

El problema oculto de la corrección: cuando el tratamiento de datos crea confianza en lugar de información

Juan Segurola

3Dresyns by Resyner Technologies S.L.

© 2026 Juan Segurola. Todos los derechos reservados.

Correspondencia: info@3Dresyns.com · www.3Dresyns.com

HEB-009 · Artículo técnico independiente · Versión v0.3

DOI: 10.5281/zenodo.22915218

Resumen

El tratamiento de datos es necesario en muchos experimentos, pero toda corrección contiene un modelo de que es el componente no deseado y como entra en la medición. La sustracción de línea base, la normalización, el suavizado, la corrección por blanco, la eliminación de fondo, la deconvolución y la eliminación de outliers pueden reducir artefactos conocidos cuando sus supuestos son válidos. También pueden suprimir estructura real, amplificar ruido, fabricar precisión aparente u ocultar que el experimento nunca separó la señal de la interferencia. Este artículo metodológico desarrolla el problema oculto de la corrección: el riesgo de que un resultado más limpio se convierta en una representación más persuasiva sin convertirse en una medición más informativa. La proposición central es que el preprocesado solo puede recuperar información cuando esa información estaba presente en la medición bruta en una forma separable; no puede crear distinciones que el experimento nunca resolvió.

Palabras clave: corrección de datos; línea base; suavizado; normalización; sustracción; deconvolución; outlier; preprocesado; incertidumbre; datos brutos

1. Propósito y límite de evidencia

HEB-001 establece: las estadísticas no pueden fabricar información que el experimento nunca generó [1]. HEB-009 aplica ese principio al preprocesado y a la corrección. Los flujos analíticos específicos de aplicación permanecen con sus titulares ER; por ejemplo, ER-253 conserva la lógica de sustracción de blanco y umbrales para extraíbles [4].

2. Toda corrección es un modelo de medición

Restar un blanco supone que blanco y muestra comparten la misma contribución de fondo. Normalizar mediante una referencia supone que la referencia captura el cambio multiplicativo que afecta a la muestra. Suavizar supone que el componente de alta frecuencia eliminado es ruido y no señal. Ajustar una línea base supone una forma para el fondo. Una corrección no es, por tanto, aritmética neutral; es un modelo incrustado en el análisis [2].

3. Cuando sea viable, la eliminación física es más fuerte que la eliminación matemática

HEB-001 describe interferencia de película delgada introducida por películas de polietileno en una configuración RTIR de doble haz. La lección general es que un artefacto debe preferentemente prevenirse o eliminarse físicamente antes de diseñar una corrección numérica a su alrededor [1]. Una corrección matemática puede ser útil, pero hereda incertidumbre tanto de la medición original como del modelo de corrección.

4. La sustracción puede aumentar la incertidumbre

Si un resultado corregido es la diferencia entre dos mediciones inciertas, importan ambas contribuciones de incertidumbre. Cuando muestra y blanco son de magnitud similar, el valor corregido puede tener una precisión relativa deficiente incluso si cada entrada parece estable. Informar solo el número corregido oculta la pérdida de información creada por la sustracción.

5. La normalización puede desplazar el error al denominador

Una razón solo es tan estable como su denominador. Si varias muestras se dividen por una única referencia que deriva, la normalización puede trasladar un error común a todos los resultados informados y crear correlación en toda la serie corregida. La aparente reducción de dispersión no demuestra, por tanto, una mejora de validez; el denominador pertenece al modelo de medición [2].

6. El suavizado intercambia resolución por varianza

El suavizado puede hacer una traza visualmente coherente y mejorar la estimación de tendencias amplias, pero también atenúa características estrechas y modifica la correlación entre puntos vecinos. Por ello deben declararse el

método, la ventana y los parámetros, y las conclusiones científicas deben seguir siendo trazables hasta los datos sin suavizar. Una curva suave no es automáticamente un experimento más preciso [3].

7. La deconvolución es inferencia restringida

Resolver señales solapadas mediante ajuste de modelos puede ser potente cuando las formas de los componentes y la física de respuesta están justificadas. Es débil cuando varias descomposiciones ajustan los mismos datos de forma similar. La salida puede contener más parámetros numéricos de los que el experimento bruto restringió de manera independiente. Por tanto, la identificabilidad de parámetros debe separarse de la bondad visual de ajuste.

8. Eliminar outliers puede borrar el mecanismo

Un valor extremo puede ser un error de transcripción, un fallo instrumental o una observación genuina procedente de un estado diferente. Eliminarlo únicamente porque resulta incómodo hace que el modelo determine los datos. La investigación debe comenzar por identidad, datos brutos, calibración, desviación de protocolo y plausibilidad física. Análisis con y sin una exclusión trascendente pueden revelar si la conclusión depende de esa decisión.

9. Las cadenas de corrección acumulan supuestos ocultos

Un pipeline que realiza sustracción de blanco, normalización, ajuste de línea base, suavizado y exclusión selectiva puede contener varios pasos individualmente plausibles cuyo efecto combinado resulta difícil de auditar. Cada transformación debe conservar un vínculo con los datos brutos, los parámetros y la justificación. Cuanto más larga sea la cadena, más importante es probar el pipeline completo con entradas conocidas o simuladas.

