



## Computational Efficiency in Continuous (and Discrete!) Time Models – Comment on Hecht and Zitzmann

Charles Driver

To cite this article: Charles Driver (2021) Computational Efficiency in Continuous (and Discrete!) Time Models – Comment on Hecht and Zitzmann, Structural Equation Modeling: A Multidisciplinary Journal, 28:5, 791-793, DOI: [10.1080/10705511.2021.1877547](https://doi.org/10.1080/10705511.2021.1877547)

To link to this article: <https://doi.org/10.1080/10705511.2021.1877547>



© 2021 The Author(s). Published with license by Taylor & Francis Group, LLC.



Published online: 04 Feb 2021.



Submit your article to this journal [↗](#)



Article views: 331



View related articles [↗](#)



View Crossmark data [↗](#)

## Computational Efficiency in Continuous (and Discrete!) Time Models – Comment on Hecht and Zitzmann

Charles Driver

Max Planck Institute for Human Development

### ABSTRACT

Continuous-time models generally imply a stochastic differential equation for latent processes, coupled to a measurement model. Various computational issues can arise, and there are different estimation approaches, with different trade-offs. It has been claimed that a SEM style continuous-time model can reduce run times for Bayesian estimations of continuous-time models from hours to minutes. However this claim is not true in the general case, but requires that individuals are characterized by the same covariance and means structure, and that the number of time points is not large. While such simplifications can be valuable, and indeed are in use in existing software when appropriate, they are in general quite restrictive. The hierarchical Bayesian form of ctsem was, for instance, developed precisely to estimate models where these restrictions do not hold. To try to shed some more light on these aspects, I discuss the related issues herein.

### KEYWORDS

Continuous time; stochastic differential equation; state space

Hecht and Zitzmann (Hecht & Zitzmann, 2020) present a Bayesian SEM style approach to the estimation of continuous time dynamic systems models, and claim dramatic computational improvements over the hierarchical Bayesian instantiation (Driver & Voelke, 2018) of the ctsem (Continuous Time Structural Equation Modeling) software (Driver et al., 2017). In the form addressed, ctsem offers a hierarchical Bayesian approach to continuous time state space modeling, wherein dynamic system and measurement equation parameters can be estimated for each individual. These parameters are modeled as arising from a higher level multivariate population distribution, as well as possible covariates. Kalman filtering procedures (likelihood calculations conditional on the previous time point) are used to integrate across the unknown states of the latent processes – in a CT (continuous time) conception, latent states are unknown at the specific instant of observation, but also unknown across the time interval between observations. In contrast to discrete time counterparts such as cross lagged panel models, the CT approach accounts for timing differences between measurement occasions, and also allows for an accurate representation of typically hypothesized data generating processes in longitudinal contexts (Aalen et al., 2012). Correcting these problems inherent to the discrete time approach thereby allows interpretable hypothesis testing and regularization opportunities with respect to a (typically) more realistic data generating process. In further contrast to typical SEM style models, ctsem allows individual differences across all parameters of the model, and individual differences in the timing of observations. The downside of such flexibility is that there is limited scope to re-use calculations from one individual for further individuals, and computational time (per likelihood evaluation) scales roughly linearly with increasing subjects.

