

Web-based supporting materials for “Generalized Pareto processes for simulating space-time extreme events” by F. Palacios-Rodríguez, G. Toulemonde, J. Carreau and T. Opitz

1. Example: Pareto processes with log-Gaussian profile process

If Gaussian process models are not well adapted to modeling extremes, they can nevertheless be used to construct flexible spatial or spatio-temporal limit models (Kabluchko et al. (2009); Engelke et al. (2015)). For instance, De Fondeville and Davison (2018) analyse the extreme rainfall in the east of Florida by fitting a spatial generalized Pareto process based on log-Gaussian processes. Sample-continuous max-stable process $\{Z(s, t)\}_{s \in \mathcal{S}, t \in \mathcal{T}}$ with unit Fréchet margins can be characterized constructively as (de Haan (1984); Schlather (2002))

$$Z(s, t) = \max_{i \geq 1} \xi_i \psi_i(s, t), \quad s \in \mathcal{S}, t \in \mathcal{T}, \quad (1)$$

where $\{\xi_i, i = 1, 2, \dots\}$ is a point process on $[0, \infty)$ with intensity $\xi^{-2} d\xi$, and $\psi_i(s, t)$ are independent copies of a nonnegative random function with $\mathbb{E}\psi_i(s, t) = 1$, and independent of $\{\xi_i\}$. Specifically, one may choose $\psi_i(s, t) = \exp\{X(s, t) - \sigma^2(s, t)/2\}$ with a centered Gaussian process $\{X(s, t)\}$ possessing variance function $\sigma^2(s, t)$. Regarding the ℓ -Pareto processes equivalent to such max-stable processes, the choice of $\ell(x) = x(s_0, t_0)$ for a fixed space-time point (s_0, t_0) is particularly interesting. In this case, the profile process $Y(s, t)$ in the generalized Pareto process is a log-Gaussian process given by $Y(s, t) \stackrel{d}{=} \exp\{X(s, t) - X(s_0, t_0) - \frac{1}{2} \text{var}(X(s, t) - X(s_0, t_0))\}$ where var denotes the variance. The idea of conditioning on a fixed component of a process is more widely known as the conditional extremes approach (Heffernan and Tawn (2004); Wadsworth and Tawn (2018)), and it arises as a special case of the cost functional ℓ .

2. Risk measures

Risk is a complex notion and can take on a variety of forms with diverse applications. Risk could be defined as the effect that lack of certainty produces on objectives (ISO (2009)). The conventional risk measure in hydrology is that of the univariate return level at level $q \in [0, 1]$ denoted as Q_q . A *return level* is a quantile, defined as the magnitude of the event that is exceeded with a probability $1 - q$; then, $1/(1 - q)$ is the associated *return period*. However,

the return level fails to give any information about the thickness of the tail of the distribution function. That is, a regulator can know only the frequency of default but not the severity of default (Denuit et al., 2005). In order to prevent the above shortcoming, another interesting risk measure was proposed in actuarial sciences, the so-called Conditional Tail Expectation (CTE) (Denuit et al., 2005). Information about the thickness of the tail of the distribution is included in the CTE defined for a given level $q \in [0, 1]$ and for an univariate random variable X by $CTE_q(X) = E(X|X > Q_q(X))$. In contrast to the return level, the CTE measure verifies the subadditivity property for continuous risks.

3. Estimator of extremal coefficient

Let $X^{(1)}, \dots, X^{(M)}$ be identically distributed random variables with unit Fréchet distribution, that is, $\mathbb{P}(X^{(k)} \leq x) = e^{-1/x}$, $x > 0$, $k = 1, \dots, M$. When the joint distribution of the random vector $(X^{(1)}, \dots, X^{(M)})^T$ follows a multivariate extreme value distribution, then the distribution function of $\max_{k=1}^M X^{(k)}$ is $e^{-\theta/x}$, $x > 0$, where $\theta = \theta(X^{(1)}, \dots, X^{(M)})$, $1 \leq \theta \leq M$, is called the extremal coefficient (Smith, 1990; Schlather and de Tawn, 2003). In practice, the coefficient θ can be interpreted as the equivalent number of asymptotically independent random variables (i.e., the effective sample size of extremes) in a random vector $(X^{(1)}, \dots, X^{(M)})$; it quantifies the dependence for extreme values. The case $\theta = 1$ represents full dependence, whereas $\theta = M$ represents full independence.

