Composite likelihood strategies for the Bayesian analysis of large hierarchical space-time models

[PRELIMINARY RESULTS – WORK IN PROGRESS]

Denis Allard, with Lionel Benoit and G. Mariethoz (U. Lausanne)

Biostatistics and Spatial Processes (BioSP), MIA division, INRA
Avignon, France

RESSTE Workshop on "Hierarchical Bayesian Models for spatio – temporal data",
Paris, 16 May, 2016

Motivation

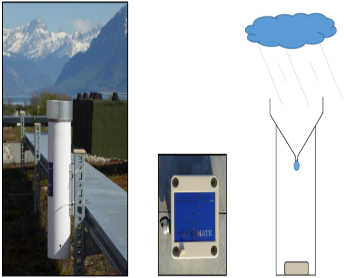
Main Objective

Analyzing and modeling precipitation **within** a radar pixel i.e. at **very fine scale**

- Fine scale modeling of precipitations: $\sim$ min; $\sim 10 - 1000$ m
- Accounting for large amount of 0 values
- Space-time dependence structure, including transport (advection)
- Large data-set
- Bayesian analysis **and** (conditional) simulation of rain events

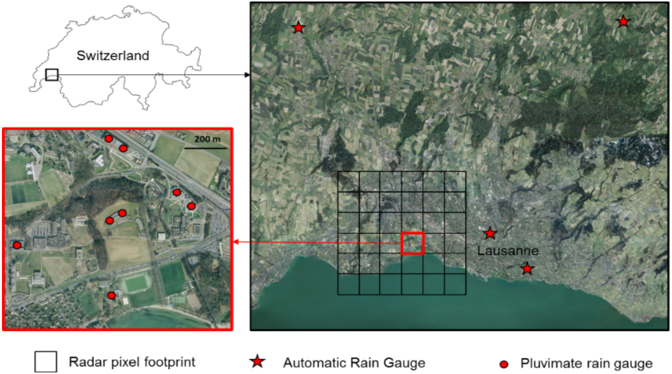
Data

- New sorts of rain gauges, called pluvimates
- Drop counting rain gauges ~ 0.01 mm resolution
- Have been calibrated with lab experiments
- Integration time is 30s

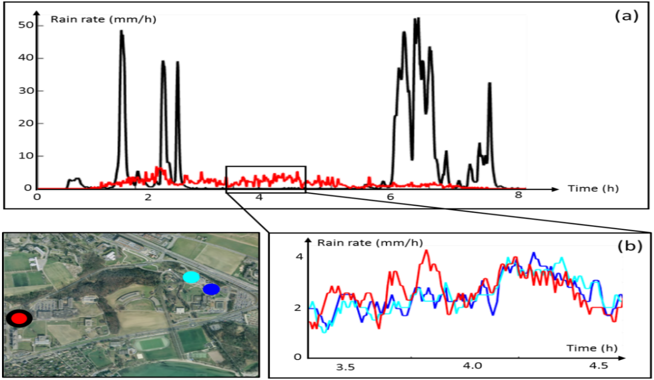


(a) (b) (c)

Data


Switzerland
200 m
Lausanne
Radar pixel footprint
Automatic Rain Gauge
Pluvimate rain gauge

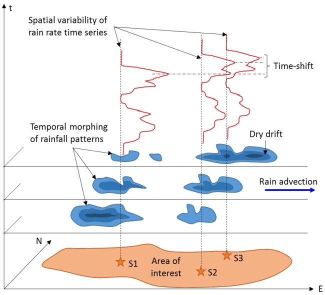
Data



Space-time rain rate fluctuations as observed by a dense network of Pluvimates. (a) Global view of a stratiform (red, 3-4 January 2016) and a convective (black, 24-25 October 2016) rain events. (b) Zoom on a two hours period for the convective event, and comparison between co-located (blue curves) and distant (red vs blue curves) measurement locations.

Conceptual model

- Fine space-time scales
  $\Longrightarrow$ stationarity
- Evolving shapes along time
  $\Longrightarrow$ space-time correlation
- Dry-drift
  $\Longrightarrow$ correlation between intensity and presence of rainfall
- Advection of clouds
  $\Longrightarrow$ transport term



Hierarchical model

Latent Gaussian field

Allcroft and Glasbey (2003), Allard and Bourotte (2014), Baxevani and Lennatsson (2015), etc...

