TIME-SERIES FORECASTING: A THEORETICAL MODEL FOR PREDICTING PERFORMANCE POTENTIAL

Joseph Walters[1], S. Kyle Travis[1], Asher Flynn[1], Paul Moqiun[1], Austin Smith[1]

[1]Center of Excellence for Sport Science and Coach Education, Department of Sport, Exercise, Recreation, and Kinesiology, East Tennessee State University, Johnson City, TN, USA

INTRODUCTION: Analyzing the training process to enhance performance is paramount in aiding understanding of athletic potential. Thus, sport scientists and coaches have unfolded novel approaches to facilitate the understanding of enhancing full athletic capabilities. Mathematical modeling may provide a means for the sport scientist to aid in this endeavor. Indeed, mathematical modeling, specifically time series forecasting, has existed for some time (Lewis, 2017). Using prior data trends over a series of time, forecasting future performance outcomes can be predictable to a level of certainty (e.g., 80 – 95 CI) (Hyndman & Athanasopoulos, 2017; Lewis, 2017). Bannister, Calvert, Savage, & Bach (1975) and T. Busso et al., (1992) pioneered the use of mathematical modeling in sport to predict physiological changes that gauge possible performance outcomes for athletes. Theoretical approaches for training adaptations in sport are based on an antagonistic understanding of training effects (i.e., positive training effects, negative training effects) that present an expected result for actual performance. Having a grasp on the basic structure and building blocks of the training process, as well as understanding system functionality for mathematical modeling could be useful for optimizing the planning of the training strategy. Therefore, the purpose of this paper was to expound upon basic mathematical models that can be easily applied to the training process to predict desired performance outcomes.

TOTAL TRAINING LOAD: One of the most useful and simplest variables sport scientists and coaches can measure is total training load (TTL). TTL can be measured internally (e.g. heart rate, questionnaire) and externally (e.g. volume load from strength training, conditioning, or competition) (Halson, 2014). TTL is one of the most important determinant factors for performance outcomes and it should be monitored during all training and competition modes over a period of weeks and months to observe training and performance trends (Gabbett, 2016). A simple method for obtaining TTL is the use of session rate of perceived exertion (sRPE) multiplied by duration (i.e., minutes) of the training session, practice, competition, or game. sRPE is a subjective, internal measure provided by the athlete from a bout of training and can provide a valid measure of difficulty from the session (Wallace, Slattery, & Coutts, 2009). This simple method is the *input* information put into the system. Once the system processes this information, it can reveal an indirect measure of the output information (i.e., positive or negative performance outcomes).

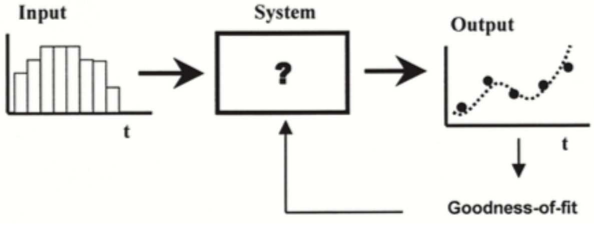
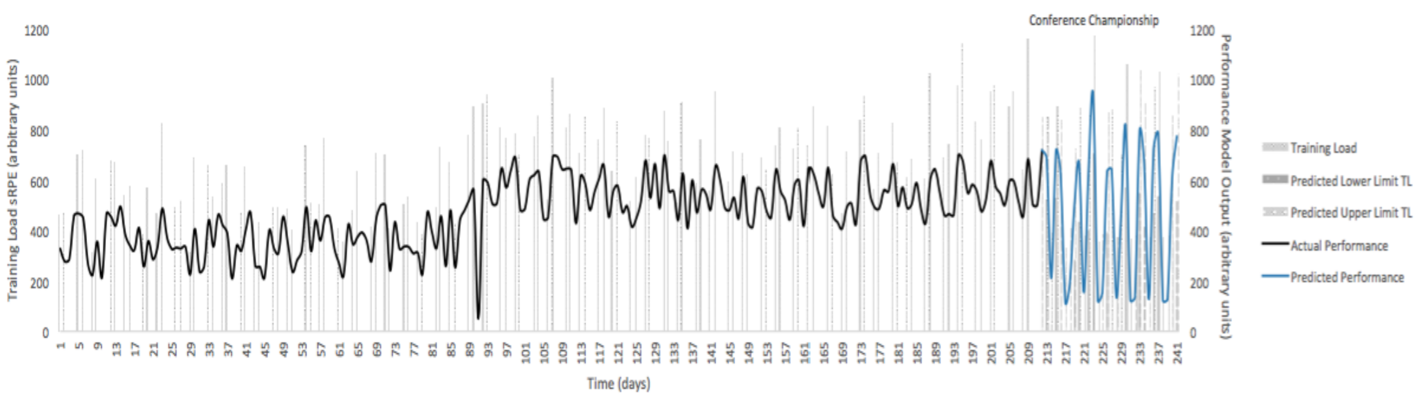