When considering threshold exceedances, extreme realizations are those that exceed a high threshold. Suppose that $(X_i^{(1)}, \dots, X_i^{(M)})^T$, $i = 1, \dots, n$ are independent and identically distributed (iid) copies of the random vector $(X^{(1)}, \dots, X^{(M)})^T$, where a threshold exceedance is observed for $X_i^{(k)}$, $1 \leq k \leq M$ if $X_i^{(k)} > u_i^{(k)}$ for some fixed threshold $u_i^{(k)}$; otherwise, the observation $X_i^{(k)}$ is considered as being left-censored at $u_i^{(k)}$. Caires et al. (2011) propose an estimator of the extremal coefficient constructed as

$$\hat{\theta} = m \left/ \sum_{i=1}^n \frac{1}{\max(X_i, u_i)} \right. \quad (2)$$

where $X_i = \max(X_i^{(1)}, \dots, X_i^{(M)})$, $u_i = \max(u_i^{(1)}, \dots, u_i^{(M)})$, and m is the number of excesses $X_i > u_i$.

For a summary of extremal dependence with respect to distance in space and time, we follow common practice and focus on pairwise extremal coefficients calculated from bivariate data $X_i = \max(X_i^{(1)}, X_i^{(2)})$ with $M = 2$ in (2). We work with two extremal coefficient functions. The spatial extremal coefficient $\theta^{spa}(h)$ measures the extremal dependence between pairs of sites

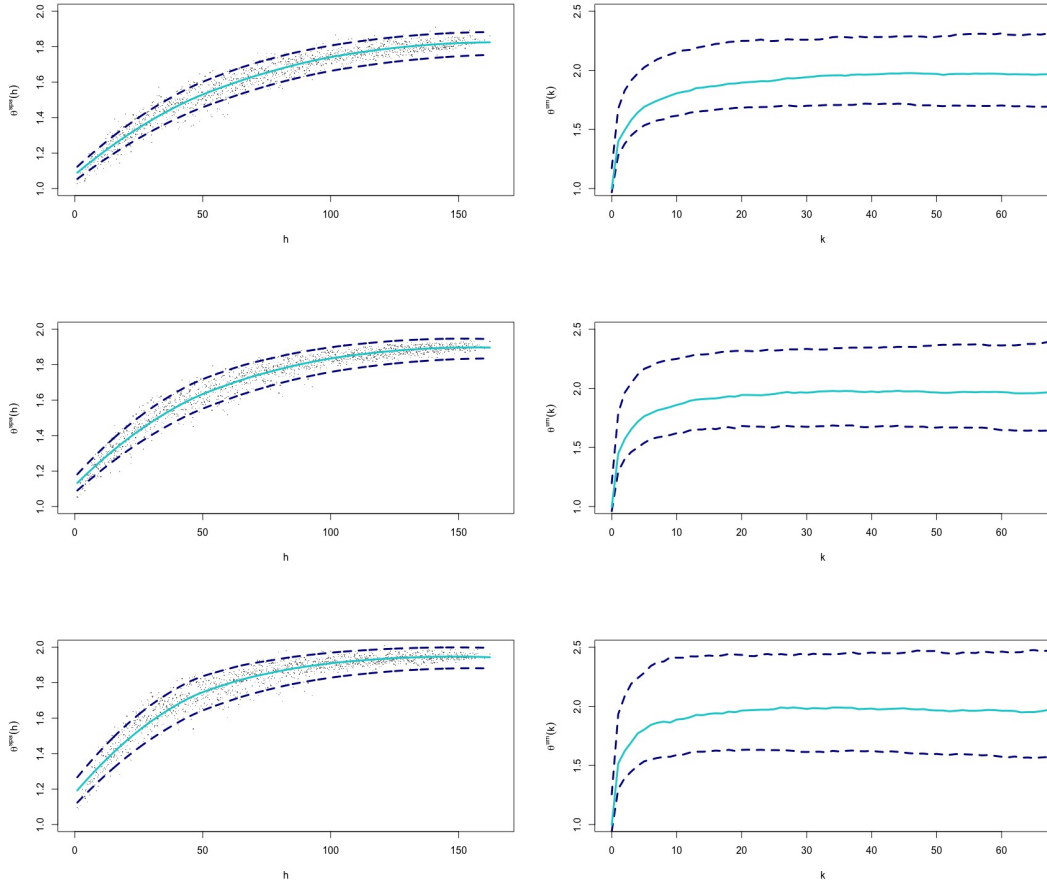


Fig. 1. Empirical extremal coefficient functions. Left: $\hat{\theta}^{spa}(h)$, based on a subsample of 1500 pairs of grid cells, with a local polynomial regression (turquoise line). Right: $\hat{\theta}^{tim}(k)$, based on pairs of spatial maxima separated by a time lag k (turquoise line). Bootstrap confidence intervals at 95% (dashed lines). Threshold values u_i in (2) are defined as the empirical q -quantile with $q = 0.98, 0.99, 0.995$ (from top to bottom).

separated by a spatial distance h at a given time. The time extremal coefficient $\theta^{tim}(k)$ measures the dependence between pairs of observations separated by a time lag k at a given site (see Section 2.2 in Chailan et al. (2017) for more details). We estimate empirical spatial extremal coefficient functions from data by considering pairs with structure $(X(s, t_i), X(s + \Delta s, t_i))$ where $\Delta s = h$, while we use $(\max_{s \in \mathcal{S}} X(s, t_i), \max_{s \in \mathcal{S}} X(s, t_i + k))$ for empirical temporal extremal coefficient functions where \mathcal{S} is the study area.

Figure 1 (left panel) presents the spatial extremal coefficient estimates. We set u_i in (2) as the maximum of the empirical q -quantiles of $X(s, t_i)$ and $X(s + h, t_i)$, where the latter two variables represent a pair of sites separated by a given spatial distance h at a given hour t_i .

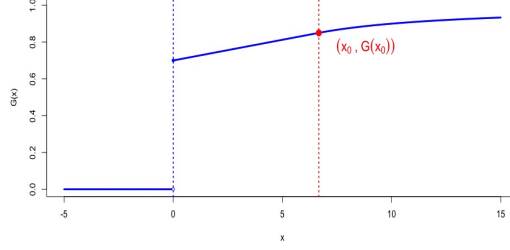


Fig. 2. Distribution function G for $p_0 = 0.7$.

The temporal extremal coefficient estimates are plotted in Figure 1 (right panel). In this case, the threshold values u_i are chosen as an empirical q -quantiles of the spatial maximum. The following values for q are used : 0.98, 0.99 and 0.995 (rows from top to bottom in Figure 1, respectively). Block bootstrap confidence intervals at 95% for both extreme coefficients are constructed by resampling blocks of hours with variable size following a geometric distribution with a mean of 300 hours (i.e. approximately 12 days) (Politis and Romano, 1994; Davis et al., 2011).

4. Application to precipitation in Southern France

4.1. Example of the normalized marginal distribution G

We choose a distribution G whose survival function \bar{G} verifies: $x\bar{G}(x) \rightarrow 1$, $x \rightarrow \infty$, and $\bar{G}(0) = 1$; we write G^{\leftarrow} for the (generalized) inverse function of G . We then define the transformation $T = T_{s,t}$ towards the standardized process X^* as follows:

$$X^*(s, t) = T(X(s, t)) = G^{\leftarrow}(F_{(s,t)}(X(s, t)))$$

where $F_{(s,t)}$ denotes the distribution function of $X(s, t)$. Following ideas in Opitz (2016), we construct G to have mass $p_0 \geq 0$ at 0, a uniform density on $(0, x_0)$, and a standard Pareto distribution for $x > x_0$ where $x_0 > 1$. The junction point x_0 is chosen to ensure continuity of the density of G for $x > 0$. Figure 2 provides an illustration of G function when $p_0 = 0.7$.

$$G(x) = \begin{cases} 0, & x < 0, \\ p_0, & x = 0, \\ p_0 + \frac{(1-p_0)^2}{4}x, & 0 < x \leq 2/(1-p_0), \\ 1 - 1/x, & x > 2/(1-p_0). \end{cases} \quad (3)$$

4.2. Spatio-temporal extremes episodes for spatial maximum

By considering the spatial maximum cost functional, Figure 3 shows the original rainfall data process $X(s, t)$ and the final uplifted process $W(s, t)$ for a subset of the selected extreme episodes from Table 1 in main document.