- Spatio-temporal coordinates $(\mathbf{s}, t) \in (D \times T)$
- Precipitations $R_m(\mathbf{s}, t)$ arise from a latent, standardized, stationary Gaussian random field $Y(\mathbf{s}, t)$ with

$$Cov(Y(\mathbf{s}, t), Y(\mathbf{s}', t')) = \rho(\mathbf{s}' - \mathbf{s}, t' - t; \boldsymbol{\eta}),$$

  where $\rho(\cdot, \cdot, \boldsymbol{\eta})$ is a spatio-temporal covariance function, with parameters $\boldsymbol{\eta}$
- Marginal transformation

$$R_m(\mathbf{s}_i, t_i) = \psi(Y(\mathbf{s}_i, t_i) + \epsilon_i) = \left(\frac{Y(\mathbf{s}_i, t_i) + \epsilon_i - a_0}{a_1}\right)^{1/a_2}, \quad i = 1, \dots, n,$$

  with $\epsilon_i \sim \mathcal{N}(0, \sigma_\epsilon^2)$ i.i.d and $\boldsymbol{\alpha} = (a_0, a_1, a_2)^t$.

$Y(\mathbf{s}, t)$ is hidden, since not observed when $R_m(\mathbf{s}, t) = 0$.

Spatio-temporal covariance model

Non separable spatio-temporal covariance function

Gneiting (2002), Gneiting et al. (2007), etc.
Bourotte et al. (2016) in the multivariate case.

- Advection is modeled by a single vector $\mathbf{V}$ with

$$\rho(||\mathbf{s}' - \mathbf{s}|| - \mathbf{V}(t' - t), |t' - t|) = \rho_L(||\mathbf{s}' - \mathbf{s}||, |t' - t|). \qquad (1)$$

- Non separable model for ρ_L, with

$$\rho_L(\mathbf{h}, u) = \frac{1}{(u/d)^{2\delta} + 1} \exp\left\{ \frac{-(|||\mathbf{h}||/c)^{2\gamma}}{\{(u/d)^{2\delta} + 1\}^{\beta\gamma}} \right\} \quad (\mathbf{h}, u) \in \mathbb{R}^2 \times \mathbb{R}, \qquad (2)$$

with $\boldsymbol{\eta} = (c, d, \beta, \gamma, \delta, S_V, \theta_V, \sigma_\epsilon^2)$.

Likelihood

Let $R_m(\mathbf{s}_1, t_1), \ldots, R_m(\mathbf{s}_n, t_n)$ be the rainfall measurements.

- For $i = 1, \ldots, n$ we write

$$Z(\mathbf{s}_i, t_i) = \begin{cases} \psi^{-1}(R_m(\mathbf{s}_i, t_i)) & = & a_0 + a_1 R_m(\mathbf{s}_i, t_i)^{a_2} & \text{when } R_m(\mathbf{s}_i, t_i) > 0 \\ Z(\mathbf{s}_i, t_i) & \leq & a_0 & \text{otherwise} \end{cases}$$

- Let $I = \{i : R_m(\mathbf{s}_i, t_i) > 0\}$ and $I^0 = \{i : R_m(\mathbf{s}_i, t_i) = 0\}$ and let $\mathbf{R} = (\mathbf{R}_{I^0}, \mathbf{R}_I)$.

Likelihood

$$\ell(\mathbf{R}; \boldsymbol{\theta}) = \ell(\mathbf{R}_I; \boldsymbol{\theta}).\mathbb{P}(\mathbf{R}_{I^0} \leq \mathbf{0} \mid \mathbf{R}_I; \boldsymbol{\theta}),$$

where $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \boldsymbol{\eta})$.

