

El presupuesto oculto de error: fallos del sistema de medida, falsa precisión y diseño experimental en la ciencia de laboratorio

Juan Seguro

3Dresyns by Resyner Technologies S.L.

© 2026 Juan Seguro. Todos los derechos reservados.

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

HEB-001 · Artículo Técnico Independiente · Versión v1.0 · 20 de septiembre de 2026

Resumen

Los instrumentos modernos de laboratorio pueden mostrar muchas cifras numéricas mientras el sistema completo de medida permanece insuficientemente controlado. La calibración del instrumento, por sí sola, no establece la validez de una medición cuando la incertidumbre está dominada por la preparación de la muestra, la inestabilidad de la referencia, el posicionamiento, el acoplamiento ambiental, la deriva temporal, los procedimientos dependientes del operador, una replicación insuficiente o el exceso de cifras reportadas. Esta revisión metodológica desarrolla el concepto de un presupuesto oculto de error mediante estudios de caso documentados y retrospectivos procedentes principalmente de investigaciones fotoquímicas y de fotopolimerización realizadas a finales de la década de 1990, junto con observaciones de otros entornos de laboratorio.

Los casos incluyen la medida directa UV/visible de absorbancias muy bajas, medidas relativas del rendimiento cuántico de fosforescencia a 77 K, formación de escarcha sobre celdas ópticas criogénicas, deriva temporal de la referencia -que se convierte en sesgo cuando referencia y muestra no se miden en el mismo momento-, interferencia de película fina introducida por películas de polietileno en espectroscopia infrarroja en tiempo real de doble haz, aplicación no uniforme de recubrimientos en medidas de fotoamarilleamiento, vibración que afecta a medidas de dureza Vickers y acoplamiento del entorno térmico en microcalorimetría de flujo. También se utiliza un ejemplo contemporáneo de fotopolímeros para distinguir el grado de conversión del monómero libre residual, de los extractables/lixiviables y de la exposición biológica, ilustrando que un proxy medido con precisión no es automáticamente el mensurando pretendido. La revisión distingue explícitamente entre observaciones documentadas en fuentes contemporáneas, observaciones retrospectivas en primera persona del autor e interpretación metrológica actual.

Una segunda parte desarrolla prácticas constructivas: replicación, medias y desviaciones estándar, cifras significativas ajustadas a la incertidumbre, estimación del error puro, diseño factorial, métodos de superficie de respuesta y diseño de mezclas. El trabajo del autor de 1999 sobre mezclas de fotoiniciadores proporciona un ejemplo documentado del paso desde la experimentación de un factor cada vez hacia experimentos diseñados capaces de estimar interacciones y equilibrar respuestas múltiples. La proposición central es que el sistema de medida no es solamente el instrumento, sino el sistema acoplado formado por instrumento, espécimen, referencia, entorno, operador, protocolo y tratamiento estadístico. La sofisticación instrumental no puede compensar un componente débil no controlado, y la sofisticación estadística no puede crear información que el experimento nunca generó.

Palabras clave: incertidumbre de medida; repetibilidad; reproducibilidad; cifras significativas; falsa precisión; definición del mensurando; grado de conversión; extractables; validación de protocolo; acoplamiento ambiental; fotoquímica; RTIR; microcalorimetría de flujo; diseño factorial; diseño de mezclas; Diseño de Experimentos.

1. Propósito y frontera de evidencia

Este artículo es una revisión metodológica construida alrededor de un conjunto de casos históricos de laboratorio. Su propósito no es juzgar conductas indebidas, invalidar cuerpos completos de trabajo publicado ni atribuir culpa a personas o instituciones concretas. Su propósito es examinar cómo debilidades experimentales aparentemente elementales pueden persistir dentro de laboratorios técnicamente sofisticados y cómo esas debilidades pueden dominar la incertidumbre del resultado final.

Se utilizan tres clases de evidencia a lo largo del artículo:

1. Evidencia documentada contemporáneamente. Un método, limitación numérica o procedimiento correctivo aparece en la tesis doctoral del autor de 1999 o en una publicación derivada.
2. