

# Módulo 2. Creación de métricas desde raw data



Acá va la introducción a la lectura (justificar texto).

☰ 1. ¿Cómo se han calculado los datos que obtenemos?

☰ Referencias

# 1. ¿Cómo se han calculado los datos que obtenemos?

---

Queremos valorar la velocidad de los jugadores del equipo en el que trabajamos. En primer lugar, debemos preguntarnos qué aspecto dentro de la velocidad queremos valorar. Aunque pueda parecer una evaluación simple, el test que utilicemos condicionará la naturaleza de la velocidad que queramos evaluar. Veamos dos ejemplos. Podemos optar por utilizar «*timing gates*». Estos dispositivos permiten cronometrar el tiempo que tarda un jugador desde el inicio del *sprint* hasta que cruza la línea final. Esta tecnología habitualmente consta de dispositivos con un láser dirigido a un reflector delante de ellos. Estas dos piezas forman un par o comúnmente llamado puerta (láser y reflector) y con la correspondiente separación entre ellas, se colocan a la misma altura; el primer par en el inicio del *sprint* y el segundo al final. El jugador, habitualmente colocado ligeramente detrás del primer par de dispositivos, al iniciar el movimiento y cruzar por la primera puerta, interrumpirá la señal que está dando el dispositivo contra el reflector, lo que activará un cronómetro, el cual se detendrá en el momento en que el jugador pase por la segunda puerta. El resultado que obtendremos de esta prueba es un valor numérico, los

segundos que el jugador tardó en recorrer la distancia entre las dos puertas.

Si realizamos esta prueba con todo el equipo, podremos ver qué jugadores son más «rápidos» recorriendo esa distancia de, por ejemplo, 40 metros, es decir, que han tardado menos tiempo en recorrer la distancia marcada. Sin embargo, ¿que un jugador haya tenido el menor tiempo en la prueba, indicará que este jugador tiene la mayor velocidad máxima o punta del equipo? La respuesta es no; un jugador puede haber alcanzado una velocidad punta a los 20 metros, pero no ser capaz de mantenerla los 20 metros restantes. Por otra parte, un jugador que sea capaz de alcanzar una velocidad punta ligeramente inferior y la mantenga durante más distancia será capaz de recorrer la distancia total en menos tiempo.

Si nuestro objetivo es valorar la velocidad punta de nuestros jugadores, el *test* utilizado deberá ser diferente. Una opción puede ser utilizar dispositivos GPS. Estos dispositivos permiten medir la velocidad del jugador en cada posición registrada, ya que calculan el cambio de posición a partir de la tecnología GPS. Dependiendo del tipo y la marca comercial del dispositivo, la posición se mide a distintas frecuencias. El término de frecuencia en estos dispositivos se refiere al número de capturas por segundo. Es decir, si tenemos un dispositivo que registra a 1 Hz, indicará que tenemos un registro de la posición del jugador cada segundo; por otro lado, si medimos con una frecuencia de 10 Hz, tendremos 10 puntos registrados cada segundo. Utilizando cada uno

de estos puntos registrados, podemos ver en cuál de ellos el jugador ha conseguido una mayor velocidad y de esta manera determinar su velocidad máxima. Además, teniendo información sobre la posición y velocidad del jugador, también podríamos obtener tiempos para recorrer una distancia determinada, lo que conseguiría el mismo tipo de análisis que las «*timing gates*».

El objetivo de este ejemplo no es decidir qué tipo de test es mejor, sino destacar cuáles son las diferencias entre las medidas que recogemos. En el primer caso, obtenemos un único valor, calculado mediante un procesamiento simple como un cronómetro, mientras que, para el segundo caso, registramos todo el *sprint* para obtener información de uno de los puntos registrados. Es importante en muchas ocasiones disponer de las medidas más sencillas posibles, pero debemos conocer el potencial que tienen los dispositivos con alta frecuencia de registro.

¿Cuáles de las siguientes afirmaciones sobre los "timing gates" son correctas?