Likelihood

Log-likelihood

$$
\begin{aligned}
L(\mathbf{R};\boldsymbol{\theta}) &= L(\mathbf{R}_I;\boldsymbol{\theta}) + \log \mathbb{P}(\mathbf{R}_{I^0} \leq \mathbf{0} \mid \mathbf{R}_I;\boldsymbol{\theta}). \\
&= -0.5 \log|\boldsymbol{\Sigma}_{I,I}| - 0.5\mathbf{Z}_I^t\boldsymbol{\Sigma}_{I,I}^{-1}\mathbf{Z}_I - N_I \log(2\pi) \\
&\quad + \log \Phi_{N_{I^0}}(\mathbf{a}_0; \boldsymbol{\Sigma}_{I^0,I}\boldsymbol{\Sigma}_{I,I}^{-1}, \boldsymbol{\Sigma}_{I^0,I^0} + \boldsymbol{\Sigma}_{I_0,I}\boldsymbol{\Sigma}_{I,I}^{-1}\boldsymbol{\Sigma}_{I,I^0})
\end{aligned}
$$

with

$$\boldsymbol{\Sigma}_{I,I}[i,j] = \rho_L(\mathbf{s}_j - \mathbf{s}_i - \mathbf{V}(t_j - t_i), t_j - t_i) \quad \text{for } i \neq j \in I, \quad \text{and } \boldsymbol{\Sigma}_{I,I}[i,i] = 1 + \sigma_\epsilon^2$$

- Computing $\boldsymbol{\Sigma}_{I,I}^{-1}$ and $|\boldsymbol{\Sigma}_{I,I}|$: necessitate $\mathcal{O}(N_I^3)$ operations
- Computing $\Phi_{N_{I^0}}$, which is a N_{I^0} dimensional normal probabilities, see e.g. Genz (1992, 2004, 2009, etc.) and R package `mvtnorm`.
- Limited to moderate size datasets, ~ 1000 i.e. $\simeq$ 2h with 8 pluvimates

If larger datasets: block likelihood (BL) or alternative stratgeies

Estimation methods

Block Likelihood

- The original dataset is divided into a series of N_τ blocks, B_p of τ consecutive measurements at all sites
- For $p = 1, \ldots, N_\tau$, let

$$
\begin{aligned}
I_p &= \{i \in B_p : R_m(\mathbf{s}_i, t_i) > 0\} \\
I_p^0 &= \{i \in B_p : R_m(\mathbf{s}_i, t_i) = 0\}
\end{aligned}
$$

- The Blockwise log-likelihood, $L_{BL}(\mathbf{R};\boldsymbol{\theta})$ is

$$
\begin{aligned}
L_{BL}(\mathbf{R};\boldsymbol{\theta}) = \sum_{p=1}^{N_\tau} & \Big[-0.5\left(|\boldsymbol{\Sigma}_{I_p,I_p}| + \mathbf{Z}_{I_p}^t \boldsymbol{\Sigma}_{I_p,I_p}^{-1} \mathbf{Z}_{I_p} + N_{I_p}\log(2\pi)\right) \\
& + \log\Phi(a_0; \boldsymbol{\Sigma}_{I_p^0,I_p}\boldsymbol{\Sigma}_{I_p,I_p}^{-1}, \boldsymbol{\Sigma}_{I_p^0,I_p^0} - \boldsymbol{\Sigma}_{I_p^0,I_p}\boldsymbol{\Sigma}_{I_p,I_p}^{-1}\boldsymbol{\Sigma}_{I_p,I_p^0})\Big].
\end{aligned}
$$

Which size for the blocks ?

Bayesian setting

The full model is

$$
\begin{aligned}
R_m(\mathbf{s}_i, t_i) &= \psi(Y(\mathbf{s}_i, t_i)) \\
Y(\cdot,\cdot) &\sim \mathcal{G}(\rho) + \epsilon; \qquad \rho: \ \mathbf{V} \text{ and } \rho_L(\mathbf{h}, u) \sim \text{Gneiting class} \\
\epsilon &\sim \text{i.i.d. } \mathcal{N}(0, \sigma_\epsilon^2) \\
\sigma_\epsilon^2 &\sim \pi_\epsilon \\
\boldsymbol{\eta} &\sim \pi_{\boldsymbol{\eta}} \\
\boldsymbol{\alpha} &\sim \pi_{\boldsymbol{\alpha}} \\
\mathbf{V} &\sim \pi_{\mathbf{V}}
\end{aligned}
$$

- All parameters are independent
- Gibbs for censored $Y(\cdot,\cdot)$ within Metropolis Hastings for $\boldsymbol{\theta}$
- Uniform vague priors
- Random Walk proposals
- 5000 iterations for burn-in; 10000 iterations; 100 samples

