Approximate Bayesian Computation with Domain Expert in the Loop

Ayush Bharti[1], Louis Filstroff[1], Samuel Kaski[1,2]

Department of Computer Science, Aalto University, Finland[1]
Department of Computer Science,University of Manchester, UK[2]

October 28, 2022

Aalto University
School of Science

MANCHESTER 1824
The University of Manchester

Outline

Inference from data

Setting:

- Let data $\mathbf{y}_{\mathrm{obs}} = \{y_{\mathrm{obs},i}\}_{i=1}^{n}$ be denoted by empirical distribution $\mathbb{Q}^n$.
- Model $\mathcal{M}_\Theta = \{\mathbb{P}_\theta : \theta \in \Theta \subset \mathbb{R}^q\}$ is a parametric family of distributions.

Estimation problem: Given data $\mathbf{y}_{\mathrm{obs}}$, estimate θ s.t. $\mathbb{Q}^n$ is “closest" to $\mathbb{P}_\theta$.

Inference from data

Setting:

- Let data $\mathbf{y}_{\text{obs}} = \{y_{\text{obs},i}\}_{i=1}^{n}$ be denoted by empirical distribution $\mathbb{Q}^n$.
- Model $\mathcal{M}_\Theta = \{\mathbb{P}_\theta : \theta \in \Theta \subset \mathbb{R}^q\}$ is a parametric family of distributions.

Estimation problem: Given data $\mathbf{y}_{\text{obs}}$, estimate θ s.t. $\mathbb{Q}^n$ is “closest" to $\mathbb{P}_\theta$.

Classical estimation techniques such as

$$\text{Bayesian inference:} \quad p(\theta|\mathbf{y}_{\text{obs}}) \propto p(\mathbf{y}_{\text{obs}}|\theta)p(\theta)$$

$$\text{Maximum Likelihood (ML):} \quad \hat{\theta}_{\text{ML}} = \underset{\theta}{\operatorname{argmax}}\, p(\mathbf{y}_{\text{obs}}|\theta)$$

require access to the likelihood function.

Problem: Many models have intractable likelihoods

The likelihood function cannot be evaluated numerically, or approximated in reasonable computation time.

Therefore, standard estimation techniques are unrealizable.

Causes of intractable likelihood:

- The model is simply too complex.
- Variables that are important for model description are unobserved.
- The likelihood function has not been derived yet for a newly constructed model.

Such models are called:

- Simulators
- Implicit models
- Generative models

Simulators in the Sciences

Physical sciences and engineering:

- Population genetics [Pritchard et al., 1999]
- Ecology and evolution [Beaumont, 2010]
- Astrophysics [Akeret et al., 2015]
- Epidemiology [Kypraios et al., 2017]
- Radio communications [Bharti et al., 2021]
- Atmospheric science [Kopka et al., 2016]
- Economics [Dyer et al., 2022]

Solution: use likelihood-free inference methods based on simulating from the model

Outline

Approximate Bayesian Computation (ABC)

ABC is a likelihood-free inference method that permits sampling from the approximate posterior of a model, given that it is easy to simulate from.

Rejection ABC algorithm

- Sample $\theta^* \sim p(\theta)$
- Simulate data from model, $\mathbf{y}^* \sim \mathbb{P}_{\theta^*}$
- If $\rho(S(\mathbf{y}_{\text{obs}}), S(\mathbf{y}^*)) < \epsilon$, accept θ^*

Approximate Bayesian Computation (ABC)

ABC is a likelihood-free inference method that permits sampling from the approximate posterior of a model, given that it is easy to simulate from.

Rejection ABC algorithm

- Sample $\theta^* \sim p(\theta)$
- Simulate data from model, $\mathbf{y}^* \sim \mathbb{P}_{\theta^*}$
- If $\rho(S(\mathbf{y}_{\mathrm{obs}}), S(\mathbf{y}^*)) < \epsilon$, accept θ^*

Here

- $\rho(\cdot\,,\,\cdot)$ is a distance metric (typical choice is Euclidean distance)
- $S(\cdot)$ is the summarizing function
- ϵ is a tolerance threshold
- Accepted samples $\theta_1, \ldots, \theta_N$ are iid from the approximate posterior:

$$p(\theta|\rho(S(\mathbf{y}_{\mathrm{obs}}), S(\mathbf{y}) < \epsilon) \approx p(\theta|\mathbf{y}_{\mathrm{obs}})$$

Approximate Bayesian Computation (ABC) — contd.