---

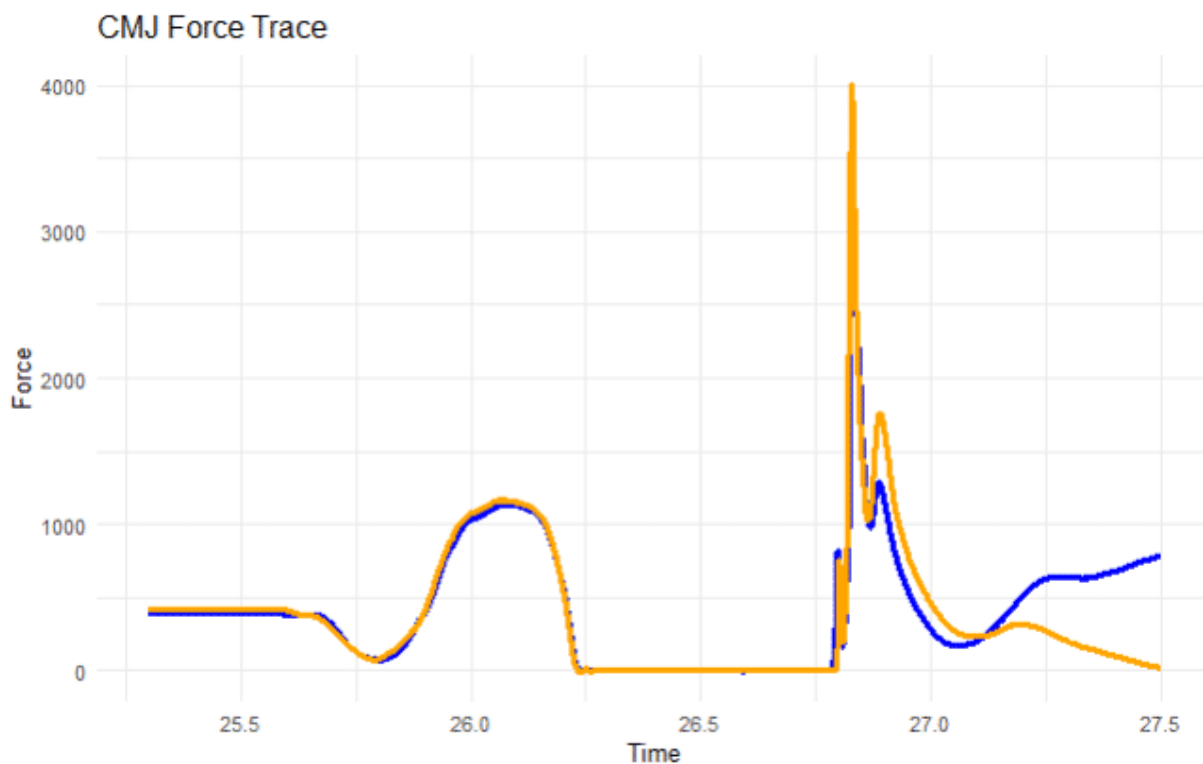
- A) Los "timing gates" constan de un láser y un reflector que forman una puerta.
- B) El cronómetro se activa cuando el jugador pasa por la segunda puerta.

- C) El cronómetro se detiene cuando el jugador pasa por la primera puerta.
- D) El jugador interrumpe la señal al pasar por la primera puerta, activando así el cronómetro.

SUBMIT

Como usuarios de los datos que nos proveen las distintas compañías de tecnología para medir variables de rendimiento físico, estamos acostumbrados a recibir y utilizar valores «calculados» de los datos recogidos a alta frecuencia. Es decir, si realizamos un *test* de salto, medido con una plataforma de fuerzas que mide a 1000 Hz, los datos que realmente están siendo recogidos por la plataforma tienen un aspecto similar al siguiente:

**Figura 1. Ejemplo de cálculo de datos – *test* de salto**



Fuente: elaboración propia

Es difícil extraer conclusiones a partir de este gráfico a simple vista; podríamos ver picos de fuerza en las distintas partes, pero obtener la información de esta manera es de poca utilidad. Es por este motivo que los distintos *software*, mediante el procesamiento de esta señal, son capaces de distinguir las distintas fases del movimiento y nos proveen información con variables ya calculadas, por ejemplo, el pico de fuerza de cada una de las fases, la duración, el impulso durante la parte concéntrica o el «*rate of force development*».

Lo mismo ocurre con los dispositivos GPS o acelerómetros; el *output* que recibimos habitualmente de las distintas compañías tecnológicas

consta de variables calculadas que posteriormente utilizaremos en el análisis. Es fundamental conocer de dónde provienen estos datos y cómo han sido calculados para determinar su utilidad en nuestro ámbito.

Los datos sin procesar, tal y como los ha registrado el dispositivo, también reciben el nombre de «*raw data*». La utilización de estos datos tiene un gran potencial para desarrollar análisis más complejos y, sobre todo, a medida.

Este es un ejemplo de las diferencias que hay entre el tipo de archivos que obtenemos después de la descarga del raw data (A) y los resultados calculados del *software* (B) para un mismo *test* (salto CMJ, 1 repetición). En el archivo *raw data* vemos pocas variables/columnas, pero muchas observaciones (más de 2100). En el caso de las variables calculadas, vemos solo una observación (1 salto) pero con tantas variables como queramos utilizar de las que dispone el software que utilizamos.

**Figura 2. Ejemplo de cálculo de datos – *test* de salto**

**A**

Time	Left	Right
0	-1.367493	-0.498093
0.001	-1.367493	-0.498093
0.002	0.632507	0.501907
0.003	-1.367493	-0.498093
0.004	0.632507	-0.498093
0.005	-1.367493	-0.498093
0.006	-0.367493	-0.498093
0.007	-0.367493	0.501907
0.008	-0.367493	-0.498093
0.009	-1.367493	-0.498093
0.01	0.632507	-0.498093
0.011	-0.367493	0.501907
0.012	0.632507	-0.498093
...	...	...
+2100 filas	+2100 filas	+2100 filas

**B**

Athlete	Test Type	Test Date	Body Weight [kg]	Trial	Concentric Mean Power / BM [W/kg]	Eccentric Mean Deceleration Force [N]	Jump Height (Flight Time) [cm]
Marco Marquez	CMJ	2/4/2022 00:00	80.94	Trial 1	36.6	1689.5	40.4

Fuente: elaboración propia

### ¿Por qué utilizar raw data?

Aunque la eficiencia y practicidad de utilizar datos calculados o procesados es clara en un contexto de análisis de datos diario, en muchas ocasiones nos encontramos con limitaciones por parte de los *software* de la tecnología que utilizamos. Nuevas investigaciones aparecen a menudo donde se presentan nuevas variables que pueden tener utilidad en nuestro contexto, métricas que tienen otras marcas, pero no la que nosotros estamos utilizando. El proceso de incorporar nuevas métricas por parte de una marca tecnológica puede ser largo y si estamos interesados en reproducir ciertos análisis que consideramos de impacto en nuestro equipo, debemos buscar la solución para realizarlos nosotros mismos.