4.3. Risk analysis

Figure 4 presents the calculated CTE at the two levels 0.98 and 0.99 for the selected extreme episodes with two choices of ℓ cost functionals (spatio-temporal mean, and spatial maximum) and with the four scale parameter values for the Pareto random variable. Here, we uplift the extreme episodes by defining the value of the scale variable R as the 0.25-, 0.5- and 0.75-quantile of the Pareto distribution with shape 1 and scale α_i , $i = 1, \dots, 4$ with α_1 corresponding to the value of the cost functional for the most extreme episode associated to t_1 , and $\alpha_2 = 2\alpha_1$, $\alpha_3 = 3\alpha_1$ and $\alpha_4 = 4\alpha_1$. Along the y-axis, the values of the CTE of each of the three original episodes are indicated.

References

- Caires, S., de Haan, L. and Smith, R. L. (2011) On the determination of the temporal and spatial evolution of extreme events. *Deltares report 1202120-001-HYE-004 (for Rijkswaterstaat, Centre for Water Management)*.
- Chailan, R., Toulemonde, G. and Bacro, J. N. (2017) A semiparametric method to simulate bivariate space-time extremes. *Ann. Appl. Statist.*, **11(3)**, 1403–1428.
- Davis, R. A., Mikosch, T. and Cribben, I. (2011) Estimating Extremal Dependence in Univariate and Multivariate Time Series via the Extremogram. *arxiv:1107.5592v1*.
- De Fondeville, R. and Davison, A. C. (2018) High-dimensional peaks-over-threshold inference. *Biometrika*, **105(3)**, 575–592.
- Denuit, M., Dhaene, J., Goovaerts, M. and Kaas, R. (2005) *Actuarial Theory for Dependence Risks: Measures, Orders and Models*. Wiley.
- Engelke, S., Malinowski, A., Kabluchko, Z. and Schlather, M. (2015) Estimation of Hüsler-Reiss distributions and Brown-Resnick processes. *J. R. Statist. Soc. B*, **77**, 239–265.
- de Haan, L. (1984) A spectral representation for max-stable processes. *Ann. Probab.*, **12(4)**, 1194–1204.

