

Cgmquantify: Python and R Software Packages for Comprehensive Analysis of Interstitial Glucose and Glycemic Variability from Continuous Glucose Monitor Data

Brinnae Bent , Maria Henriquez, and Jessilyn P. Dunn , *Member, IEEE*

Abstract—Goal: Continuous glucose monitoring (CGM) is commonly used in Type 1 diabetes management by clinicians and patients and in diabetes research to understand how factors of longitudinal glucose and glucose variability relate to disease onset and severity and the efficacy of interventions. CGM data presents unique bioinformatic challenges because the data is longitudinal, temporal, and there are infinite ways to summarize and use this data. There are over 25 metrics of glucose variability used clinically and in research, metrics are not standardized, and little validation exists across studies. The primary goal of this work is to present a software resource for systematic, reproducible, and comprehensive analysis of interstitial glucose and glycemic variability from continuous glucose monitor data. **Methods:** Comprehensive literature review informed the clinically-validated functions developed in this work. Software packages were developed and open-sourced through the Python Package Index (PyPi) and the Comprehensive R Archive Network (CRAN). *cgmquantify* is integrated into the Digital Biomarker Discovery Pipeline and MD2K Cerebral Cortex. **Results:** Here we present an open-source software toolbox called *cgmquantify*, which contains 25 functions calculating 28 clinically validated metrics of glucose and glycemic variability, as well as tools for visualizing longitudinal CGM data. Detailed documentation facilitates modification of existing code by the community for customization of input data and visualizations. **Conclusions:** We have built systematic functions and documentation of metrics and visualizations into a software resource available in both the Python and R languages. This resource will enable digital biomarker development using continuous glucose monitors.

Index Terms—Continuous glucose monitor, diabetes, digital biomarkers, glycemic variability, open-source software.

I. INTRODUCTION

Continuous glucose monitoring (CGM) systems provide real-time, dynamic glucose information by tracking interstitial glucose values throughout the day. CGMs are commonly used in Type 1 diabetes (T1D) management, with 1.2 million diabetic patients using a CGM [1], [2]. These devices have been used extensively by the T1D community, including in the Open Artificial Pancreas System Project (OpenAPS) [3], a project developed to create a patient-implemented closed loop system between a CGM and an insulin pump.

CGM data is commonly provided from CGM manufacturers as either raw glucose values (in a .csv format) or in summary reports that utilize proprietary methods to plot and summarize glucose statistics. Because these algorithms are proprietary, they cannot be externally validated by clinical researchers [4]. Additionally, the provided glucose summaries are limited and do not usually contain any information about an important clinical metric, glycemic variability.

Glycemic variability, also known as glucose variability, is an established risk factor for hypoglycemia [5] and has been shown to be a risk factor in other diabetes complications [6]. Glucose variability is mentioned in over 26000 publications indexed in PubMed at the time of this publication and is a key metric in clinical research [7]. Over 25 metrics of glucose variability have been used in the literature, which makes it difficult to examine and compare results across numerous research studies analyzing and drawing conclusions about glucose variability.

There is a need for an open-source software toolbox containing algorithms that are utilized and validated in clinical research studies. This would enable standardized glucose variability metrics and the ability to compare findings from studies that utilize different metrics of glucose variability. Its availability in an open-source programming language with a low barrier to entry will encourage researchers, clinicians, and patients alike to explore CGM data.

Manuscript received March 29, 2021; revised May 18, 2021 and August 6, 2021; accepted August 7, 2021. Date of publication August 18, 2021; date of current version September 14, 2021. This work was supported in part by the Duke MedX and in part by the Chan Zuckerberg Initiative DAF, an advised fund of Silicon Valley Community Foundation under Grant 2020-218599. The review of this letter was arranged by Editor P. Bonato. (*Corresponding author: Jessilyn P. Dunn.*)

Brinnae Bent and Jessilyn P. Dunn are with the Department of Biomedical Engineering, Duke University, Durham, NC 27705 USA (e-mail: brinnae.bent@duke.edu; jessilyn.dunn@duke.edu).

Maria Henriquez is with the Department of Statistical Science, Duke University, Durham, NC 27705 USA (e-mail: maria.henriquez@duke.edu).