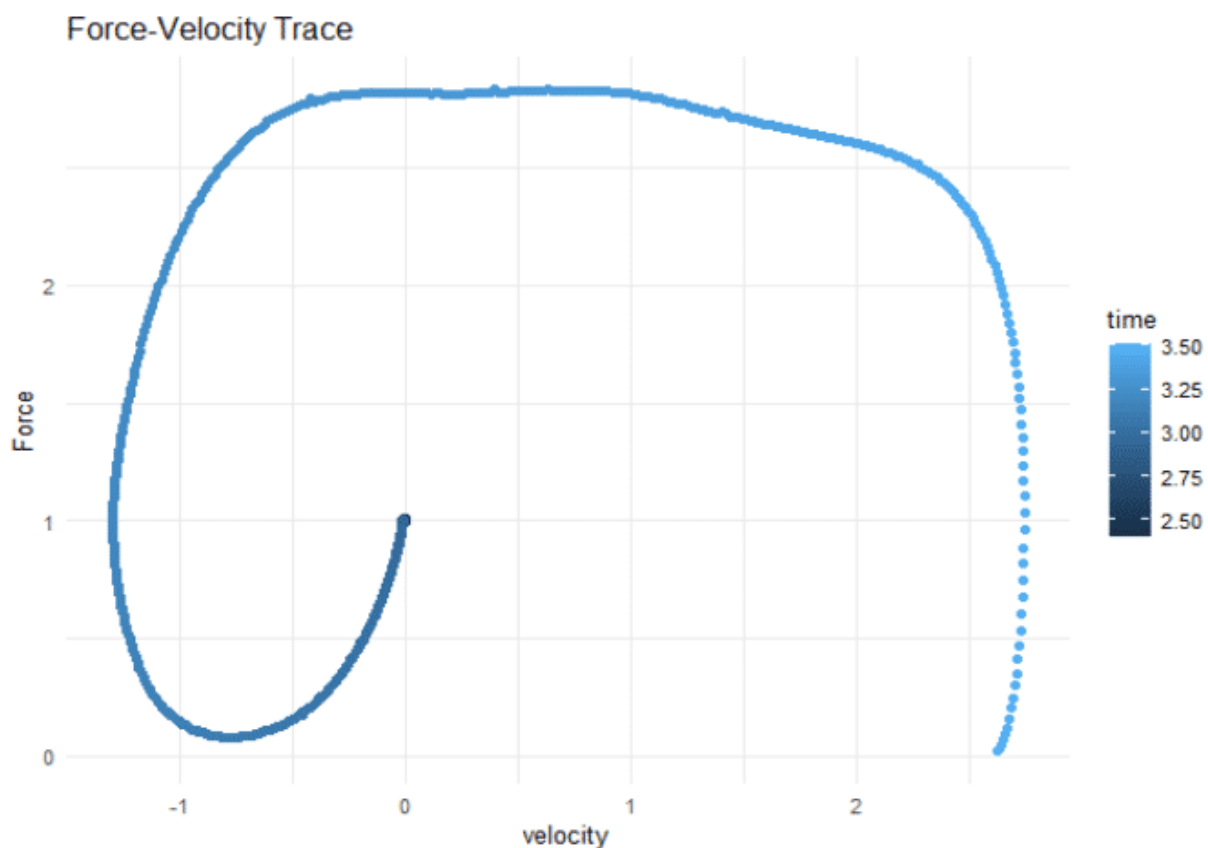
El *raw data* y el conocimiento del procesamiento y análisis del mismo facilitarán la consecución de dicho objetivo, nos darán mayor flexibilidad en el análisis que queramos realizar y será a medida.

## Ejemplos de análisis con raw data —

En el material de video veremos cómo trabajar con el *raw data*, desde las características específicas de la importación de los datos hasta métodos para ser eficientes en su análisis. Los ejemplos que veremos a continuación son muestras de cómo utilizar el *raw data* para replicar investigaciones y profundizar en el análisis sin tener las limitaciones de los distintos *software* tecnológicos.

El CMJ ha sido un *test* ampliamente utilizado para valorar los cambios en el rendimiento, la fatiga y los procesos de readaptación de los jugadores (Bishop et al., 2023). Habitualmente se utilizan variables calculadas (valores como el RSI, «*eccentric duration*», RFD) para valorar los cambios, pero Gathercole et al. (2015) proponen un método alternativo para valorar la fatiga. Utiliza el *raw data* de los distintos saltos para calcular una serie de variables derivadas a partir de los valores que proporciona la plataforma de fuerza y visualiza los valores calculados y cómo van cambiando a lo largo del salto. De esta manera, en lugar de utilizar un único valor, podemos ver cómo el jugador cambia su estrategia de salto según el nivel de fatiga. Si estuviéramos interesados en utilizar este tipo de análisis, deberíamos verificar si el *software* que utilizamos proporciona este tipo de visualización. Si no fuera el caso, podemos descargar el *raw data*, realizar los cálculos pertinentes y visualizarlo para su análisis posterior.

**Figura 3. Ejemplo de uso de raw data**



Fuente: elaboración propia

---

Otro ejemplo de la flexibilidad en el análisis del raw data lo podemos encontrar en la utilización de datos GPS. Un dispositivo que registre información durante toda la sesión de entrenamiento o el partido es una gran fuente de información. Hemos destacado en módulos anteriores la utilidad del registro del desempeño físico utilizando estos dispositivos y existe extensa literatura que justifica su uso para la valoración del rendimiento en distintos deportes (Caro et al., 2022).

