

**Analysis of Power Output and Training Stress in Cyclists:
The Development of the BikeScore™ Algorithm**

**Dr. Philip Friere Skiba
PhysFarm Training Systems LLC**

Acknowledgments:

The author wishes to express his gratitude to all of the amateur, elite, and professional athletes who have made their training and racing data available to him. Without these data, the development of new technology would be a much more difficult (if not impossible) enterprise.

Legal Notes:

BikeScore is a trademark of PhysFarm Training Systems LLC. Any commercial use of the algorithm requires a license from PhysFarm Training Systems, LLC. Non-commercial / academic use is permitted free of charge provided that such use is properly referenced.

TSS and NP were first developed by Dr. Andrew Coggan, and are trademarks claimed by PeaksWare LLC. The BikeScore system does not calculate NP or TSS, nor do any of the other products of PhysFarm Training Systems.

This work is Copyright 2008 by Dr. Philip Friere Skiba and PhysFarm Training Systems, LLC.

Introduction:

There exists a dose-response relationship between training stimulus and adaptation of the athlete (Bannister et al 1975, Busso 2003). Training load can be expressed simply as:

$$\text{Training load} = \text{Intensity} \cdot \text{Duration (Eq. 1)}$$

It is clear that different types of stimuli will effect different physiologic responses. It is less clear how to compare/quantify differing stimuli and their ability to affect the same response. A number of systems have been proposed, the most widely used of which is TRIMPS, which was devised by Dr. Eric Banister in the 1970's. Simply put, Banister sought to relate an easily measured parameter (heart rate) to lactate production through the use of a population study. This made a great deal of sense, perhaps even more so today as it is now widely accepted that the work rate at lactate threshold (defined as a rise of serum lactate of 1 mmol/L over exercise baseline) is the primary determinant of endurance exercise performance (Coyle 1988, 1999)

$$\text{TRIMPS} = \text{Duration} \cdot \text{Average HR during exercise} \cdot \text{A HR-dependant, intensity based weighting factor (Eq. 2)}$$

The benefit of Banister's system is that it takes into consideration the observation that higher workloads are more metabolically taxing (exponentially so, via the weighting factor) than lower workloads of equivalent duration (Bannister 1996). However, it is still dependent upon the measurement of heart rate, which is variable based on factors such as hydration, rest, illness, or cardiac drift. Furthermore, though HR is dependent upon workload, it may take minutes to stabilize when that workload changes. Because of these complicating factors, it would be preferable to measure work rate directly.

In 2003, Dr. Andrew Coggan refined Banister's concept by developing a system that also incorporated lactate response to workload. This system related the *change* in lactate concentration with the *change* in an *objective* measure of exercise intensity: power output, which can be directly measured by on-bike power meters.

Coggan devised a mathematical algorithm similar to that of Bannister, called the Training Stress Score (TSS).

$$\text{TSS} = \text{Exercise duration} \cdot \text{Average power} \cdot \text{Power-dependent, intensity weighting factor (Eq. 3)}$$

The power dependent intensity weighting factor was derived directly from a plot of blood lactate concentration as a percentage of concentration at threshold against % of threshold power. His work indicated a near 4th power relationship between the two.

The elegance of Coggan's system is that while it successfully relates lactate concentration to power output, it is not dependent upon invasive tests. In 1988, Coyle et al. illustrated

that the highest power output or pace an athlete can maintain over the course of an hour long exercise task is highly correlated with LT. Thus, to determine threshold intensity, the athlete need only perform such a test and use the resulting average power in the calculations. This 1-hour power has since been dubbed “functional threshold power” (Coggan 2003, 2006)

One potential hurdle in the analysis of power meter data is the often stochastic nature of the information. This is in stark contrast to HR data, for instance, which varies with a relatively predictable half-life. For instance, upon cessation of an effort, HR falls rather slowly over a period of 30 seconds to a minute. In contrast, when a cyclist stops pedaling, the power output immediately falls to zero. However, the physiologic response to the stress applied falls with a similar time course as the HR in the above example. This must be accounted for in any effort to calculate the physiologic strain imposed by the stress of a given exercise task. To solve this problem, Coggan used a 30 second moving average to smooth the power data, sensible given the many physiologic processes that have ~30 second half-lives (e.g. HR, plasma epinephrine concentration, ventilation, etc). However, these processes decay exponentially, rather than linearly. This is an area of potential / theoretical improvement which we examined.