Digital Object Identifier 10.1109/OJEMB.2021.3105816

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see <https://creativecommons.org/licenses/by/4.0/>

TABLE 1.
COMPARISON TO OTHER OPEN-SOURCE CGM SOFTWARE

CGM Software	Number of metrics that are clinically validated	Functions for plots?	Python	R	Access
cgmquantify	28	Yes	Yes	Yes	PyPi CRAN DBDP MD2K
EasyGV[8]	12	No	No	No	Need to request access
cgmanalysis [9]	21	Yes	No	Yes	CRAN
CGManalyzer [10]	7	Yes	No	Yes	CRAN

Previous open-source resources have been implemented in Excel [8] and R [9], [10] (Table 1). There is currently no comprehensive resource for CGM data in Python, the third most common programming language used globally and the leading language among newcomers to programming [11]. Additionally, previous implementations of open source CGM data analysis present limited metrics of glucose variability. Further, these methods are typically developed for a research specific purpose and without sufficient documentation to be extended to other purposes.

The primary goal of this work is to provide a comprehensive, open-source software resource available across both the Python and R programming languages for *systematic and reproducible* analysis of continuous glucose monitor data.

II. MATERIALS AND METHODS

cgmquantify is an open-source software Python package and R package composed of 25 functions with 28 clinically validated metrics of glucose and glucose variability, as shown in Table 2. These packages are published under the MIT license.

Customizable visualizations (Fig. 1, Fig. 2) are also included as easy to implement functions.

cgmquantify is version controlled through GitHub, the Python Package Index (PyPI), and the Comprehensive R Archive Network (CRAN). This allows for single-line installation in either language. Source code, an extensive user guide, and issue tracking are available on GitHub to facilitate ease of use and enable customization based on user needs. Tests are available in GitHub under the tests subdirectory to allow for manual testing of all functions. Results of tests of the current software versions are available in GitHub. The current software passes all tests. The community can expand and contribute code to cgmquantify through the DBDP (cite).

The cgmquantify python package is available in PyPi (<https://pypi.org/project/cgmquantify/>). Source code can be found at <https://github.com/DigitalBiomarkerDiscoveryPipeline/cgmquantify>. The cgmquantify R package is available

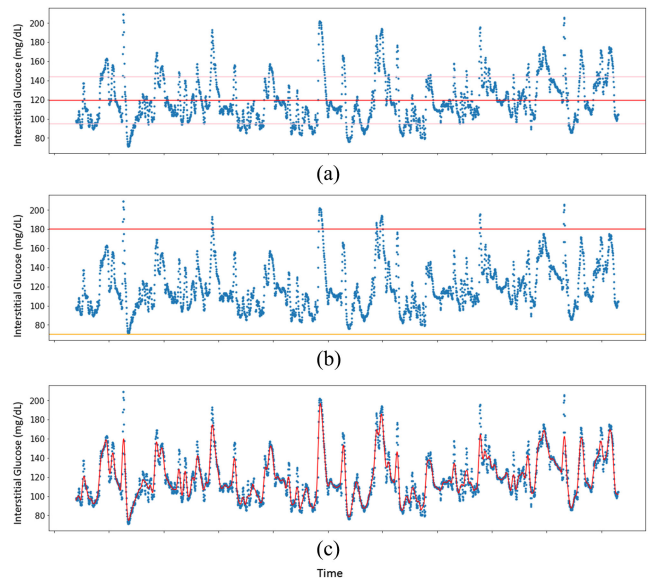


FIGURE 1. Visualizing longitudinal CGM data with the cgmquantify Python package. Shown are a visualization with indicators of the mean interstitial glucose level (red) and 1 SD from the mean (pink) (a), a visualization with indicators of hyperglycemic (>180 mg/dL glucose, red) and hypoglycemic (<70 mg/dL glucose, yellow) (b), and a plot with LOWESS smoothing of the glucose data (c).

in CRAN (<https://CRAN.R-project.org/package=cgmquantify>) and source code can be found at <https://github.com/DigitalBiomarkerDiscoveryPipeline/cgmquantify-R>.

III. RESULTS

We have included import functions to format data for use with the cgmquantify package. These functions currently support Dexcom and Abbott Freestyle Libre CGM devices, with ongoing work adding data import functions for other CGM manufacturers (e.g., Medtronic). Our user guide also outlines how new data can be easily formatted data to make it compatible with the functions in cgmquantify.

Functions are available for all of the commonly studied glucose and glucose variability metrics (Table 2). Additionally, functions for data visualization of the longitudinal CGM data are provided. These visualizations are easily customizable and can include personalized or standard clinical thresholds for hyper- and hypoglycemia. We have also implemented a function that enables LOWESS smoothing over the CGM data to facilitate viewing and interpretation (Fig. 1).

We have integrated cgmquantify in the Digital Biomarker Discovery Pipeline, an open-source software resource aimed at making digital biomarker development and digital medicine more accessible [12]. cgmquantify has also been integrated in MD2K Cerebral Cortex in the MD2K Center of Excellence for turning mobile sensor data into reliable and actionable health information [13]. The cgmquantify software package is already being used in several research projects at multiple institutions.

TABLE 2.
GLUCOSE AND GLUCOSE VARIABILITY METRICS

Metric	Description	Equation
interdaySD[6], [7]	Interday standard deviation of glucose	$\sigma_{interday} = \sqrt{\frac{\sum(G_i - \mu)^2}{N}}$ Where N = total days, G = glucose value
interdayCV[7]	Interday coefficient of variation of glucose	$CV_{interday} = \frac{\sigma_{interday}}{\mu}$
intradaySD Mean[7]	Intraday standard deviation of glucose (mean across all days)	$\sigma_{intraday_{mean}} = \sum_N \sigma_i$
intradayCV Mean[7]	Intraday coefficient of variation of glucose (mean across all days)	$CV_{intraday_{mean}} = \sum_N CV_i$
intradaySD Median[15]	Intraday standard deviation of glucose (median across all days)	$\sigma_{intraday_{median}} = \text{median}(\sigma_i)$
intradayCV Median[15]	Intraday coefficient of variation of glucose (median across all days)	$CV_{intraday_{median}} = \text{median}(CV_i)$
intradaySD Standard Deviation[15]	Intraday standard deviation of glucose (standard deviation across all days)	$\sigma_{intraday_{standard\ deviation}} = SD(\sigma_i)$
intradayCV Standard Deviation[15]	Intraday coefficient of variation of glucose (standard deviation across all days)	$CV_{intraday_{standard\ deviation}} = SD(CV_i)$
CONGA24[6], [16]	Continuous overall net glycemic action over 24 hours	$CONGA24 = SD(G_t - G_{t-24hours})$
GMI[10]	Glucose management indicator	$GMI = 3.31 + (0.02392 * \mu \text{ (mg/dL)})$
HbG1[6], [7]	High Blood Glucose Index	$\sum \frac{r_1}{n}$ $r_1 = 22.7 * (f(G_i))^2 \text{ if } f(G_i) \leq 0, r_1 = 0 \text{ otherwise}$ $f(G_i) = \ln(G_i)^{1.084} + 5.381$
LBGI[6], [7]	Low Blood Glucose Index	$\sum \frac{r_h}{n}$ $r_h = 22.7 * (f(G_i))^2 \text{ if } f(G_i) > 0, r_h = 0 \text{ otherwise}$ $f(G_i) = \ln(G_i)^{1.084} + 5.381$
ADRR[7]	Average Daily Risk Range, assessment of total daily glucose variations within risk space	$ADRR = \frac{\sum_{all\ days} (LR^j + HR^j)}{N\ days}$ where $LR^j = \max(r_l)$ and $HR^j = \max(r_h)$
J-index[6], [17]	Measure of both the mean level and variability of glycemia	$J = 0.001 * (\mu + \sigma)^2$
MAGE[6], [18]	Mean amplitude of glucose excursions (default = 1SD)	1. Local minima/maxima determined 2. Assess max/min pairs against SD 3. If difference from min to max > SD, mean is retained 4. Otherwise excluded Troughs are retained and summed
MGE[6]	Mean of glucose outside range (default = 1SD)	$\mu_{glucose\ outside\ \#SD\ of\ mean}$ Where # is set, default is 1 SD
MGN*	Mean of glucose inside range (default = 1SD)	$\mu_{glucose\ inside\ \#SD\ of\ mean}$
MODD[7], [16]	Mean of daily differences in glucose	$MODD = \frac{\sum G_t - G_{t-24hours} }{total\ matched\ observations}$
TIR[19]	Time spent in range (minutes), default = 1SD	$TIR = \sum_N \text{time inside } \#SD\ of\ mean$
TOR[19]	Time spent outside range (minutes), default = 1SD	$TOR = \sum_N \text{time outside } \#SD\ of\ mean$
POR[9]	Percent of time spent outside range	$POR = \frac{TOR}{total\ time} * 100\%$
PIR[9]	Percent of time spent inside range, default = 1SD	$PIR = \frac{TIR}{total\ time} * 100\%$
eA1c[20]	Estimated A1c (according to American Diabetes Association)	$eA1c = \frac{(46.7 + \mu)}{28.7}$
meanG[16]	Mean glucose over all days	$\mu = \frac{\sum_N \bar{x}_i}{N}$
medianG[9], [10]	Median glucose over all days	median (Glucose)
minG[9]	Minimum glucose over all days	min (Glucose)
maxG[9]	Maximum glucose over all days	max (Glucose)
Q1G[9]	First quartile glucose value over all days	first quartile (Glucose)
Q3G[9]	Third quartile glucose value over all days	third quartile (Glucose)