Figure 1. Adapted from Busso and Thomas, 2006. Schematic representation of the procedure for specific characterization of a dynamical process. The model parameters are estimated by fitting model output to actual previous data.

This model is dependent on the two-component input information (i.e., sRPE x duration) to derive the mean potential output by fitting to the already collected data. By utilizing these factors, sport scientists and coaches could predict internal load responses from external workloads to determine the potential performance level. Composed of two antagonistic first-order transfer functions, this model allows proper representation of the training responses derived from an average of the input in order to predict an accurate performance level for an athlete or team to contest. The quality of this prediction is measured by the mean differences between each factor from the previous weeks of training with a variation of 4.9% agreeing with the variations described by Torrents, Balagué, Perl and Schöllhorn (2007). Torrents et al. (2007) also state that actual performance variation can change up to 10%. The actual performance data is measured by team ranking, schedule strength, total points per game or practice, point attempts, etc. However, the authors suggest observing and inputting data over a full season, repeatedly, to reduce the variation within the output prediction to increase accuracy and goodness-of-fit.

THEORETICAL MODEL: The experimental data presented in this report is theoretical and based on training load variations that would be observed in team sports that tend to have extended seasons (i.e., basketball). The forecast method used is a Box and Cox transformation, ARMA errors, Trend and Seasonal components, also known as the BATS model. This approach is a combination of other models (e.g., moving average, Holt-Winters) and can capture seasonal trends in daily, weekly, and monthly sequences and improves forecast accuracy concerning daily analysis. This can be particularly useful for sport scientists and coaches allowing accurate adjustments to daily workloads (e.g., practice, strength training sessions, conditioning sessions, etc.). By utilizing the prediction forecast, the upcoming weeks in training can allow for



appropriate load management, therefore enhancing preparedness.

Figure 2. Daily TL variations and actual performance trending towards conference championship.

DISCUSSION: Because computational power has magnified in recent decades, time series forecasting can be used in sport with fundamental knowledge of data science. Sport scientists and coaches can use the above model to aid them in daily and weekly forecasting of training. This model can be considered an extension of the classical Banister model; however, the variation for performance is based on the TTL. The most frequently used model in the current literature predicting performance effects from various forms of TTL is from Bannister et al., (1975). Accentuating the importance of TTL, the authors highlight various investigations that use numerous methods for this variable. For example, Banister, Carter, & Zarkadas (1999) validated the performance of his model with an elite triathlete, Calvert, Banister, Savage, & Bach (1976) presented a case of a swimmer, studies in cross-country skiers were done by Candau, Busso, & Lacour (1992), constructing training programs through previous training and performance trends from Fitz-Clarke, Morton, & Banister (1991), and running performance trends analyzed by Morton, Fitz-Clarke, & Banister (1990). Additional recent contributions for utilizing mathematical modeling include predicting exercise and performance by Thierry Busso & Thomas (2006), as well as using regression techniques, neural networks, and times series forecasting by Pfeiffer (2005) and Pfeiffer & Hohmann (2012). Applying computational intelligence in sports for predicting performance have been defended indubitably (Fister, Ljubič, Suganthan, Perc, & Fister, 2015).