The “approximation" in Bayesian inference arises due to

- use of tolerance threshold in accepting parameter samples
- summarizing the data into a few statistics. If $S(\cdot)$ is a *sufficient statistic* of $\mathbf{y}$, then

$$p(\theta|\rho(S(\mathbf{y}^*), S(\mathbf{y}_{\text{obs}})) < \epsilon) = p(\theta|\rho(\mathbf{y}^*, \mathbf{y}_{\text{obs}}) < \epsilon)$$

Approximate Bayesian Computation (ABC) — contd.

The “approximation" in Bayesian inference arises due to

- use of tolerance threshold in accepting parameter samples
- summarizing the data into a few statistics. If $S(\cdot)$ is a *sufficient statistic* of $\mathbf{y}$, then

$$p(\theta|\rho(S(\mathbf{y}^*), S(\mathbf{y}_{\text{obs}})) < \epsilon) = p(\theta|\rho(\mathbf{y}^*, \mathbf{y}_{\text{obs}}) < \epsilon)$$

Ingredients required for implementing an ABC algorithm:

- distance metric $\rho(\cdot, \cdot)$
- summary statistics $S(\cdot)$
- tolerance threshold ϵ

$$\rho(S(\mathbf{y}_{\text{obs}}), S(\mathbf{y}^*)) < \epsilon$$

Approximate Bayesian Computation (ABC) — contd.

Choosing ϵ

- In practice, select ϵ as a small percentile of the simulated distances, i.e. given $\{(\boldsymbol{\theta}_i^*, S(\mathbf{y}_i^*))\}_{i=1}^M$, accept the ϵM samples of $\boldsymbol{\theta}_i^*$ with the least $\rho(S(\mathbf{y}), S(\mathbf{y}_i^*))$

Choose $\rho(\cdot, \cdot)$

- Euclidean distance for summary-based ABC
- Integral probability metrics such as maximum mean discrepancy, Wasserstein distance

Approximate Bayesian Computation (ABC) — contd.

Choosing ϵ

- In practice, select ϵ as a small percentile of the simulated distances, i.e. given $\{(\boldsymbol{\theta}_i^*, S(\mathbf{y}_i^*))\}_{i=1}^M$, accept the ϵM samples of $\boldsymbol{\theta}_i^*$ with the least $\rho(S(\mathbf{y}), S(\mathbf{y}_i^*))$

Choose $\rho(\cdot, \cdot)$

- Euclidean distance for summary-based ABC
- Integral probability metrics such as maximum mean discrepancy, Wasserstein distance

Choosing $S(\cdot)$

- The choice of statistics is non-trivial as it involves trade-off between
    1. information loss due to summarization
    2. curse of dimensionality
- Fundamental unsolved problem in ABC

Choosing Summary Statistics

In practice, domain experts manually handcraft and select statistics:

- laborious and time-consuming
- involves multiple trial-and-error steps
- takes up majority of the time of likelihood-free inference projects

Hence, domain knowledge is vital for constructing summaries.

Choosing Summary Statistics

In practice, domain experts manually handcraft and select statistics:

- laborious and time-consuming
- involves multiple trial-and-error steps
- takes up majority of the time of likelihood-free inference projects

Hence, domain knowledge is vital for constructing summaries.

Existing methods given a pool of candidate statistics:

- Subset selection (Joyce & Marjoram, 2008)
- Projection techniques (Fearnhead & Prangle, 2012)
- Regression adjustment (Beaumont et al., 2010)

However, performance of these methods degrade when

- number of available model simulations is limited (low-simulation regime)
- model is misspecified, i.e., $\mathbb{Q}^n \notin \mathcal{M}_\Theta$