Verdadero o Falso: La incorporación de nuevas métricas por parte de una marca tecnológica suele ser rápida, lo que facilita la realización de análisis de datos de impacto en nuestro equipo.

---

Verdadero

Falso

SUBMIT

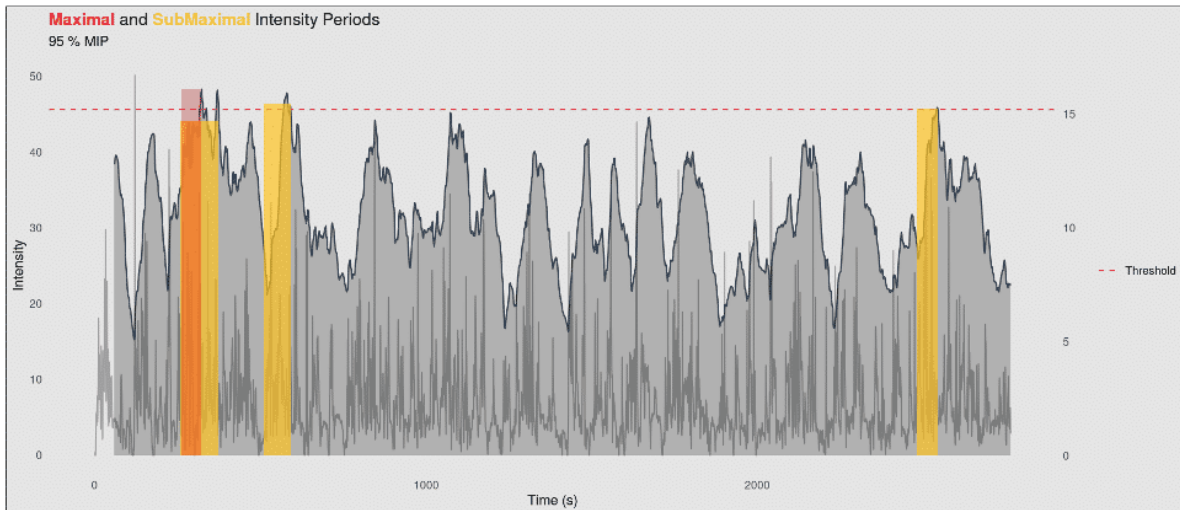
En los últimos años, se han desarrollado distintos análisis para valorar las fases en las que el jugador experimenta mayores exigencias condicionales durante los partidos (García et al., 2020). Estas fases también son conocidas como períodos de máxima intensidad. Conocer la intensidad de estas fases es de gran importancia para poder preparar a los jugadores durante el entrenamiento, ya que se ha demostrado que la intensidad media del partido subestima las demandas específicas de cada posición. A partir de este enfoque de la valoración del rendimiento se han desarrollado análisis complementarios, como es el caso de los «*submaximum intensity*

*periods»* (SubMIP) (Caro et al., 2022). Este enfoque quiere demostrar que estas fases no ocurren de manera aislada, sino que se producen de manera repetida a una intensidad similar a esos valores máximos. Además, son específicas para cada posición, lo que justifica su uso en el control de la carga de entrenamiento.

De nuevo, este tipo de investigaciones son relativamente recientes, por lo que muchos de los *software* de análisis de datos GPS no permiten el cálculo de estas variables. Esta limitación puede ser corregida si usamos el raw data de los dispositivos y replicamos los métodos que describen los estudios.

En el siguiente gráfico vemos una visualización del análisis de los períodos submáximos de un partido de fútbol. Encontraremos este análisis en el material de video, pero cabe destacar que el uso del raw data no se trata exclusivamente de visualizar una señal o métrica (como en el ejemplo anterior), sino que a partir del análisis y sus cálculos podemos obtener valores resumen. En la imagen podemos ver el número de períodos submáximos, su intensidad y su duración. Estos valores nos pueden servir de referencia para diseñar tareas que repliquen las demandas de la competición.

**Figura 4. Análisis de los períodos submáximos de un partido de fútbol**



id	event_duration	intensity	start_time	end_time
1	117.50	14.71	256.70	374.20
2	81.30	15.47	510.40	591.70
3	63.40	15.25	2480.50	2543.90

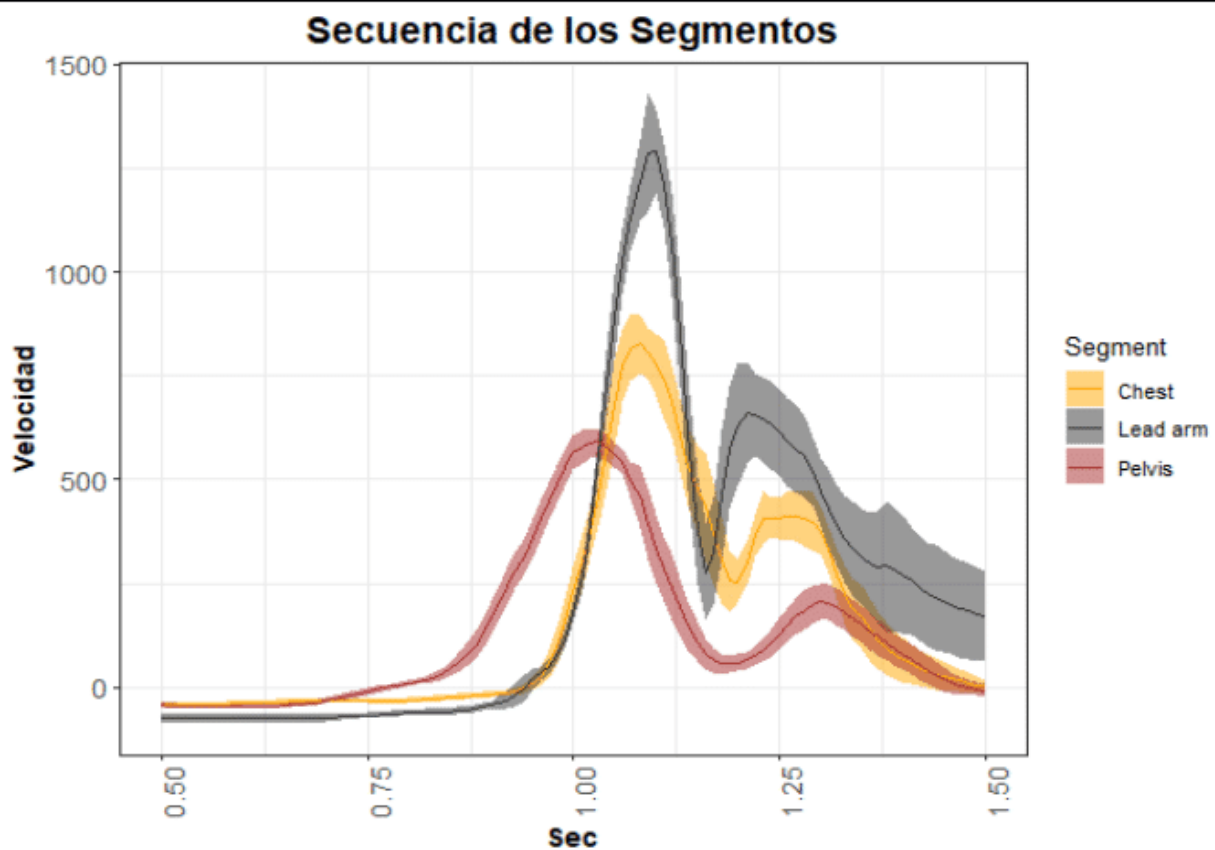
Fuente: elaboración propia

Finalmente, veamos un ejemplo donde también visualizaremos la información proporcionada por el dispositivo, pero en este caso utilizando más de un dispositivo a la vez. Se trata de acelerómetros colocados en distintas partes del cuerpo (cadera, escápulas y brazo) durante un movimiento deportivo. En este caso, queremos evaluar cómo los distintos segmentos del cuerpo se mueven en relación con el resto durante un revés de un jugador de tenis. Podríamos utilizar los resultados resumen/calculados que nos proporciona el dispositivo, por ejemplo, la velocidad angular máxima de cada uno de los dispositivos, pero no estaríamos respondiendo la pregunta planteada. Queremos evaluar la secuencia de los distintos segmentos, es decir, qué parte del cuerpo consigue una velocidad máxima antes, qué parte le sigue y qué parte es la última en alcanzar su pico de velocidad. Visualizando

la señal del *raw data* a lo largo del tiempo registrado del movimiento de los tres acelerómetros, podemos obtener el siguiente gráfico. Este tipo de análisis cinemáticos son muy comunes en deportes de mayor complejidad técnica.

Vemos cómo el pico de la cadera se sitúa más a la izquierda que el resto (antes en el tiempo), luego consigue su pico el torso y finalmente el brazo. Por lo tanto, nos acerca a responder la pregunta. Las distintas alturas de los picos también nos dan información complementaria que puede ser calculada a partir del *raw data*.

**Figura 5. Ejemplo de información obtenida de más de un dispositivo**



Estos tres ejemplos únicamente sirven para mostrar tipos de análisis que podemos realizar. El punto más importante a tener en cuenta es el objetivo de nuestros análisis, es decir, cuál es la pregunta que queremos responder. El *raw data* será una herramienta más de la que podremos disponer para acercarnos a la solución del problema que se nos plantee.