Data

Synthetic data

Situation mimicking real data

- 400 60s-measurements. $\Rightarrow$ 6h40m
- 9 pluvimates on 1000m $\times$ 1000m regular grid
- ρ_L: $c = 10000m,\ \gamma = 0.3,\ d = 1000s,\ \delta = 0.5,\ \beta = 0.5,\ \sigma_\epsilon = 0.2$
- **V**: $Sv = 5m/s,\ \theta_v = 0$
- Transform: $a_0 = -0.1,\ a_1 = 0.4,\ a_2 = 0.8$
- 80 blocks of size 5 (BL5); 4 blocks of size 100 (BL100); 1 block of size 400 (BL400)

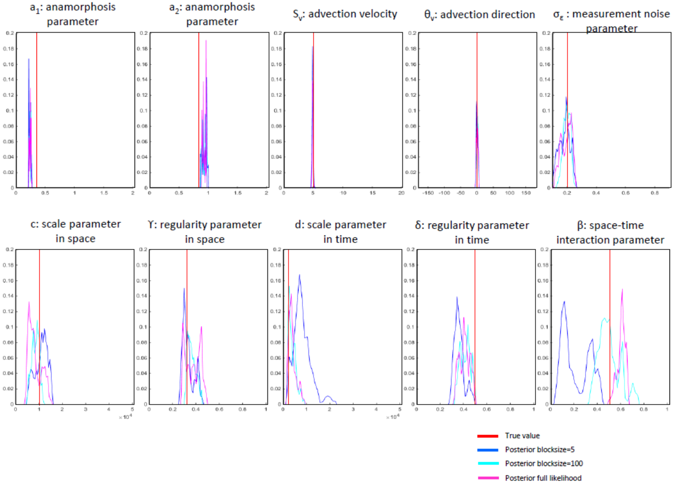
Real data

- 1000 30s-measurements. $\Rightarrow$ 8h.
- 8 pluvimates on 1000m $\times$ 1000m
- 50 blocks of size 20

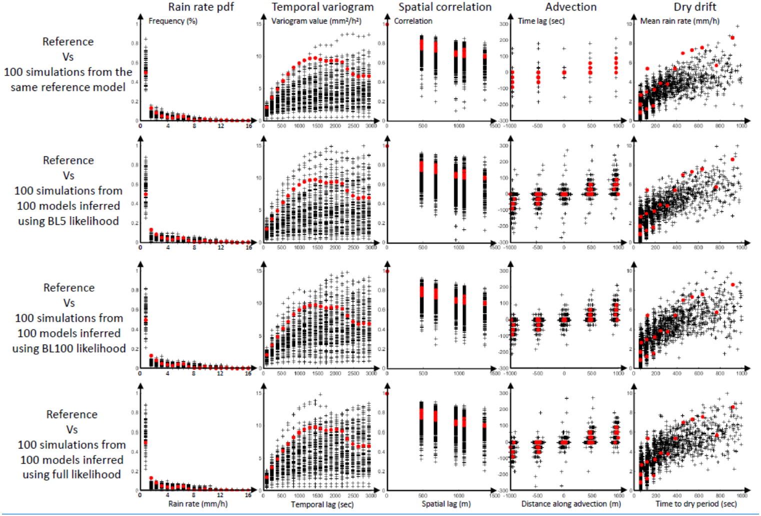
CPU

- Matlab; 20 cores machine for matrix inversion (Intel Xéon CPU E5-2699 v4, 2.2GHz)
- From 45m (BL5) to 550m (BL100) and 1300m (BL400)