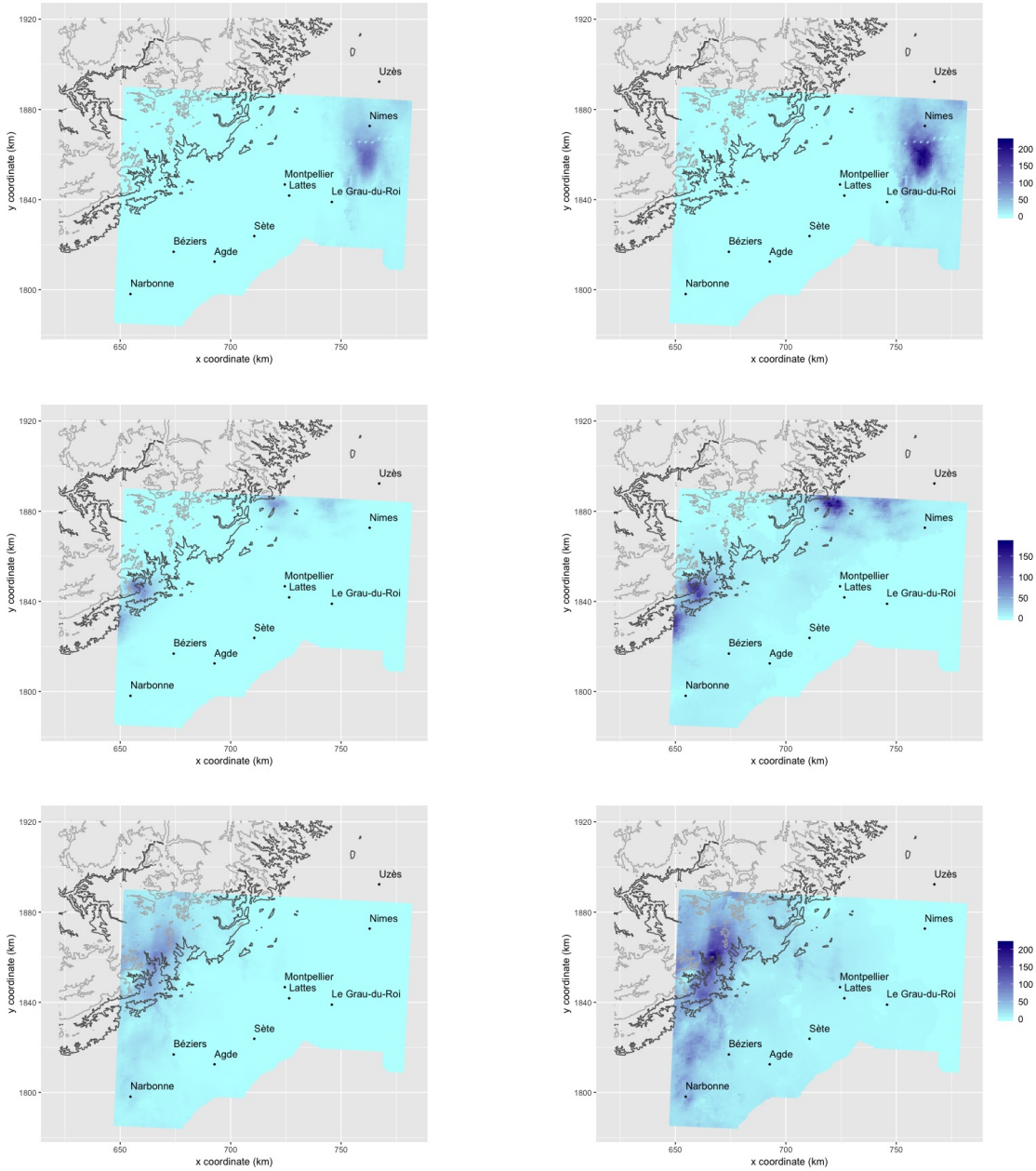


Fig. 3. Original precipitation data $X(s, t)$ (left column) and uplifted episodes $W(s, t)$ (right column) based on the spatial maximum $\ell_t^{(2)}$. First row: extreme episode associated to t_1 , here shown for $t = 2005-09-06$, 15:00:00; second row: same for t_3 and $t = 1999-08-28$, 16:00:00; third row: same for t_6 and $t = 2006-10-11$, 20:00:00.