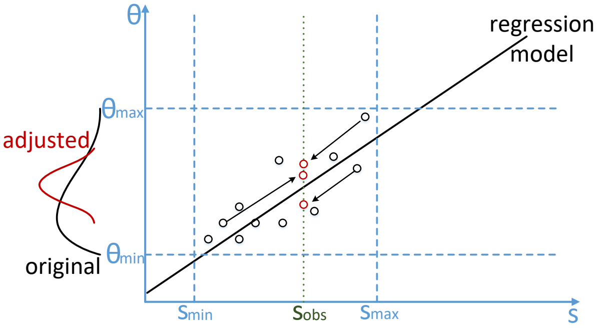
Outline

Case study: Regression-ABC methods

Given accepted samples $(\theta_i, \mathbf{s}_i)_{i=1}^{n_\epsilon}$, regression-ABC methods account for the difference between the simulated and observed statistic by adjusting the parameter values using model

$$\theta_i = \varphi(\mathbf{s}_i) + \varepsilon_i, \quad i = 1, \ldots, n_\epsilon.$$

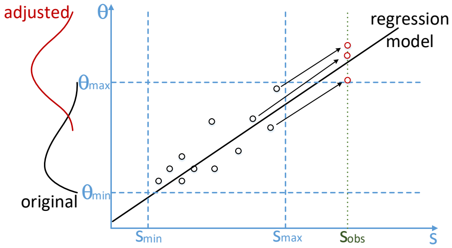
- $\mathbf{s}$: statistics vector
- $\varphi(\cdot)$: conditional expectation $\mathbb{E}[\theta|\mathbf{s}]$
- ε_i: the residuals
- Adjusted samples: $\tilde{\theta}_i = \hat{\varphi}(\mathbf{s}_{\text{obs}}) + \hat{\varepsilon}_i$



Failure of regression-ABC methods

Model misspecification

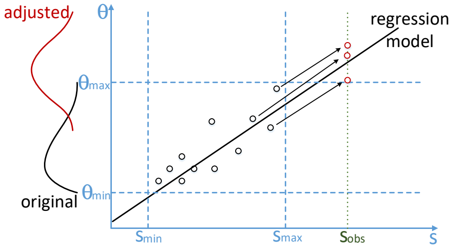
- Regression-ABC can yield erroneous results under misspecification.
- Approx. posterior can lie beyond prior range.

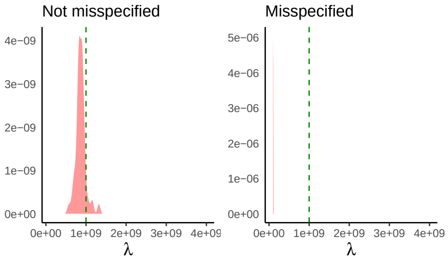


Failure of regression-ABC methods

Model misspecification

- Regression-ABC can yield erroneous results under misspecification.
- Approx. posterior can lie beyond prior range.

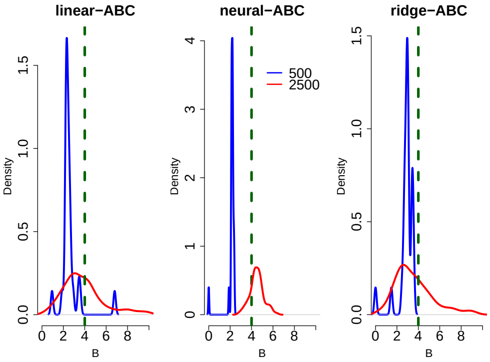




Failure of regression-ABC methods

Low-simulation regime

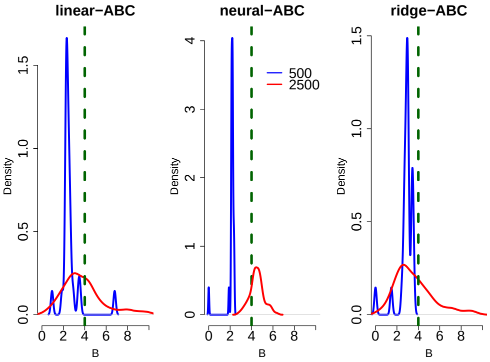
- When model is expensive, number of simulations are limited.
- Regression is susceptible to overfitting.
- ABC posteriors can be concentrated far from the true value.



Failure of regression-ABC methods

Low-simulation regime

- When model is expensive, number of simulations are limited.
- Regression is susceptible to overfitting.
- ABC posteriors can be concentrated far from the true value.