## Consideraciones en la utilización del *raw data*

### VOLUMEN DE DATOS

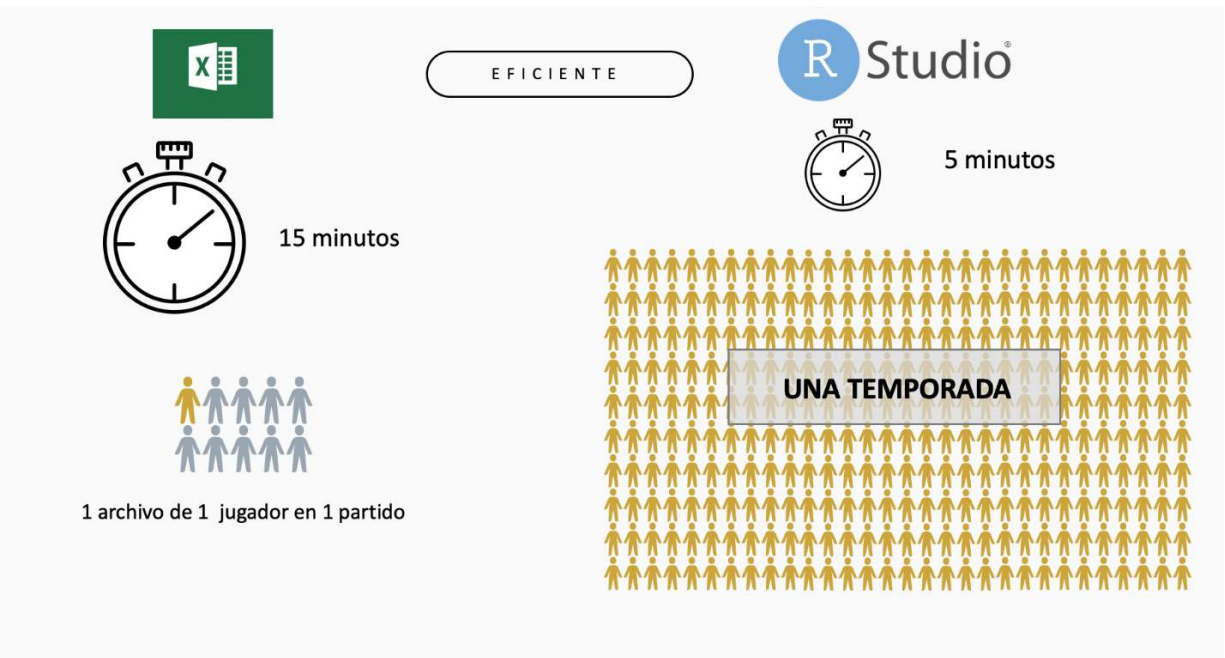
En primer lugar, hay que destacar el alto volumen de datos con los que trabajamos cuando utilizamos *raw data*. Hemos visto la comparativa en tablas de los test de saltos, la diferencia entre el número de observaciones de un salto para un jugador. Si se trata de un análisis que queremos realizar de forma diaria para todos los jugadores de nuestro equipo, el volumen de datos crece exponencialmente. Además, cuantas más variables queramos calcular o modificaciones al *raw data*, la complejidad de los datos también será mayor.

Por este motivo, es necesario utilizar herramientas que permitan ser eficientes en el manejo de estas estructuras de datos. RStudio se sitúa de nuevo como una opción ideal para desarrollar los análisis que queramos con el *raw data*.

Para visualizar las capacidades de RStudio comparado con herramientas más habituales, vamos a representar una comparativa del mismo proceso de análisis, realizado mediante Excel y RStudio para ver las diferencias. En esta comparativa se pretende realizar cálculos de los SubMIP (ejemplo que vimos anteriormente) para un equipo de fútbol. Los mismos cálculos se realizaron con Excel y RStudio, los cuales consistían en:

- importar los datos a Excel o RStudio;
- aplicar filtrado en la señal de velocidad y aceleración (lo veremos a continuación);
- calcular variables complementarias;
- aplicar los promedios móviles para calcular los SubMIP;
- identificar los SubMIP;
- calcular las características de cada período;
- resumir las características de la sesión.

**Figura 6. RStudio vs. Excel**



Fuente: elaboración propia

En el esquema podemos ver claramente la eficiencia de ambas herramientas. Mientras que en Excel necesitamos 15 minutos para completar el análisis de una sola sesión por un jugador, debido a la necesidad de utilizar múltiples libros de Excel para soportar la demanda computacional de las operaciones, con RStudio podemos analizar una temporada entera en 5 minutos para todos los jugadores del equipo. Queda clara la eficiencia de RStudio para desarrollar tareas con este tipo de datos.

¿Cuáles de las siguientes tareas se realizaron tanto con Excel como con RStudio en el análisis comparativo de los SubMIP para un equipo de fútbol?

- A) Importar los datos a Excel o RStudio.
- B) Aplicar filtrado en la señal de velocidad y aceleración.
- C) Generar gráficos tridimensionales de las trayectorias de los jugadores.
- D) Calcular variables complementarias.
- E) Aplicar los promedios móviles para calcular los SubMIP.

SUBMIT

## Filtrado de datos

Este segundo aspecto es más complejo, pero de gran importancia cuando trabajamos con raw data: se trata del filtrado de los datos. Cuando trabajamos con dispositivos que registran a alta frecuencia, a menudo la señal que obtenemos contiene el llamado «ruido». Este ruido se refiere a pequeñas variaciones en el dato registrado que