Another potential issue lies in the definition of threshold power. Athletes are often unwilling/unable to undertake an hour-long test / time trial in order to obtain a satisfactory measurement. Additionally, the terminology “threshold” can be problematic, as athletes often have their own idea of what this means to them. Dr. Coggan has previously suggested the use of Monod and Scherrer’s Critical Power model in the absence of 40k TT / true 1 hour maximal power data. We examined the use of this paradigm in place of 40k TT data.

Finally, training metrics such as these have been demonstrated to be useful as the input functions for systems-based performance modeling and prediction equations. We examined differences in performance modeling / prediction based on the use of each metric.

Calculation of xPower and BikeScore:

Using the a protocol modified from that first described by Coggan (2003, 2006), we calculated our alternative stress metrics (xPower and BikeScore) via the following protocol.

1. Calculate Critical Power per the method of Monod (1960), using 3 minute and 20 minute exercise tests.
2. Analyze the data from a workout, computing a 25s exponentially weighted moving average for power.
3. Raise the values in step 2 to the 4th power.
4. Average for the values from step 3.
5. Take the 4th root of step 4. This is the xPower.
6. Divide xPower by Critical Power from step 1 to get the Relative Intensity (RI).
7. Multiply the xPower by the duration of the workout in seconds to obtain a “normalized work” value in joules.
8. Multiply value obtained in step 7 by the RI to get a raw BikeScore.
9. Divide the values from step 8 by the amount of work performed during an hour at Critical Power.
10. Multiply the number from step 9 by 100 to obtain the final BikeScore.

This calculation appears laborious at first glance, however, inexpensive software has been developed which automates the process (http://www.physfarm.com?page_id=12).

Comparison of the 30 sec Moving Average with a 25 sec Exponentially Weighted Moving Average to Address Variability:

As intimated above, a 30 second moving average (30s MA) does not account for the fact that physiologic response to stress decays with an exponential rather than linear time course. A logical / intuitive solution would be to address this with an algorithm that provided an appropriate decay pattern. Theoretically, an exponentially weighted moving average should provide a more realistic model. (This was in fact the initial basis for the development of the xPower and BikeScore algorithms.) We selected a 25 second exponentially weighted moving average (25s EWMA) based upon the fact that this should track consistently with oxygen kinetics.

Figure 1 demonstrates the difference between the smoothing methods for a 20k TT. Note that with some minor exceptions, the 25s EWMA (red) and 30s MA (blue) show a similar response pattern to the stress (power output, yellow). For this effort, AP was 198W, xPower was 236W, and NP as calculated by Coggan’s original methodology was 239W.