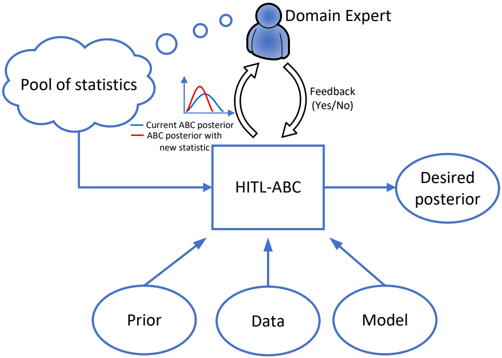


Can we mitigate these issues by leveraging domain expertise?

Outline

Human-in-the-loop (HITL) ABC

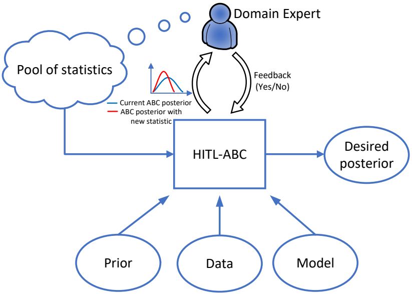
- Involve the expert in the inference procedure
- Elicit domain knowledge about statistics
- Assumptions:
  - Expert knowledge is tacit
  - Querying the expert is costly



Problem setting

Human-in-the-loop (HITL) ABC

- Pool of statistics: $\mathcal{S} = \{s_1, s_2, \ldots, s_w\}$
- Indicator of inclusion or exclusion of s_j: $\gamma_j \in \{0, 1\}$
- With each vector **s**, we have associated $\gamma = [\gamma_1, \ldots, \gamma_w]^\top$
- ABC posterior: $p^{\epsilon}_{\mathrm{ABC}}(\theta|\mathbf{y}_{\mathrm{obs}}, \gamma)$
- Let γ^* represent the “desired" statistics vector
- **Goal:** converge towards γ^* by querying expert about $\mathcal{S}$



Expert feedback model

Human-in-the-loop (HITL) ABC

Formulate expert feedback as a probabilistic model

- Expert provides binary feedback $f_j \in \{0, 1\}$ about s_j
- Model f_j as a noisy version of γ_j:

$$\gamma_j \sim \text{Bernoulli}(\rho_j),$$
$$f_j|\gamma_j \sim \gamma_j \text{Bernoulli}(\pi) + (1 - \gamma_j)\text{Bernoulli}(1 - \pi).$$

- $\pi \in [0, 1]$: the level of noise
- ρ_j: prior probability of selecting s_j
- Let $\mathcal{F}$ be the set of feedback obtained
- $p(\boldsymbol{\gamma}|\mathcal{F})$ available in closed-form

Expert feedback model

Human-in-the-loop (HITL) ABC

Formulate expert feedback as a probabilistic model

- Expert provides binary feedback $f_j \in \{0, 1\}$ about s_j
- Model f_j as a noisy version of γ_j:

$$\begin{aligned} \gamma_j &\sim \text{Bernoulli}(\rho_j), \\ f_j|\gamma_j &\sim \gamma_j \text{Bernoulli}(\pi) + (1 - \gamma_j)\text{Bernoulli}(1 - \pi). \end{aligned}$$

- $\pi \in [0, 1]$: the level of noise
- ρ_j: prior probability of selecting s_j
- Let $\mathcal{F}$ be the set of feedback obtained
- $p(\boldsymbol{\gamma}|\mathcal{F})$ available in closed-form

ABC posterior based on $\mathcal{F}$:

$$p^{\epsilon}_{\text{ABC}}(\theta|\mathbf{y}_{\text{obs}}, \mathcal{F}) := \sum_{\boldsymbol{\gamma} \in \{0,1\}^w} p^{\epsilon}_{\text{ABC}}(\theta|\mathbf{y}_{\text{obs}}, \boldsymbol{\gamma}) p(\boldsymbol{\gamma}|\mathcal{F})$$

Sequential experimental design

Human-in-the-loop (HITL) ABC

We design a sequential Bayesian experiment to select next statistic to query.

At iteration $k+1$, we select s_{j^*}, where

$$j^* = \underset{j \notin \mathcal{J}_k}{\arg\max} \quad \mathrm{KL}[p^{\epsilon}_{\mathrm{ABC}}(\theta|\mathbf{y}_{\mathrm{obs}}, \mathcal{F}_k, \tilde{\mathcal{F}}_j) \,||\, p^{\epsilon}_{\mathrm{ABC}}(\theta|\mathbf{y}_{\mathrm{obs}}, \mathcal{F}_k)]$$

- $\mathcal{J}_k$: set of indices of statistics queried after k iterations
- Estimate KL from samples as per [Perez-Cruz, 2008]

Sequential experimental design

Human-in-the-loop (HITL) ABC

We design a sequential Bayesian experiment to select next statistic to query.