Hecht and Zitzmann's claimed computational improvements largely arise from restricting individual differences to those which can be easily integrated out (i.e. intercept parameters), and ensuring that all individuals share the same pattern of observation timing. In such a case, the model implied means and covariance matrix only needs to be computed for a single individual, and scaling up to many individuals is relatively inexpensive. Hecht and Zitzmann briefly point out that they did not examine these issues, but appear content to propagate general conclusions regarding the superiority of their favored approach regardless. This approach they favor is actually what is used in the original mixed effects implementation of ctsem (Driver et al., 2017), which used OpenMx to implement CT models using a RAM (McArdle, 2005) formulation of SEM. In the case that individuals all share the same observation timing pattern, then OpenMx detects that there is no need to re-compute the model expectations, and performance can be very fast, as seen in the results from Hecht and Zitzmann. The only change required for Bayesian estimation then is amending a prior on top of the likelihood calculation; then, one may use a variety of estimation approaches (i.e. various forms of optimization and/or sampling) to attain results of interest regarding the model fit. Beyond the individual differences aspect, a further complicating issue for SEM style approaches, in which the multivariate likelihood equation for each subject includes every time point, is that as the number of time points grows, so too does the necessary matrix inversion. As such, as time points increase, computational costs increase faster (roughly  $O(n^3)$ , dependent on algorithms), and at some point, filtering approaches become more efficient. The complexity of filtering approaches typically scales linearly with number of time points, because the conditional likelihood calculation only involves (at most) the

number of distinct measurement variables used at a specific time point. The exact number of measurement occasions when filtering approaches become computationally more efficient will depend on the specific context (e.g., number of variables, time points, possibility to re-use calculations, computational libraries), but is often not very high. Hecht and Zitzmann do offer some speculative suggestions that their favored SEM style approach is most beneficial with larger numbers of subjects, but miss the point that this also depends on restrictive assumptions regarding individual differences, and limited time points. For more on filtering versus SEM approaches, see (Chow et al., 2010) and (Hunter, 2018).

So, the examples that Hecht and Zitzmann examined are perfect candidates for a SEM style approach, as there are lots of subjects, no individual differences in the expected means and covariances, and few time points. However, improved performance of a SEM approach under such conditions is not specific to CT models – the same would hold true across all forms of computational modeling when you contrast an approach with individual differences in expectations versus an approach without. Such computational differences are, as mentioned, already leveraged in software such as OpenMx (Neale et al., 2016) to maximize performance.

Regarding ctsem specifically, in its present state (v3.3.8) it offers a range of estimation approaches, and the ability to easily enable priors when Bayesian (or some form of penalized) estimation is desired. The default is for maximum likelihood estimation using a form of hybrid (discrete observations, continuous time) extended Kalman filter (Mazzoni, 2008), in which individuals system and measurement parameters may all vary as correlated random effects, fixed effects via covariates, or some combination. In such a case, the random components of the individual differences are integrated out by extending the latent states. The maximum likelihood is sought using optimization approaches, with standard errors and confidence intervals based on the estimated Hessian at the maximum. For improved uncertainty quantification, adaptive importance sampling (Oh & Berger, 1992) can be requested, which uses the maximum likelihood (or posterior mode when priors are used) and estimated asymptotic covariance matrix as an initial proposal distribution. Limitations to this approach include the fact that the likelihood is only approximate when nonlinear random effects (e.g., individual differences in a correlation parameter) are requested, and the adaptive importance sampler (when switched on) has trouble converging when large numbers of parameters are involved. The best approach (given unlimited time/computational resources) is typically the one Hecht and Zitzmann focus on, which uses the Bayesian sampler offered by Stan. In this case the nonlinear random effects are approximated more accurately because the individual parameters are sampled directly and not based on linearizations, and the sampler in Stan offers sophisticated diagnostics. However, as noted, with large numbers of subjects or large systems this quickly becomes intractable, and the alternatives discussed are preferable, at least during model development.

It is true that the current range of options is not perfect for such limited (no covariates, no timing differences, random effects on intercepts only, few time points) cases as in the examples shown by Hecht and Zitzmann – tailoring an approach for a specific