However, sport scientist and coaches that are considering using time series forecasting should understand that certain models have limitations in prediction accuracy, and choosing the correct model may require trial and error to determine the best fit. As previously suggested, collecting more data points will increase the accuracy of the model and therefore correct any outstanding errors. This report shows that it is possible to obtain an exceptional fit with the respected experimental data, calculating high and low TTL relating to performance. This model can be used to predict any sort of performance level depending on how the investigator defines each variable. Mathematical modeling provides an additional tool to construct training plans, minimize fatigue, predict performance outcomes, and optimize an individual athlete or team's ability to compete at their highest level.

REFERENCES

Banister, E. W., Carter, J. B., & Zarkadas, P. C. (1999). Training theory and taper: validation in triathlon athletes. *European Journal of Applied Physiology and Occupational Physiology*, *79*(2), 182–191. https://doi.org/10.1007/s004210050493

Bannister, E., Calvert, T., Savage, M., & Bach, T. (1975). A systems model of training for athletic performance. *Aust J Sports Med*, *7*, 57–61.

Busso, T., Häkkinen, K., Pakarinen, A., Kauhanen, H., Komi, P. V., & Lacour, J. R. (1992). Hormonal adaptations and modelled responses in elite weightlifters during 6 weeks of training. *European Journal of Applied Physiology and Occupational Physiology*, *64*(4), 381–386. https://doi.org/10.1007/BF00636228

Busso, Thierry, & Thomas, L. (2006). Using mathematical modeling in training planning. *International Journal of Sports Physiology and Performance*, *1*(4), 400–405.

Calvert, T. W., Banister, E. W., Savage, M. V., & Bach, T. (1976). A Systems Model of the Effects of Training on Physical Performance. *IEEE Transactions on Systems, Man, and Cybernetics*, *SMC-6*(2), 94–102. https://doi.org/10.1109/TSMC.1976.5409179

Candau, R., Busso, T., & Lacour, J. R. (1992). Effects of training on iron status in cross-country skiers. *European Journal of Applied Physiology and Occupational Physiology*, *64*(6), 497–502.

Fister, I., Ljubič, K., Suganthan, P. N., Perc, M., & Fister, I. (2015). Computational intelligence in sports: Challenges and opportunities within a new research domain. *Applied Mathematics and Computation*, *262*(Supplement C), 178–186. https://doi.org/10.1016/j.amc.2015.04.004

Fitz-Clarke, J. R., Morton, R. H., & Banister, E. W. (1991). Optimizing athletic performance by influence curves. *Journal of Applied Physiology (Bethesda, Md.: 1985)*, *71*(3), 1151–1158.

Hyndman, R. J., & Athanasopoulos, G. (2017). *Forecasting: principles and practice* (2nd ed.). Retrieved from https://www.otexts.org/fpp

Lewis, N. D. (2017). *Automated Time Series Forecasting Made Easy with R: An intuitive Step by Step Introduction for Data Science*. www.AusCov.com.

Morton, R. H., Fitz-Clarke, J. R., & Banister, E. W. (1990). Modeling human performance in running. *Journal of Applied Physiology (Bethesda, Md.: 1985)*, *69*(3), 1171–1177.

Pfeiffer, M. (2005). Modeling the Relationship between Training and Performance - A Comparison of Two Antagonistic Concepts. *International Journal of Computer Science in Sport*, *Volume 7*(2), 13–32.

Pfeiffer, M., & Hohmann, A. (2012). Applications of neural networks in training science. *Human Movement Science*, *31*(2), 344–359. https://doi.org/10.1016/j.humov.2010.11.004

Wallace, L. K., Slattery, K. M., & Coutts, A. J. (2009). The ecological validity and application of the session-RPE method for quantifying training loads in swimming. *Journal of Strength and Conditioning Research*, *23*(1), 33–38. https://doi.org/10.1519/JSC.0b013e3181874512