At iteration $k+1$, we select s_{j^*}, where

$$j^* = \underset{j \notin \mathcal{J}_k}{\arg\max} \quad \text{KL}[p^{\epsilon}_{\text{ABC}}(\theta|\mathbf{y}_{\text{obs}}, \mathcal{F}_k, \tilde{\mathcal{F}}_j) \,||\, p^{\epsilon}_{\text{ABC}}(\theta|\mathbf{y}_{\text{obs}}, \mathcal{F}_k)]$$

- $\mathcal{J}_k$: set of indices of statistics queried after k iterations
- Estimate KL from samples as per [Perez-Cruz, 2008]
- **Stopping criterion:** when utility of next statistic < δ
- **Output:** ABC posterior $p^{\epsilon}_{\text{ABC}}(\theta|\mathbf{y}_{\text{obs}}, \hat{\gamma})$, where

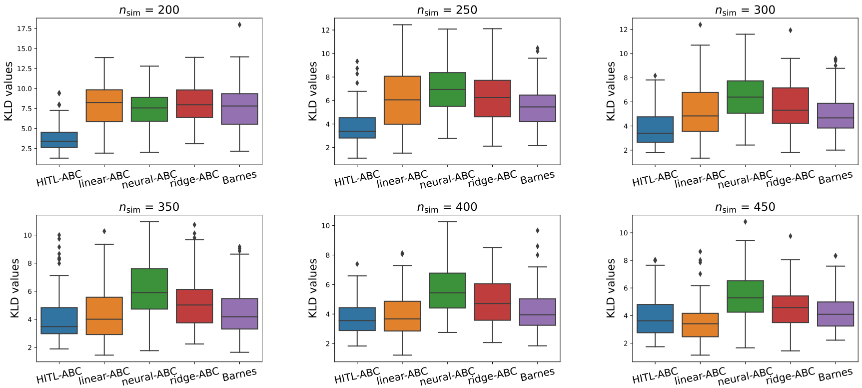
$$\hat{\gamma}_{k,j} = \begin{cases} \underset{\gamma_j \in \{0,1\}}{\arg\max}\, p(\gamma_j|f_j), & \text{if } j \in \mathcal{J}_k \\ 0, & \text{otherwise.} \end{cases}$$

Outline

Results

Low-simulation regime

- Model: g-and-k distribution (4 parameters)
- Total 15 statistics in the pool (4 informative, 6 correlated, 5 noisy)
- HITL-ABC outperforms other methods for $n_{\mathrm{sim}} \leq 350$, otherwise at par.
- Lack of simulations is compensated by expert's feedback.

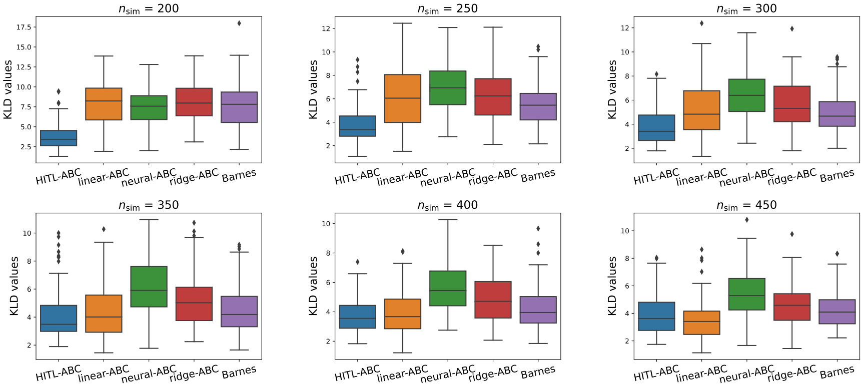


Results

Low-simulation regime

Table: Average number of expert feedback required.