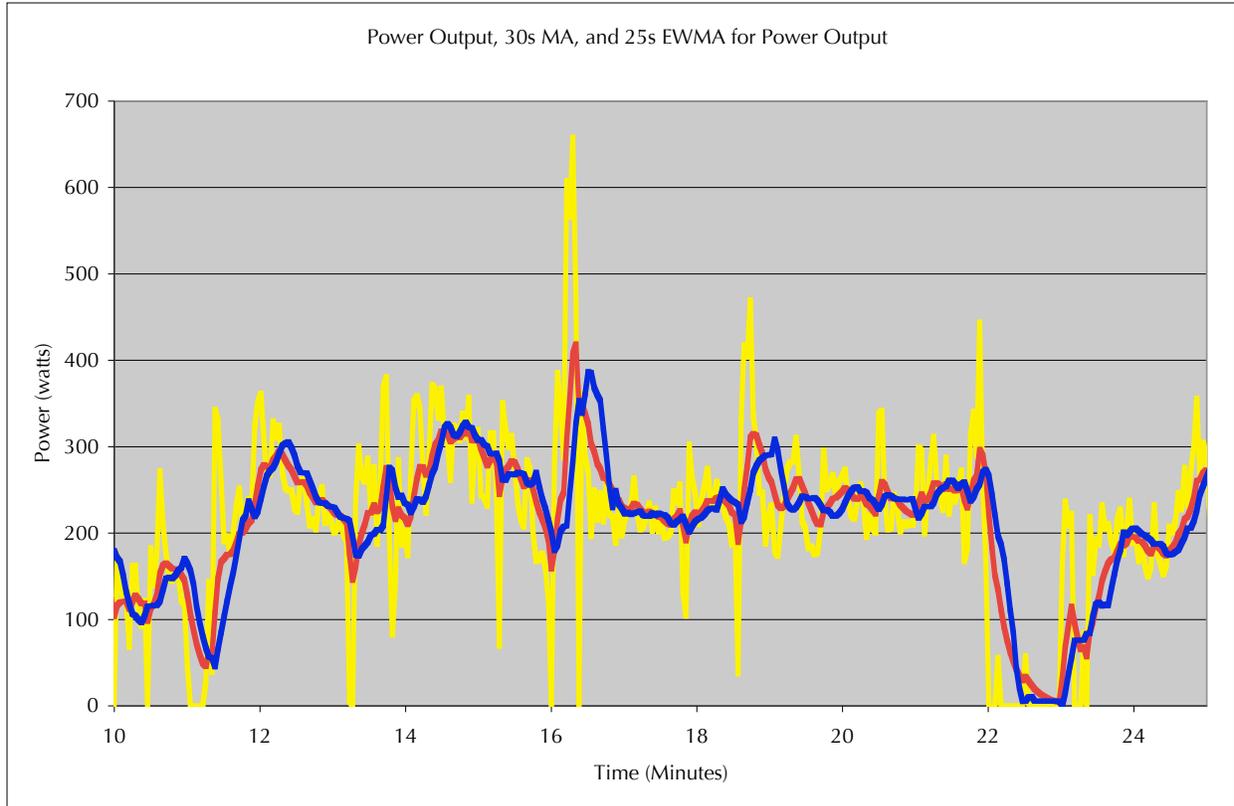


Figure 1: Instantaneous power output (yellow), 25s exponentially weighted moving average for power output (red), and 30s moving average for power output (blue) for a 20k TT. There is little apparent difference between smoothing methods.

When applied to a portion of an interval workout, similar results are observed. (Figure 2, AP=184W, xPower=200W, and NP=204). However, it is easier to observe the differences between smoothing methods. In this case, it would seem that while the 30s MA more closely tracks the stress (e.g. the raw power output), the 25s EWMA might be more representative of the *physiologic response or strain imposed by the athlete's effort*.

