknesses. Here we consider the question: given complete data on the first 500 values of $X(t)$, could one have devised a sampling scheme, so as to observe the remaining 133 points only sporadically, thus saving the expense and effort of reconstructing the entire record? We attempt to do so using our robust method, and compare the results with what could have been attained using other sampling schemes, perhaps tailored to a particular autocorrelation structure determined from the initial 500 observations.

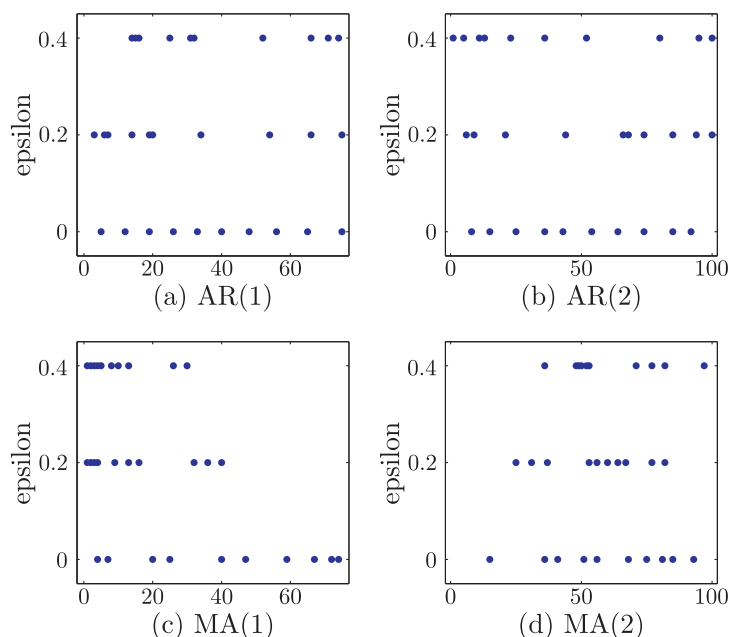Shumway and Verosub (1992) fit an MA(1) model to $\{X(t)\}$, although the

Figure 4. Minimax designs for varying ε and four base models; $n = 10$, $N = 100$. (a), (c) $N_0 = 75$; (b), (d) $N_0 = 100$.

diagnostic plots suggest ARMA(1,1) as a plausible competitor. This uncertainty justifies working in a neighbourhood $\mathcal{F}_\varepsilon$ of the MA(1) spectral density. We took $\varepsilon = .25$ and used the MA parameter $\theta = -.87$, which is the estimate obtained from the fit to the first 500 data points. We then constructed designs of size $n = 25$, used them to sample the $N = 133$ time points set aside, and then computed $\widehat{X}(t)$ as at (2.6), for $t = 501, 502, \ldots, 633$.

The designs studied were the robust design, minimax in $\mathcal{F}_\varepsilon$; the 'optimal' (for $\mathcal{F}_0$) design; the uniform design. The corresponding values of the sample MSE,

$$\widehat{\text{AMSPE}} = \frac{1}{N} \sum_{t=1}^{N} \left(\widehat{X}(500 + t) - X(500 + t) \right)^2,$$

were computed, and are displayed along with the designs in Figure 5(a).

We repeated the above procedure, instead predicting the final 233 values of the series from a sample of $n = 30$ time points, using parameter estimates obtained from the first 400 values; the results are plotted in Figure 5(b). Both procedures were then repeated, with the MA(1) fit replaced by an ARMA(1,1) fit – Figures 5 (c), (d).