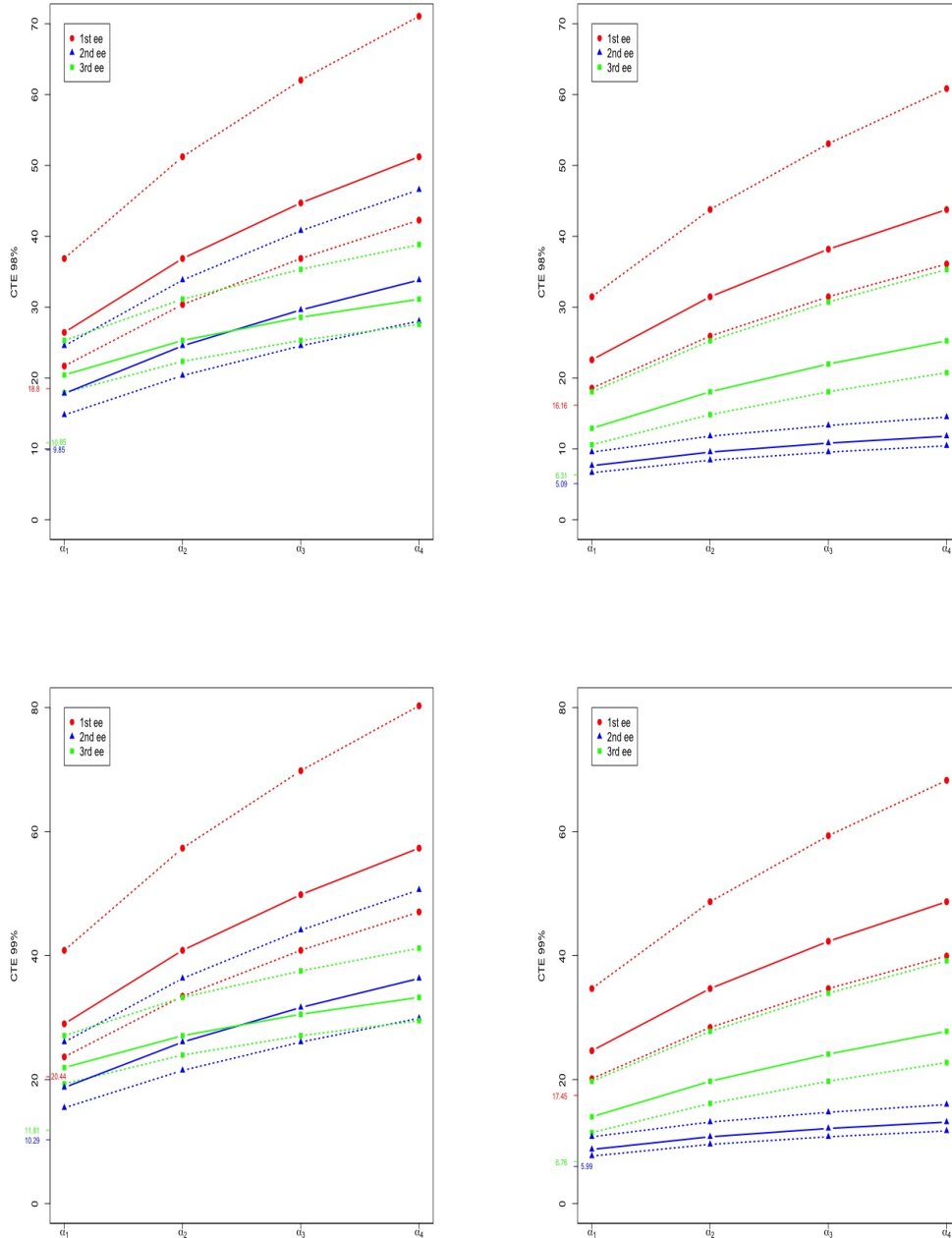


Fig. 4. Conditional-Tail-Expectation at 98% (first row) and at 99% (second row). Left: spatio-temporal average cost functional. Right: spatial maximum cost functional. The legend indicates the extreme episode (ee). For each episode, the lines correspond to different uplifting levels using the 0.25-, 0.5- and 0.75-quantile (from bottom to top) of the Pareto distribution of the scaling variable with shape 1 and scale $\alpha_i, i = 1, \dots, 4$

- Heffernan, J. E. and Tawn, J. A. (2004) A conditional approach for multivariate extreme values. *J. R. Statist. Soc. B*, **66(3)**, 497–546.
- ISO (2009) *ISO 31000- Risk Management: Principles and Guidelines*. Int. Organ. for Stand.
- Kabluchko, Z., Schlather, M. and de Haan, L. (2009) Stationary max-stable fields associated to negative definite functions. *Ann. Probab.*, **37(5)**, 2042–2065.
- Opitz, T. (2016) Modeling asymptotically independent spatial extremes based on laplace random fields. *Spatial Statistics*, **16**, 1–18.
- Politis, D. N. and Romano, J. P. (1994) The stationary bootstrap. *J. Am. Statist. Ass.*, **89**, 1303–1313.
- Schlather, M. (2002) Models for Stationary Max-Stable Random Fields. *Extremes*, **5(1)**, 33–44.
- Schlather, M. and de Tawn, J. A. (2003) A Dependence Measure for Multivariate and Spatial Extreme Values: Properties and Inference. *Biometrika*, **90(1)**, 139–156.
- Smith, R. L. (1990) Max-stable processes and spatial extremes. *Preprint. University of Surrey*.
- Wadsworth, J. L. and Tawn, J. A. (2018) Spatial conditional extremes. <https://www.lancaster.ac.uk/wadswojl/CSE-paper.pdf>.