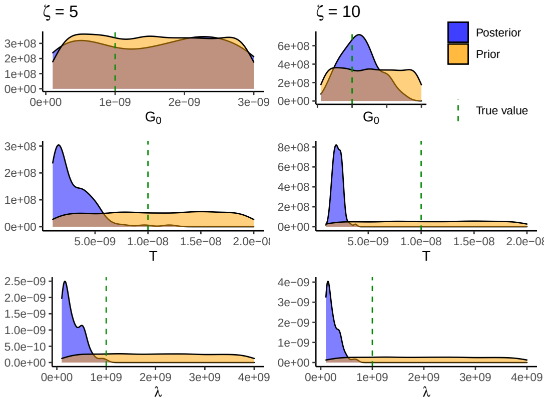
n_{sim}	200	250	300	350	400	450
HITL-ABC	**10.1**	**8.5**	**8.3**	**6.3**	**6.0**	**6.3**
Random	13.8	13.6	13.4	13.3	13.1	13.4



Results

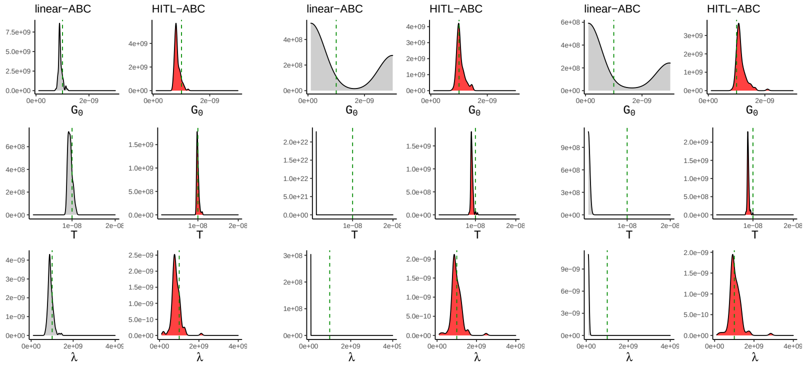
Model misspecification

- Radio propagation model (high-dimensional, complex-valued time series)
- Expert is shown inference results, and can detect misspecified statistics
- Parameters $[G_0, T, \lambda]$, 6 statistics (one mismatch)
- ζ: misspecification level



Results

Model misspecification



(a) $\zeta = 0$ (b) $\zeta = 5$ (c) $\zeta = 10$

- When $\zeta > 0$, performance of linear regression ABC seriously degrades
- Removing mismatched statistic improves performance

Conclusion

- We introduce the first ABC method that actively leverages domain knowledge from experts in order to select summary statistics.
- With fairly limited effort from the expert (answering yes/no when presented with a few statistics), we are able to outperform the regression-ABC methods in situations where the simulation budget is low.
- Involving the experts in the ABC method gives us the opportunity to handle misspecified models, something the existing methods fail in.

Akeret, J., Refregier, A., Amara, A., Seehars, S., and Hasner, C. (2015).
Approximate Bayesian computation for forward modeling in cosmology.
Journal of Cosmology and Astroparticle Physics, 2015(08):043–043.

Beaumont, M. A. (2010).
Approximate Bayesian computation in evolution and ecology.
Annual Review of Ecology, Evolution, and Systematics, 41(1):379–406.

Bharti, A., Briol, F.-X., and Pedersen, T. (2021).
A general method for calibrating stochastic radio channel models with kernels.
IEEE Transactions on Antennas and Propagation, pages 1–1.

Dyer, J., Cannon, P., Farmer, J. D., and Schmon, S. (2022).
Black-box bayesian inference for economic agent-based models.

Kopka, P., Wawrzynczak, A., and Borysiewicz, M. (2016).

Application of the approximate Bayesian computation methods in the stochastic estimation of atmospheric contamination parameters for mobile sources.
Atmospheric Environment, 145:201–212.

Kypraios, T., Neal, P., and Prangle, D. (2017).
A tutorial introduction to bayesian inference for stochastic epidemic models using approximate bayesian computation.
Mathematical Biosciences, 287:42–53.

Perez-Cruz, F. (2008).
Kullback-leibler divergence estimation of continuous distributions.
In Proceedings of the IEEE International Symposium on Information Theory, pages 1666–1670.

Pritchard, J. K., Seielstad, M. T., Perez-Lezaun, A., and Feldman, M. W. (1999).
Population growth of human y chromosomes: a study of y chromosome microsatellites.
Molecular Biology and Evolution, 16(12):1791–1798.