Although minimization of the MSE at one particular model is not the goal of a robust procedure, it is nonetheless encouraging that the robust designs always outperformed the optimal designs, and usually the uniform designs as well, with
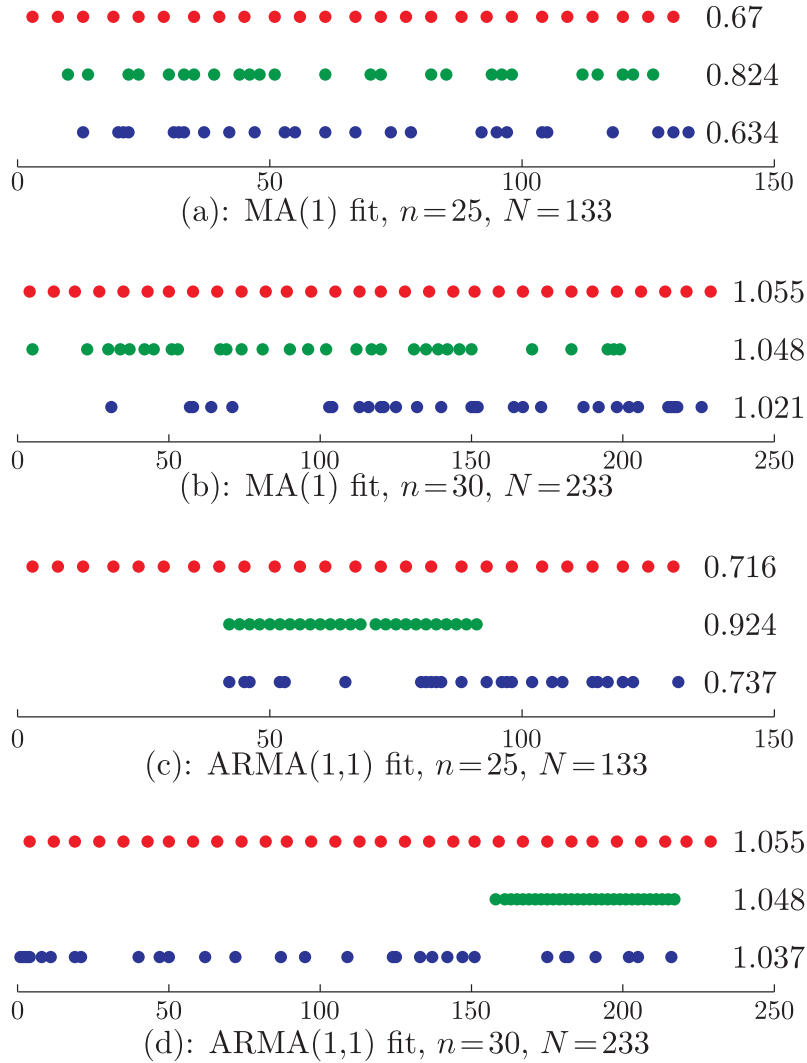
Figure 5. Designs for the glacial varve study – minimax (bottom), optimal for the base model (middle) and uniform (top) – together with the sample AMSPE values.

respect to mean squared error. In light of the comparisons in §3 we do not recommend the routine use of uniform sampling when the underlying model is in doubt.

5. Summary and Conclusions

We have given a method for the construction of sampling designs for a second-order stationary stochastic process that are robust against possible misspecifi-

cations of the underlying mechanism – the spectral density, or autocorrelation function – generating the process. The loss incurred by using a particular design strategy has been measured through the mean squared prediction error, averaged over the range of time points of interest to the investigator. Under a condition on the class of possible designs – a condition which appears to hold universally – we have obtained an explicit expression for the maximum loss which is realized in the class of departures from the fitted model. Even if the condition fails this maximum loss can still be computed, albeit with considerably more difficulty. The maximum loss has then been minimized, to obtain the final designs, via a genetic algorithm. One can see from Figures 2 and 3 – an observation which has been repeated in all other examples at which we have looked – that the least favourable spectrum tends to be more uniform than the base spectrum f_0. As well, the least favourable autocorrelations are smaller in absolute value, sometimes differing in sign from those of the base model. Although the detailed explanation for this lies within Theorem 1, it is a phenomenon which could as well arise through a mixing of f_0 with the uniform spectral density of white noise – in line with the findings of Hosoya (1978) discussed in §1, and those of Welsh and Wiens (2013), who found in a survey sampling context that "the least favourable distribution for modelling dependence is actually the distribution under independence." Further support for this interpretation is that when the base model is one of independence, so that $f_0(\omega) \equiv 1/(2\pi)$, then also $f_{**}(\omega) \equiv 1/(2\pi)$.

Through comparisons such as we have detailed in §3 we see that an investigator who designed with complete faith in his assumed spectrum could be seriously misled, even under small departures from this assumption. Thus methods as outlined here should be routinely used, with at least a small but positive value of ε.