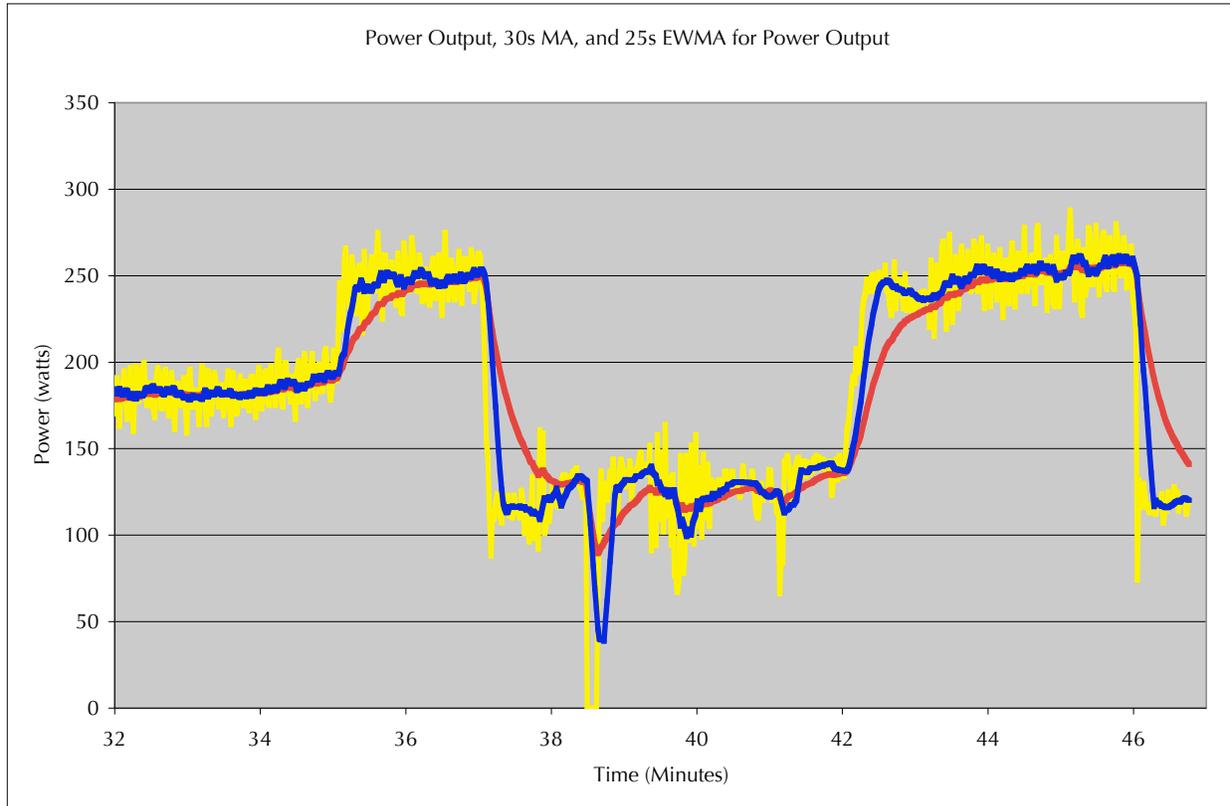


Figure 2: Instantaneous power output (yellow), 25s exponentially weighted moving average for power output (red), and 30s moving average for power output (blue) for a portion of an interval workout. Note the difference between smoothing methods. Note the difference in decay between the red and blue lines after each interval.

Addressing “Threshold”:

The aforementioned threshold power has proven to be problematic as, at least in our experience, athletes often neglect to test their threshold power with sufficient regularity. This seems to be largely due to the fact that an hour long test at maximal effort is exceedingly difficult and may require two or more days to completely recover from. Alternatives have been proposed, including 95% of the power maintained for an all out 20 minute test (Allen and Coggan, 2006). However, such arbitrary treatments of the data convey a false sense of precision and may lead to inappropriate training decisions.

A viable alternative comes in from of the Critical Power algorithm, first proposed in the 1960’s by Monod and Scherrer, which looks at a series of tests between a few minutes and 20 to 30 minutes. A plot is made of the number of joules generated at different time points, and a best fit line connecting them is calculated. The slope of this line is calculated (j/s or watts) which is dubbed the “Critical Power”; theoretically, a power the athlete can maintain indefinitely without fading.

This model almost certainly provides an overestimate in two senses: both the purely anaerobic/nonrenewable component of the effort (equal to the y-intercept of the line), and the fact that an exercise task at Critical Power could be maintained indefinitely. This said, it turns out that the Critical Power is quite close to 1 hour maximal / “threshold” power.

We have found two important factors in the development of BikeScore. First, a good result can be had by simply using the results of a short (3 minute) and a long (20 minute) test, provided that both the short and long test are undertaken at maximal effort. This method was found to have an excellent correlation to measured 40k power in a small group of elite / professional triathletes ($r^2 > 0.95$, $n = 5$, data not shown), and a group of amateur triathletes ($r^2 > 0.92$, $n = 10$, data not shown), with Critical Power seeming to overestimate 40k TT power in the latter more than the former. This leads us to believe that the difference between measured 1 hour maximal power and the calculated Critical Power are, as a practical matter, negligible. We have further found that the Critical Power paradigm provides a useful framework to evaluate athletes without being overly imposing / detrimental to their training schedule.

Differences in Modeling Ability:

We calculated an impulse-response model using commercially available software (RaceDay Performance Predictor, PhysFarm Training Systems LLC, Clark NJ), comparing Coggan’s TSS metric with BikeScore (Figure 3a and 3b).

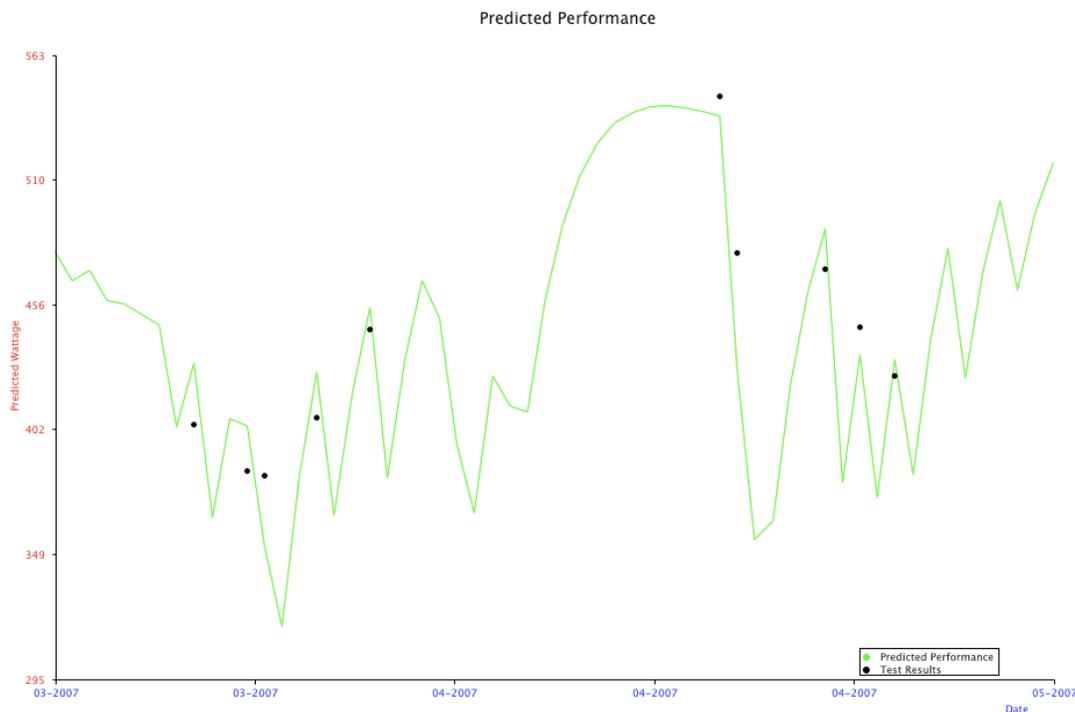


Figure 3a: Modeled (green) vs. predicted performance (black dots) using TSS as the input function. A high degree of correlation may be observed ($r^2 = .7551$).