*indicates previously unidentified metrics of glucose variability that are similar to clinically validated metrics

IV. DISCUSSION

cgmquantify is a package that simplifies the process of calculating interstitial glucose metrics from CGM data and thus allows for easy comparison across different research studies that use different metrics summarizing glucose and glucose

variability. Functions have been developed using equations from clinically validated research studies such that users can compare their results to previous findings. The cgmquantify package is easily implemented with a one-line installation and contains an extensive user guide in both the Python and R languages. Detailed documentation facilitates modification of existing code

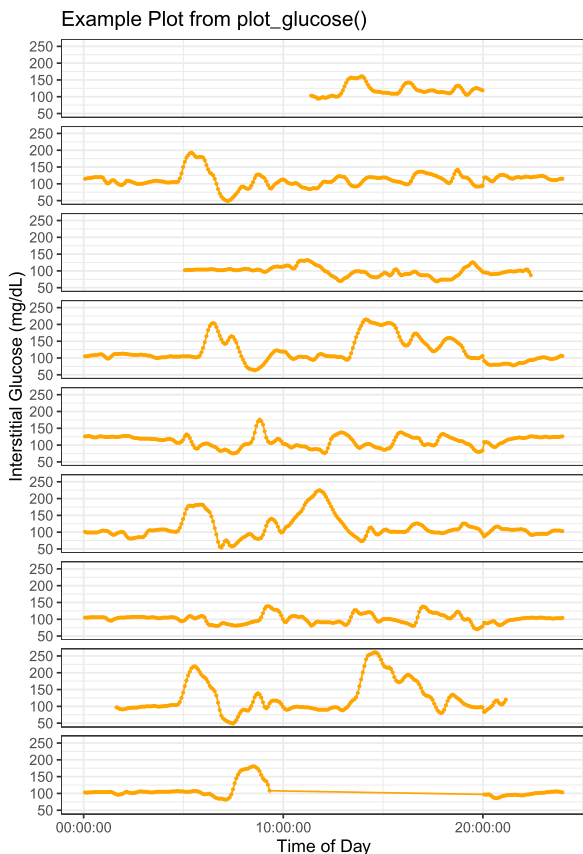


FIGURE 2. Example plot from the `plot_glucose()` function from the `cgmquantify` R package to visualize longitudinal CGM data. Shown is a visualization available in the `cgmquantify` R package that enables visualization of CGM data by time of day for each day specified.

for customization of input and visualizations. With the aim of build a community of developers to contribute to the literature in this burgeoning field, we anticipate that many researchers will not only find this resource useful for their own work but will contribute code updates and improvements through the DBDP to ensure that this package remains both relevant and useful for the community.

This is a much-needed resource for the community of researchers, clinicians, and patients using CGM. Currently, little is understood about the relationships between glucose and glucose variability metrics from CGM data to diseases including but not limited to prediabetes, Type 2 diabetes, and severity of symptoms in T1D. As more researchers and clinicians start looking to CGM data to answer these questions, the need for a standardized resource available in the most common programming languages is necessary. As we have seen with the Open APS community, analysis of CGM data is not limited to researchers and clinicians but includes patients as well [14]. By providing this toolbox as an open-source resource, we hope to encourage patients to interact with their own data, determine personalized insights, and make meaningful contributions to the digital health landscape.

While the primary use case for `cgmquantify` is in research, in future directions we may incorporate a graphical user interface to enable ease of use outside of the Python or R programming languages.

V. CONCLUSION

The `cgmquantify` software resource enables comprehensive analysis of continuous glucose monitor data with clinically validated metrics and easy to implement visualizations. Future implementations include incorporating food diary and physical activity data into the visualizations. The fields of digital biomarker development and diabetes research will benefit greatly from this resource.