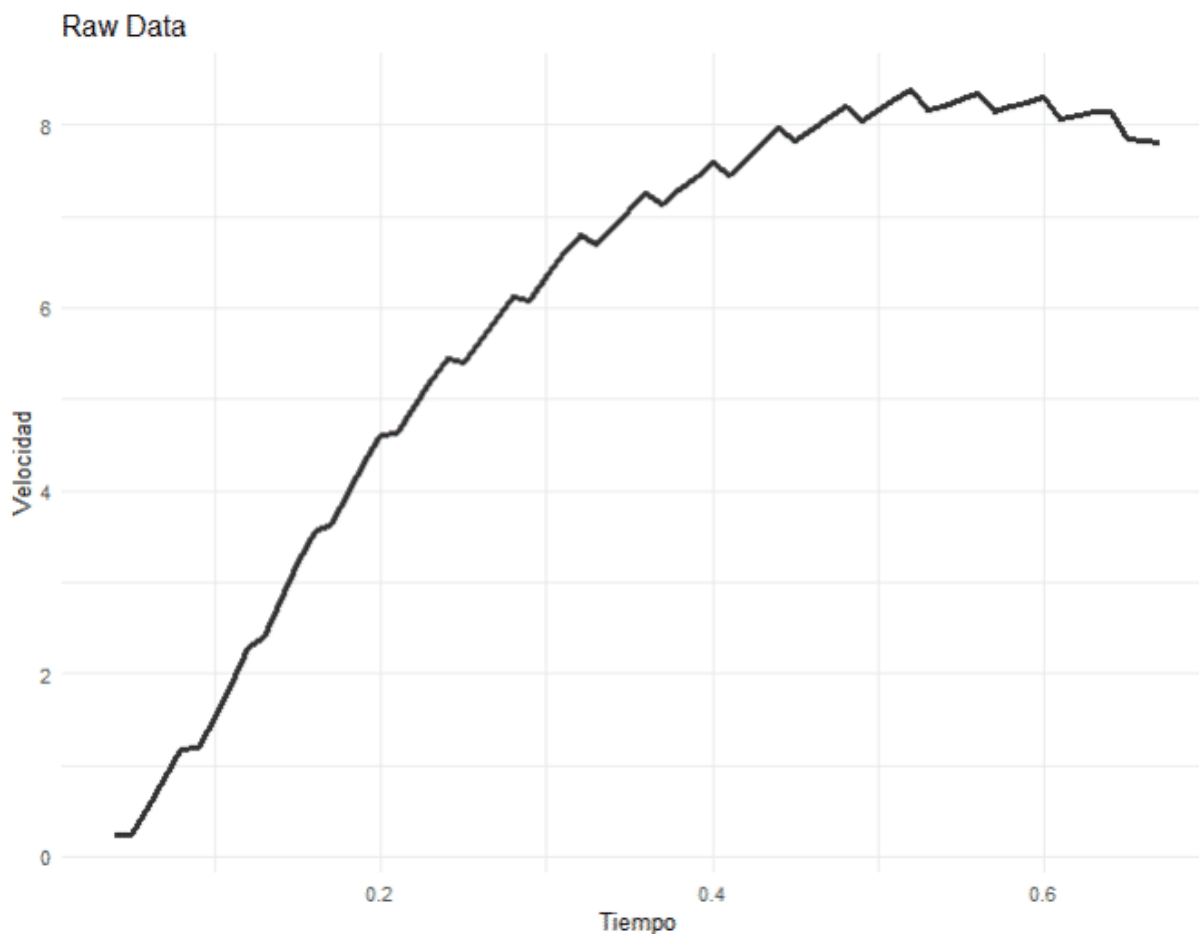
interfieren en la calidad de la información. Este ruido puede afectar en los pasos posteriores del procesamiento y análisis de datos, y es común e inherente al tipo de captura de señal.

Para corregir ese ruido y obtener una señal que sea válida para utilizar, se utilizan métodos de filtrado de la señal que permiten reducir el ruido sin perder la información relevante. Se trata de un concepto complejo y podemos acudir a la literatura aplicada a nuestro contexto para profundizar sobre este contenido (Hoppe et al., 2018), pero de manera simplificada podríamos referirnos a la aplicación de filtrado de una variable como aquella que permite reducir el ruido (eliminar picos indeseados o variaciones excesivas en los datos) sin perder información relevante (mantener la forma o tendencia de los datos a lo largo de la serie temporal).

En la siguiente imagen podemos ver la señal de velocidad durante un *sprint* registrada por un radar. Se puede observar cómo la señal contiene cierto ruido (pequeños incrementos y decrementos a lo largo del tiempo). Si conocemos el tipo de movimiento que estamos registrando, un *sprint* lineal, podemos asumir que el jugador está desplazándose mientras incrementa la velocidad (aceleración positiva constante). Si vemos la señal donde hay incrementos y decrementos de la velocidad, estaría indicando que el jugador desacelera y acelera

constantemente, lo que no estaría representando el movimiento que queremos registrar.

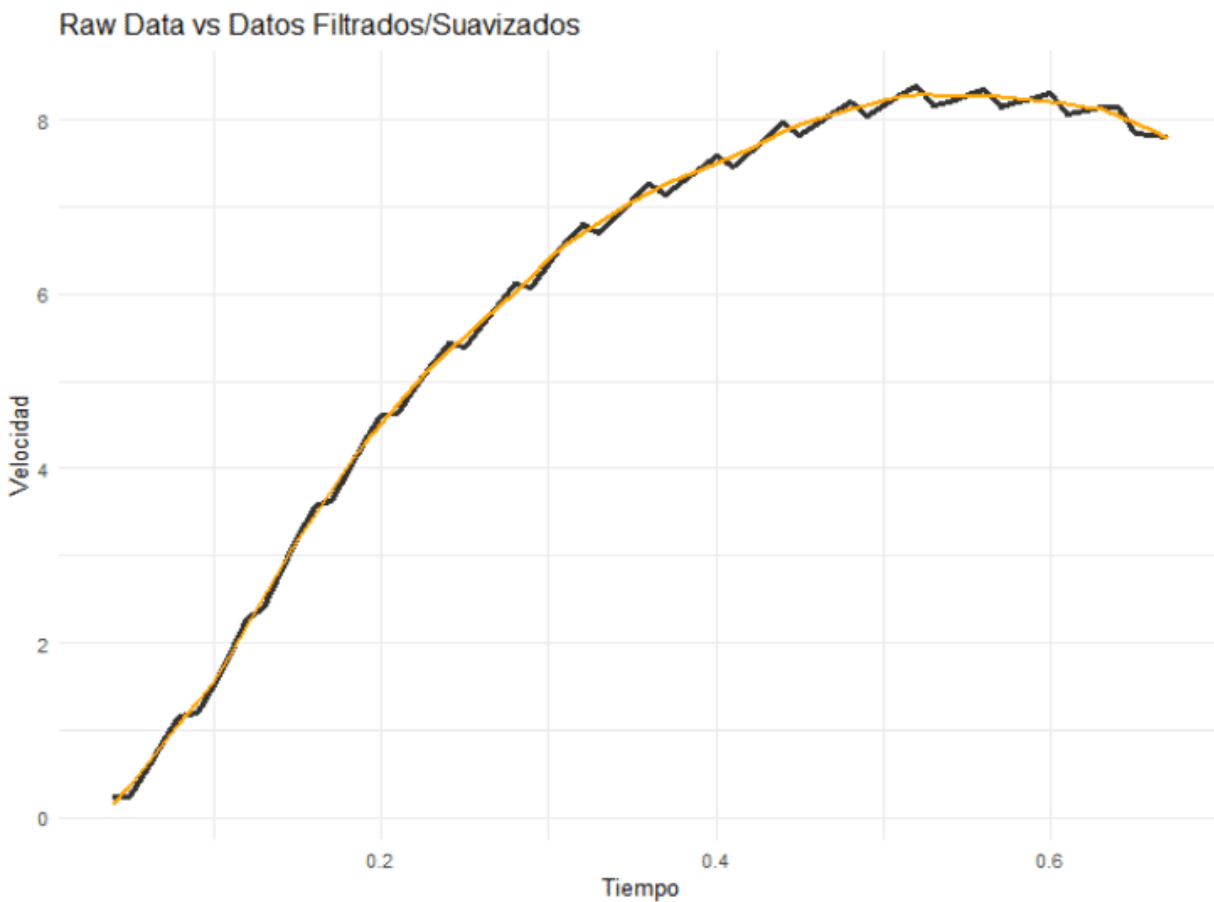
**Figura 7. Señal de velocidad durante un *sprint* registrada por un radar**