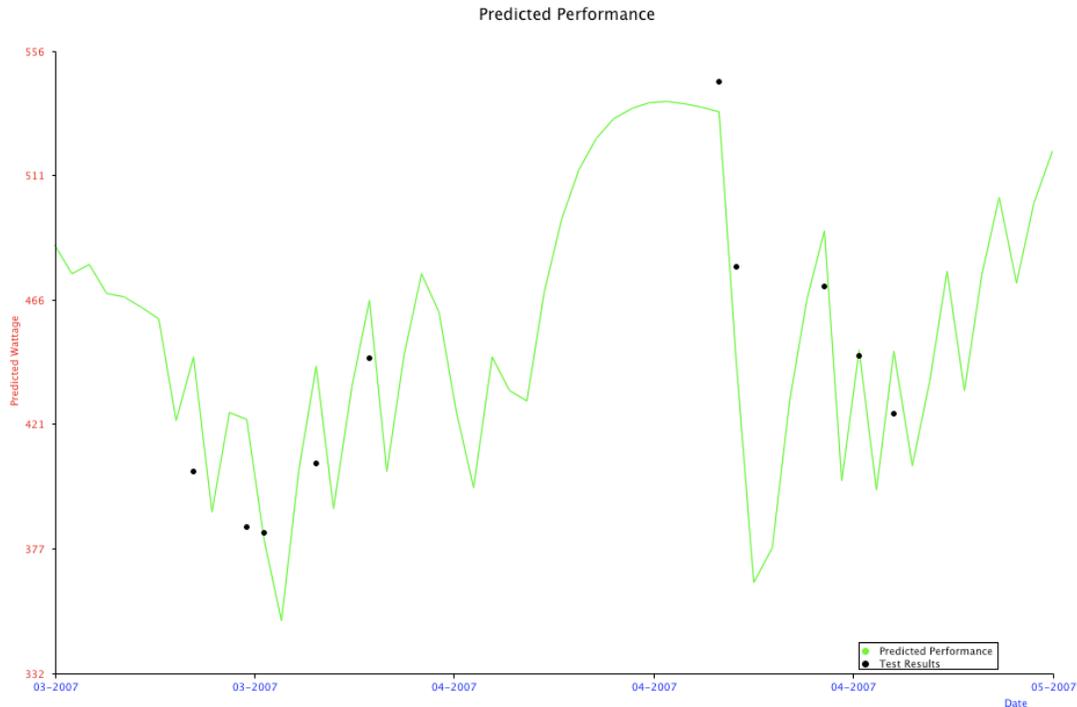


Figure 3b: Modeled (green) vs. predicted performance (black dots) using BikeScore as the input function. Though a high degree of correlation may be observed, there is no significant difference as compared to TSS.

As is evident from the graphical representations above, there is no significant difference in model fit using the different metrics and both models delivered the same parameters ($k_1 = 1$, $k_2 = 5$, $T_1 = 16$, $T_2 = 3$). This indicates either would likely deliver equal performance in modeling applications, and in an examination of a number of additional cases, we have yet to find a situation where one metric appears to be a substantial improvement over the other. (This is not necessarily surprising given the quality of results possible using a metric as crude as TRIMPS, which *also* yields similar model performance).

Conclusions:

Although it remains to be seen whether the above changes represent a substantial (indeed any) improvement over Dr. Coggan's rather robust solution, we suspect that more formal studies will demonstrate no significant difference between these two different solutions. In point of fact, the author was one of the first people to validate Coggan's metric for cycling and a novel variation that permitted similar calculations for running (Skiba, 2006 & 2007a), and to advance a similar protocol for the calculation of the stress of swimming and cross country skiing (Skiba 2007b & 2007c). However, BikeScore was developed and released as a possible incremental improvement (which would also be be freely licensed for noncommercial / academic / open-source use), and for the moment seems to be at least as useful as the work which originally inspired it. We hope this will encourage further work in the field and the investigation of methods to improve these sorts of tools.

References:

- Banister EW, Calvert TW, Savage MV. A systems model of training for athletic performance. *Aust. J. Sports Med* 1975; 7:57-61.
- Banister EW. Modeling elite athletic performance. In: MacDougall JD, Wenger HA, Green HJ, eds. *Physiological Testing of the High-Performance Athlete*. Champaign, IL: Human Kinetics; 1996: 403-424.
- Busso T. Variable dose-response relationship between exercise training and performance. *Med Sci Sports Exerc.* 2003; 35(7):1188-1195.
- Coyle EF, Coggan AR, Hopper MK. Determinants of endurance in well-trained cyclists. *J Appl Physiol.* 1988; 64 (6): 2622-30.
- Coyle EF. Physiological determinants of endurance performance. *Journal of Science and Medicine in Sport.* 1999; 2(3): 181-189, 1999.
- Coggan, Andrew R. Training and racing using a power meter: an introduction. *USA Cycling*, 2003. Available at <http://www.midweekclub.com/articles/coggan.pdf>. Accessed September 20, 2005.
- Coggan, Andrew R. Making sense out of apparent chaos: Analyzing on the bike power data. In: *The Science of Cycling: Transforming research into practical applications for athletes and coaches*. Highlighted symposium, American College of Sports Medicine 53rd Annual Meeting. May 31, 2006.
- Allen, H and Coggan AR. *Training and racing with a power meter*. Boulder, CO: VeloPress; 2005.
- Skiba, Philip Friere. Quantification of Training Stress in Distance Runners. *Arch Phys Med Rehabil* 87:29, 2006.
- Skiba, Philip Friere. Evaluation of a Novel Training Metric in Trained Cyclists. *Med Sci Sports Exerc* 39:5, Supplement, 2007.
- Skiba, Philip Friere. Evaluation of a Novel Training Metric in a Trained Triathlete. *Clin J Sports Med* (In Press). AOASM Conference Lecture, April 2007.
- Skiba, Philip Friere. Calculation of Optimal Taper Characteristics in an Amateur Triathlete. In review. 2008.