Acknowledgements

Appendix: Proof of Theorem 1

Proof of (i). Let

$$\psi_a(b) = \int_{-\pi}^{\pi} [f_0(\omega) + a(H_n(\omega) - b)]^+ \, d\omega,$$

and let b_L and b_U be the minimum and maximum values of the (continuous) function $H_n(\omega)$, $\omega \in [-\pi, \pi]$. Then for $a > 0$ and $b \in [b_L, b_U]$,

$$f_0(\omega) + a(H_n(\omega) - b_U) \leq [f_0(\omega) + a(H_n(\omega) - b)]^+ \leq f_0(\omega) + a(H_n(\omega) - b_L),$$

so that $\psi_a(b)$ ranges monotonically over $[\psi_a(b_U), \psi_a(b_L)]$ as b ranges over $[b_L, b_U]$. Now an integration of the terms in the inequalities

$$f_0(\omega) + a(H_n(\omega) - b_U) \le f_0(\omega) \le f_0(\omega) + a(H_n(\omega) - b_L)$$

shows that $1 \in [\psi_a(b_U), \psi_a(b_L)]$. Thus for each $a > 0$ there is $b = b(a) \in [b_L, b_U]$ for which $\psi_a(b) = 1$; this is (2.2a).

Take $I_a(\omega)$ to be the indicator of the event that $f_0(\omega) + a(H_n(\omega) - b(a)) \ge 0$. In this notation (2.2a) is

$$\int_{-\pi}^{\pi} (1 - I_a(\omega)) f_0(\omega)\, d\omega = \int_{-\pi}^{\pi} I_a(\omega) \{a(H_n(\omega) - b(a))\}\, d\omega. \qquad (A.1)$$

Equation (2.2b) is $\phi(a) = \varepsilon^2$, where

$$\phi(a) = \int_{-\pi}^{\pi} (1 - I_a(\omega)) f_0^2(\omega) + I_a(\omega) \{a(H_n(\omega) - b(a))\}^2\, d\omega.$$

We show that any $\varepsilon^2 \le \|f_0\|_{L_2}^2$ is in the range of the continuous function $\phi(a)$. First observe that as $a \downarrow 0$, both sides of (A.1) must tend to zero; this can only be the case if $(1 - I_a(\omega)) \to 0$ for a.e. ω at which $f_0(\omega) > 0$. Thus also $\phi(a) \to 0$ as $a \downarrow 0$.

Now note that

$$\phi(a) \ge \int_{-\pi}^{\pi} \min\left[f_0^2(\omega), \{a(H_n(\omega) - b(a))\}^2 \right] d\omega.$$

For almost every ω, since $b(a)$ is bounded,

$$a\,|H_n(\omega) - b(a)| \ge a\,||H_n(\omega)| - |b(a)|| \to \infty,$$

as $a \to \infty$. We say 'almost every' to exclude those ω for which $|H_n(\omega)| = \lim_{a\to\infty} |b(a)|$; since $H_n(\omega)$ is a non-constant linear combination of cosines at integer multiples of ω such points have measure zero.

If $\sqrt{\int_{-\pi}^{\pi} f_0^2(\omega)\, d\omega} = \|f_0\|_{L_2} < \infty$ then by the Dominated Convergence Theorem,

$$\begin{aligned}
\lim_{a\to\infty} \phi(a) &\ge \lim_{a\to\infty} \int_{-\pi}^{\pi} \min\left[f_0^2(\omega), \{a(H_n(\omega) - b(a))\}^2 \right] d\omega \\
&= \int_{-\pi}^{\pi} \lim_{a\to\infty} \min\left[f_0^2(\omega), \{a(H_n(\omega) - b(a))\}^2 \right] d\omega \\
&= \int_{-\pi}^{\pi} f_0^2(\omega)\, d\omega.
\end{aligned}$$

If $\|f_0\|_{L_2} = \infty$ then Fatou's Lemma leads to the same conclusion – the existence of a pair $(a, b(a))$ satisfying (2.2), as long as $0 \le \varepsilon \le \|f_0\|_{L_2}$.