Fuente: elaboración propia

Con las consideraciones anteriores, se decide aplicar un filtrado a la señal de velocidad para reducir el ruido, manteniendo la mayor cantidad de información posible.

**Figura 8. Señal de velocidad durante un *sprint* registrada por un radar – aplicación de filtrado**



Fuente: elaboración propia

En la imagen anterior, podemos ver la señal filtrada, en ocasiones también llamada suavizada (línea naranja), comparada con el *raw data* (línea negra). El filtrado de los datos cumple el objetivo marcado: reduce el ruido innecesario del registro manteniendo la información relevante, tendencia que sigue la línea.

Como hemos comentado anteriormente, se trata de un procedimiento bastante complejo, ya que existen múltiples tipos de filtros y parámetros para regular el resultado que queremos. Si deseamos replicar estudios en la literatura o utilizar el *raw data* de alguna de las tecnologías que utilicemos (GPS, plataformas de fuerzas, acelerómetros), es fundamental acudir a la documentación o la literatura para conocer qué tipos de filtros utilizaron, ya que, de lo contrario, los resultados que podemos obtener serían distintos.

A modo de resumen podemos destacar cuáles son las posibilidades y ventajas de utilizar raw data en nuestros análisis:

- Detalle: ser más precisos en la evaluación de los resultados que obtenemos, identificar distintas fases con mayor precisión y conocer de qué manera se ha producido el resultado final.
- Adaptable/customizable: permite elaborar análisis con los parámetros que nosotros indiquemos, ofreciendo mayor versatilidad y potencial a las posibilidades del análisis.
- Transparencia: en muchas ocasiones, los métodos para procesar los datos no están descritos en la literatura, por lo que estamos limitados a conocer de dónde provienen los resultados. Utilizando análisis

proprios con el *raw data* nos permite ser más claros en el proceso que estamos utilizando y detectar posibles aspectos de mejora.

- Repetibilidad: el punto anterior ayuda a que podamos utilizar los mismos métodos en el futuro, realizar comparaciones más precisas, ya que también es común que las distintas compañías modifiquen cambios en el procesamiento de los datos a lo largo del tiempo, lo que impide que las comparaciones con datos más recientes sean fiables.

CONTINUAR

## Referencias

---

**Bishop, C., Jordan, M., Torres-Ronda, L., Loturco, I., Harry, J., Virgile, A., Mundy, P., Turner, A., & Comfort, P.** (2023). Selecting metrics that matter: comparing the use of the countermovement jump for performance profiling, neuromuscular fatigue monitoring and injury rehabilitation testing. *Strength and Conditioning Journal*, 45(5), 545-553.

**Caro, E., Campos-Vázquez, M. Á., Lapuente-Sagarra, M., & Caparrós, T.** (2022). *Analysis of professional soccer players in competitive match play based on submaximum intensity periods*. PeerJ, 10, e13309.

**García, A., Castellano, J., Mendez-Villanueva, A., Gómez-Díaz, A., Cos, F., & Casamichana, D.** (2020). *Physical Demands of Ball Possession Games in Relation to the Most Demanding Passages of a Competitive Match*. *Journal of Sports Science & Medicine*, 19, 1-9.

**Gathercole, R. J., Stellingwerff, T., & Sporer, B. C.** (2015). *Effect of Acute Fatigue and Training Adaptation on Countermovement Jump Performance in Elite Snowboard Cross Athletes*. *Journal of Strength and Conditioning Research*, 29(1), 37-46.

**Hoppe, M. W., Baumgart, C., & Freiwald, J.** (2018). *Estimating external loads and internal demands by positioning systems and innovative data processing approaches during intermittent running activities in team and racquet sports. Sport-Orthopädie - Sport-Traumatologie - Sports Orthopaedics and Traumatology, 34, 3-14.*

CONTINUAR