problem will often yield performance gains over general purpose approaches. Focusing on the filtering approach for the modern form of ctsem was a design decision to allow for greater flexibility in directions where typical SEM is very limited. However, my hope for the future would be to see works that either extend current software to the specific use cases desired, develop algorithms for detecting when specific algorithms should be used, or offer genuine improvements to either a) the underlying computational procedures and libraries, or b) computationally simpler modeling approaches that offer similar or improved options for scientific inference. To give an example of one specific aspect, the major computational bottleneck in ctsem (and, as far as I understand, other stochastic differential-equation-based modeling approaches) is the need to repeatedly solve the stochastic differential equation for different time intervals and system parameters. The primary approach ctsem uses for this is based on matrix exponentiation, as this is near perfect (up to accuracy of the exponential algorithm) for linear systems and can have higher accuracy for moderately non-linear systems (Hochbruck et al., 1998). However, this is likely inefficient for large and sparse (i.e. few connections) systems, and ‘could’ be considered unnecessarily accurate for many cases. At present one alternative is available within ctsem, using a Taylor–Heun-approximation of the vector field and a modified Gauss–Legendre-scheme for the covariance (Mazzoni, 2008), but testing of this is very limited and it is unclear under which circumstances ‘good enough’ answers are obtained. For those interested in such projects, I would encourage making contact with those developing the approaches and software directly, as there are always many possibilities and many nuances!

## References

- Aalen, O. O., Røysland, K., Gran, J. M., & Ledergerber, B. (2012). Causality, mediation and time: A dynamic viewpoint. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 175, 831–861. <https://doi.org/10.1111/j.1467-985X.2011.01030.x>
- Chow, S.-M., Ho, M.-H. R., Hamaker, E. L., & Dolan, C. V. (2010). Equivalence and differences between structural equation modeling and state-space modeling techniques. *Structural Equation Modeling*, 17, 303–332. <https://doi.org/10.1080/10705511003661553>
- Driver, C. C., Oud, J., & Voelke, M. C. (2017). Continuous time structural equation modeling with r package ctsem. *Journal of Statistical Software, Articles*, 77, 1–35. <https://doi.org/10.18637/jss.v077.i05>
- Driver, C. C., & Voelke, M. C. (2018). Hierarchical bayesian continuous time dynamic modeling. *Psychological Methods*, 23, 774. <https://doi.org/10.1037/met0000168>
- Hecht, M., & Zitzmann, S. (2020). A computationally more efficient bayesian approach for estimating continuous-time models. *Structural Equation Modeling: A Multidisciplinary Journal*, 27, 1–12. <https://doi.org/10.1080/10705511.2020.1779069>
- Hochbruck, M., Lubich, C., & Selhofer, H. (1998). Exponential integrators for large systems of differential equations. *SIAM Journal on Scientific Computing*, 19, 1552–1574. <https://doi.org/10.1137/S1064827595295337>
- Hunter, M. D. (2018). State space modeling in an open source, modular, structural equation modeling environment. *Structural Equation Modeling: A Multidisciplinary Journal*, 25, 307–324. <https://doi.org/10.1080/10705511.2017.1369354>
- Mazzoni, T. (2008). Computational aspects of continuous–discrete extended Kalman-filtering. *Computational Statistics*, 23, 519–539. <https://doi.org/10.1007/s00180-007-0094-4>
- McArdle, J. J. (2005). *The development of the ram rules for latent variable structural equation modeling*. In A. Maydeu-Olivares & J. J. McArdle (Eds.), *Multivariate applications book series. Contemporary*

- psychometrics: A festschrift for Roderick P. McDonald* (p. 225–273). Lawrence Erlbaum Associates Publishers.
- Neale, M. C., Hunter, M. D., Pritikin, J. N., Zahery, M., Brick, T. R., Kirkpatrick, R. M., Estabrook, R., Bates, T. C., Maes, H. H., & Boker, S. M. (2016). Openmx 2.0: Extended structural equation and statistical modeling. *Psychometrika*, 81, 535–549. <https://doi.org/10.1007/s11336-014-9435-8>
- Oh, M.-S., & Berger, J. O. (1992). Adaptive importance sampling in monte carlo integration. *Journal of Statistical Computation and Simulation*, 41, 143–168. <https://doi.org/10.1080/00949659208810398>