Proof of (ii): Given the constants in (i), $f_* (\cdot; \varepsilon) \in \mathcal{F}_\varepsilon$. Let f be any other member of $\mathcal{F}_\varepsilon$ and define, for $s \in [0, 1]$,

$$f_{(s)}(\omega) = (1 - s) f_* (\omega; \varepsilon) + s f (\omega) \in \mathcal{F}_\varepsilon.$$

The function

$$\zeta(s) = \int_{-\pi}^{\pi} \left\{ (H_n(\omega) - b) f_{(s)}(\omega) - \frac{1}{2a} \left(f_{(s)}(\omega) - f_0(\omega) \right)^2 \right\} d\omega$$

is concave on $[0, 1]$ – it is quadratic in s with negative leading coefficient – and has derivative

$$\zeta'(0) = \int_{-\pi}^{\pi} (f(\omega) - f_* (\omega; \varepsilon)) \left\{ (H_n(\omega) - b) - \frac{1}{a} (f_* (\omega, \varepsilon) - f_0(\omega)) \right\} d\omega.$$

On the set $\mathcal{S}$ where $f_* (\omega; \varepsilon) > 0$ we have

$$(H_n(\omega) - b) - \frac{1}{a} (f_* (\omega, \varepsilon) - f_0(\omega)) = 0;$$

on the complement $\mathcal{S}^c$ of this set we have $f_0(\omega) + a (H_n(\omega) - b) \leq 0$. Thus

$$\zeta'(0) = \frac{1}{a} \int_{\mathcal{S}^c} f(\omega) \left\{ a (H_n(\omega) - b) + f_0(\omega) \right\} d\omega \leq 0.$$

It follows that $\zeta(0) \geq \zeta(1)$ for any $f \in \mathcal{F}_\varepsilon$, so

$$\text{AMSPE}(f_*) - \text{AMSPE}(f) = \int_{-\pi}^{\pi} H_n(\omega) f_* (\omega, \varepsilon) d\omega - \int_{-\pi}^{\pi} H_n(\omega) f(\omega) d\omega$$

$$\geq \frac{1}{2a} \left\{ \|f_* (\cdot; \varepsilon) - f_0(\cdot)\|_{L_2}^2 - \|f(\cdot) - f_0(\cdot)\|_{L_2}^2 \right\} \geq 0.$$

References

Anscombe, F. J. (1960). Rejection of outliers. *Technometrics* **2**, 123-147.

Christakos, G. (1992). *Random Fields Models in Earth Sciences*. Academic Press.

De Geer, G. (1912). A geochronology of the last 12,000 years. *Report of the* 11*th Geological Congress* **1**, 241-253.

De Geer, G. (1940). Geochronologia suecica principles. *Svenska Betenskapsakad. Handl.*, 3:18, 367p.

Diaconis, P. (1988). Bayesian numerical analysis. *Statistical Decision Theory and Related Topics IV* **1** (Edited by S. S. Gupta and J. O. Berger), 163-175. Springer-Verlag.

Franke, J. (1985). Minimax robust prediction of discrete time series. *Z. Wahrsch. Verw. Gebiete* **68**, 337-364.

Hosoya, Y. (1978). Robust linear extrapolations of second-order stationary processes. *Ann. Probab.* **6**, 575-584.

Journel, A. and Huijbregts, C. (1978). *Mining Geostatistics.* Academic Press.

Matheron, G. (1963). Principles of Geostatistics. *Economic Geology* **58**, 1246-1266.

Mukherjee, B. (2003). Exactly optimal sampling designs for processes with a product covariance structure. *Canad. J. Statist.* **31**, 69-87.

Samarov, A. M. (1987). Robust spectral regression. *Ann. Statist.* **15**, 99-111.

Shumway, R. H. and Verosub, K. L. (1992). State space modeling of paleoclimatic time series. *Proceedings of the* 5*th International Meeting on Statistical Climatology*, 22-26.

Wein, A. and Srinivasan, L. (2013). InterML: A fast, robust algorithm for estimating signals with finite rate of innovation. *IEEE Trans. Signal Process.* **61**, 5324-5336.

Welsh, A. H. and Wiens, D. P. (2013). Robust model-based sampling designs. *Statist. Comput.* **23**, 689-701.

Department of Mathematical and Statistical Sciences; University of Alberta, Edmonton, Alberta; Canada T6G 2G1.

E-mail: rabhi@ualberta.ca

Department of Mathematical and Statistical Sciences; University of Alberta, Edmonton, Alberta; Canada T6G 2G1.

E-mail: doug.wiens@ualberta.ca