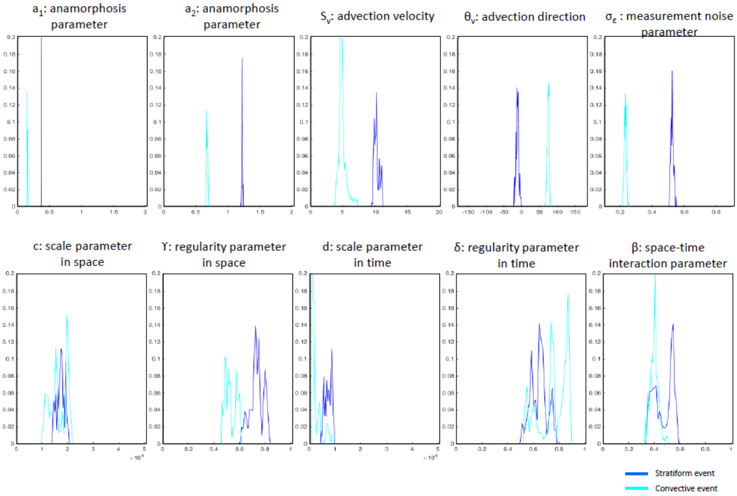
Synthetic data – Posteriors



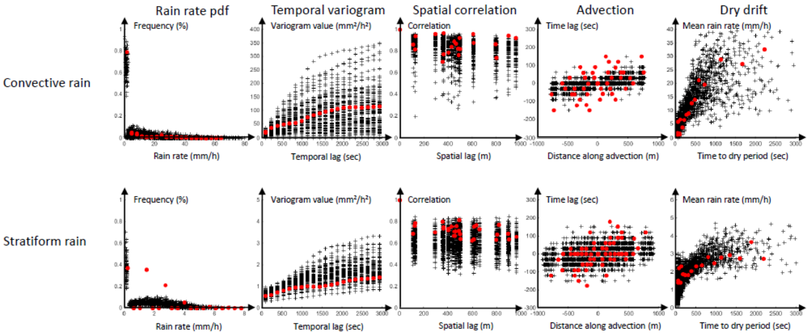
Synthetic data – Simulations



Real data – Posteriors



Real data – Simulations



Alternative methods/approaches

Sparse precision matrices

- Some Matérn spatial covariance matrices induce sparse precision matrices Lindgren et al. (2011)
- Efficient parallel implementation possible ⇒ larger datasets

Composite likelihood

- products of smaller likelihoods Varin et al. (2011) ⇒ easy to compute, e.g. pairs or smaller blocks
- Pairwise Likelihood (PL) is efficient and unbiased, see e.g. Bevilacqua and Gaetan (2015), Bourotte et al. (2016)
- But PL is under dispersed ⇒ cannot be used directly in a Bayesian setting
- Adjustments and re-calibration proposed in Ribatet et al. (2012) and Stoehr and Friel (2015).

Pairwise Composite Likelihood

- Pairwise marginal Gaussian log-likelihoods computed on pairs of data (Bevilacqua and Gaetan, 2015)
$$L_{PL}(\mathbf{s}_i, \mathbf{s}_j, t_i, t_j; \theta).$$
- The Weighted Pairwise log-Likelihood (WPL) is thus (Bevilacqua and Gaetan, 2015)
$$\text{wpl}(\theta) = \sum_{(i,j,\alpha,\beta)} L_{PL}(\mathbf{s}_i, \mathbf{s}_j, t_i, t_j; \theta) w_{ij}, \quad \theta \in \Theta,$$
- The computational cost is $\mathcal{O}\left(N^2\right)$.
- Can be further reduced to $\mathcal{O}(KN)$ by only considering pairs such that
$$(\|\mathbf{h}\|, u) \leq (d_S, d_T).$$
- Three different expressions for $L_{PL}(\mathbf{s}_i, \mathbf{s}_j, t_i, t_j; \theta)$ depending on dry or wet conditions for i and j.

Curvature adjustment

In the context of spatial extremes, where full likelihood is not accessible, Ribatet et al. (2012) proposed to adjust L_{PL}:

- Let us define

$$H(\boldsymbol{\theta}_0) = -\mathbb{E}[\nabla^2 L_{PL}(\boldsymbol{\theta}_0; \mathbf{R})]$$

  and

$$J(\boldsymbol{\theta}_0) = -\mathbb{V}[\nabla L_{PL}(\boldsymbol{\theta}_0; \mathbf{R})]$$

- Under usual regularity conditions

$$\sqrt{n}\{H(\boldsymbol{\theta}_0)J^{-1}(\boldsymbol{\theta}_0)H(\boldsymbol{\theta}_0)\}^{1/2}(\boldsymbol{\theta}_0 - \hat{\boldsymbol{\theta}_{PL}}) \rightarrow N(0, \mathbf{I})$$

- Curvature adjustment

$$L_{curv}(\boldsymbol{\theta}) = L_{PL}(\hat{\boldsymbol{\theta}}_{PL} + \mathbf{C}(\boldsymbol{\theta}_0 - \hat{\boldsymbol{\theta}}_{PL})),$$

  with

$$\mathbf{C}^t H(\boldsymbol{\theta}_0)\mathbf{C} = H(\boldsymbol{\theta}_0)J^{-1}(\boldsymbol{\theta}_0)H(\boldsymbol{\theta}_0)$$

- Gibbs and Metropolis-Hastings algorithms are then adapted using L_{curv} instead of L_{PL}.