REFERENCES

- [1] H. Wolinsky, "Continuous glucose monitors: The long and winding road to acceptance, coverage managed care magazine," *Managed Care*, 2018. Accessed: Feb. 15, 2020. [Online]. Available: <https://www.managedcaremag.com/archives/2018/7/continuous-glucose-monitors-long-and-winding-road-acceptance-coverage>
- [2] W. V. Tamborlane *et al.*, "Continuous glucose monitoring and intensive treatment of type 1 diabetes," *N. Engl. J. Med.*, vol. 359, no. 14, pp. 1464–1476, Oct. 2008.
- [3] "OpenAPS.org – #WeAreNotWaiting to reduce the burden of type 1 diabetes." Accessed: Feb. 15, 2020. [Online]. Available: <https://openaps.org/>
- [4] J. C. Goldsack *et al.*, "Verification, analytical validation, and clinical validation (V3): The foundation of determining Fit-for-purpose for Biometric Monitoring Technologies (BioMeTs)," *npj Digit. Med.*, vol. 3, pp. 55, 2020.
- [5] G. E. Umpierrez and B. P. Kovatchev, "Glycemic variability: How to measure and its clinical implication for type 2 diabetes," *Amer. J. Med. Sci.*, vol. 356, no. 6, pp. 518–527, Dec. 2018.
- [6] F. J. Service, "Glucose variability," *Diabetes*, vol. 62, no. 5, pp. 1398–1404, May 2013.
- [7] B. P. Kovatchev, "Metrics for glycaemic control — From HbA1c to continuous glucose monitoring," *Nat. Rev. Endocrinol.*, vol. 13, no. 7, pp. 425–436, Jul. 2017.
- [8] N. R. Hill, N. S. Oliver, P. Choudhary, J. C. Levy, P. Hindmarsh, and D. R. Matthews, "Normal reference range for mean tissue glucose and glycemic variability derived from continuous glucose monitoring for subjects without diabetes in different ethnic groups," *Diabetes Technol. Ther.*, vol. 13, no. 9, pp. 921–928, Sep. 2011.
- [9] T. Vigers *et al.*, "cgmanalysis: An R package for descriptive analysis of continuous glucose monitor data," *PLoS One*, vol. 14, no. 10, Oct. 2019, Art. no. e0216851.
- [10] X. D. Zhang, Z. Zhang, and D. Wang, "CGManalyzer: An R package for analyzing continuous glucose monitoring studies," *Bioinformatics*, vol. 34, no. 9, pp. 1609–1611, May 2018.
- [11] P. Krill, "Python popularity reaches an all-time high | infoworld," 2019. Accessed: Feb. 12, 2020. [Online]. Available: <https://www.infoworld.com/article/3401536/python-popularity-reaches-an-all-time-high.html>
- [12] B. Bent *et al.*, "The digital biomarker discovery pipeline: An open source software platform for the development of digital biomarkers using mHealth and wearables data," *J. Clin. Transl. Sci.*, no. 11, pp. 1–28, 2020.
- [13] "M. D.2K. GitHub," *GitHub*, 2020. Accessed: Feb. 25, 2020. [Online]. Available: <https://github.com/MD2Korg/>
- [14] M. De Groot, M. Drangsholt, F. Martin-Sanchez, and G. Wolf, "Single subject (N-of-1) research design, data processing, and personal science," *Methods Inf. Med.*, vol. 56, no. 6, pp. 416–418, Feb. 2017.
- [15] P. Cho *et al.*, "Expanding the definition of intraday glucose variability," *Diabetes*, vol. 69, pp. 73–LB, 2020.
- [16] R. A. Rawlings, H. Shi, L. H. Yuan, W. Brehm, R. Pop-Busui, and P. W. Nelson, "Translating glucose variability metrics into the clinic via continuous glucose monitoring: A graphical user interface for diabetes evaluation (CGM-GUIDE)," *Diabetes Technol. Ther.*, vol. 13, no. 12, pp. 1241–1248, Dec. 2011.
- [17] J. M. Wojcicki, "'J'-index: A new proposition of the assessment of current glucose control in diabetic patients," *Horm. Metab. Res.*, vol. 27, no. 1, pp. 41–42, 1995.
- [18] S. Suh and J. H. Kim, "Glycemic variability: How do we measure it and why is it important?," *Diabetes Metab. J.*, vol. 39, no. 4, pp. 273–282, 2015.
- [19] B. Kovatchev, "Glycemic variability: Risk factors, assessment, and control," *J. Diabetes Sci. Technol.*, vol. 13, no. 4, pp. 627–635, Jul. 2019.
- [20] "eAG/A1C conversion calculator American Diabetes Association." Accessed: Feb. 15, 2020. [Online]. Available: https://professional.diabetes.org/diapro/glucose_calc