10. La prueba de fallo complementario para el tratamiento de datos

Para cada corrección deben formularse dos preguntas: ¿qué fallo pretende suprimir? ¿Qué nuevo fallo puede introducir la corrección? Este es el principio de fallo complementario de HEB-001 aplicado al análisis [1]. Una corrección debe desafiarse mediante controles de sobrecorrección, infracorrección, sensibilidad a parámetros y conservación de características conocidas.

11. La incertidumbre de la corrección debe propagarse, no ocultarse

Un factor de corrección, modelo de fondo o valor de referencia tiene su propia incertidumbre. Si ese término se estima a partir de datos limitados, su incertidumbre puede estar correlacionada entre muchas observaciones corregidas. Tratar los valores corregidos como independientes puede, por tanto, subestimar la incertidumbre compartida. Por ejemplo, normalizar toda una serie mediante una única referencia incierta introduce un denominador común. Todos los puntos corregidos pueden desplazarse juntos si la referencia es errónea. La precisión aparente de las diferencias punto a punto y la incertidumbre de la escala absoluta son entonces preguntas diferentes. Un análisis defendible declara que incertidumbres son compartidas, cuáles son independientes y que transformación creó la dependencia.

12. La sensibilidad a las elecciones de preprocesado es evidencia sobre robustez

Cuando existen varias ventanas, líneas base o restricciones de ajuste razonables, la estabilidad de la conclusión científica frente a esas elecciones es informativa. Si un pico, tendencia o diferencia entre grupos existe solo para una configuración estrecha de preprocesado, la conclusión puede ser un artefacto del análisis. Si el resultado sobrevive a un intervalo declarado de configuraciones razonables y permanece visible en los datos brutos, aumenta la confianza. Esto no justifica probar muchos pipelines y seleccionar el que produce el resultado preferido. Las elecciones y las comprobaciones de robustez deben estar definidas por el problema de medición, idealmente antes de inspeccionar la comparación final.

13. Algunas pérdidas de información son irreversibles

Ningún postprocesado puede recuperar un valor de detector saturado, un promedio espacial no resuelto, contenido de señal fuera del ancho de banda efectivo del instrumento o un estado de probeta que nunca se registró. La interpolación solo puede estimar un punto ausente bajo un modelo; la deconvolución solo puede separar componentes en la medida en que la medición contenga estructura distinguible. El problema oculto de la corrección se vuelve grave cuando el análisis trata una pérdida irreversible de información como si fuera una molestia eliminable. La salida científica correcta puede ser una limitación o repetir el experimento, no un algoritmo más elaborado.

14. Conclusiones

El tratamiento de datos está justificado cuando su modelo de medición es explícito, comprobable y proporcional a la información contenida en los datos brutos. Las correcciones deben conservar las observaciones brutas, propagar la incertidumbre y mantener sensibilidad a elecciones alternativas plausibles de parámetros. Un conjunto de datos visualmente mejorado no es necesariamente un conjunto de datos epistemicamente mejorado. El punto final correcto no es la curva más limpia, sino la inferencia más defendible.

Referencias

- [1] Seguro, J. The Hidden Error Budget: Measurement-System Failure, False Precision and Experimental Design in Laboratory Science. HEB-001. 3Dresyns Independent Technical Articles (2026). DOI: 10.5281/zenodo.22862642.
- [2] Joint Committee for Guides in Metrology. Evaluation of Measurement Data - Guide to the Expression of Uncertainty in Measurement. JCGM 100:2008 with Amendment 1:2026, Nonlinearity in measurement models. Base DOI: 10.59161/JCGM100-2008E; amendment DOI: 10.59161/PPDI3267.
- [3] Heckert, N. A.; Filliben, J. J.; Croarkin, C. M.; Hembree, B.; Guthrie, W. F.; Tobias, P.; Prinz, J. Handbook 151: NIST/SEMATECH e-Handbook of Statistical Methods. National Institute of Standards and Technology, 2002.
- [4] Seguro, J. Analytical Evaluation Thresholds for Photopolymer Extractables: Coverage, Identification, Quantitation and Uncertainty. ER-253. 3Dresyns Engineering Publications (2026).

Limite de conocimiento

HEB-009 trata las consecuencias epistémicas generales del preprocesado y la corrección. No prescribe un pipeline analítico específico de aplicación ni sustituye ER-253 u otras revisiones de cualificación específicas de método.

Limitaciones

Este artículo metodológico es una síntesis acotada de los corpus controlados HEB y Engineering Review. No informa de un nuevo conjunto de datos experimental, no establece umbrales numéricos universales y no convierte ejemplos de un dominio en especificaciones aplicables a todas las aplicaciones. Las afirmaciones que requieren cualificación específica de aplicación permanecen en la revisión técnica o estudio primario correspondiente.