Curvature adjustment

Gaussian process with exponential covariance: 3 parameters

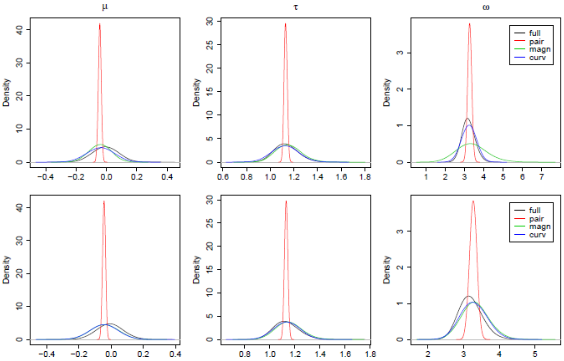


Figure 3: *Comparison between the marginal full posterior (black), the marginal pairwise posterior (red) and the marginal adjusted pairwise posterior densities based on the magnitude (green) and curvature (blue) adjustments. The posterior distributions are derived from $n = 50$ realisations of a Gaussian process having an exponential covariance function with $\mu = 0$, $\tau = 1$ and $\omega = 3$ and observed at $K = 20$ locations. Top row: Metropolis–Hastings algorithm. Bottom row: Adaptive adjusted Gibbs sampler.*

Future work

- Extend / adapt this work to our dataset
- Many more parameters: $3 \to 12$
- Some regularity conditions might be lost
- Requires estimates $\hat{\boldsymbol{\theta}}_{PL}$ and estimates for $H(\boldsymbol{\theta}_0)$ and $J(\boldsymbol{\theta}_0)$

Looks promising but challenging

Bibliography

Ailliot, P., Allard, D., Monbet, V., and Naveau, P. (2015). Stochastic weather generators: an overview of weather type models. *Journal de la Société Fran caise de Statistique*, 156(1).

Allard, D. and Bourotte, M. (2014). Disaggregating daily precipitations into hourly values with a transformed censored latent Gaussian process. *preprint*.

Baxevani, A. and Lennatsson, J. (2015). A spatiotemporal precipitation generator based on a censored latent gaussian field. *Water Resources Research*.

Bevilacqua, M. and Gaetan, C. (2015). Comparing composite likelihood methods based on pairs for spatial Gaussian random fields. *Statistics and Computing*, 25(5):877–892.

Bourotte, M., Allard, D., and Porcu, E. (2016). A flexible class of non-separable cross-covariance functions for multivariate space-time data. *Spatial Statistics*, 18, 125–146.

Gneiting, T. (2002). Nonseparable, stationary covariance functions for space-time data. *Journal of the American Statistical Association*, 97(458):590–600.

Gneiting, T., Genton, M., and Guttorp, P. (2007). Geostatistical space-time models, stationarity, separability and full symmetry. In Finkenstädt, B., Held, L., and Isham, V., editors, *Statistical Methods for Spatio-Temporal Systems*, pages 151–175. Chapman & Hall/CRC.

Gneiting, T., Kleiber, W., and Schlather, M. (2010). Matérn cross-covariance functions for multivariate random fields. *Journal of the American Statistical Association*, 105(491):1167–1177.

Lindgren, F., Rue, H., Lindström, J. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(4), 423-498.

Ribatet, M., Cooley, D., Davison, A. C. (2012). Bayesian inference from composite likelihoods, with an application to spatial extremes. *Statistica Sinica*, 813-845.

Stoehr, J., Friel, N. (2015). Calibration of conditional composite likelihood for Bayesian inference on Gibbs random fields. (AISTATS).

Varin, C., Reid, N., and Firth, D. (2011). An overview of composite likelihood methods. *Statistica Sinica*, 21(1): 5–42.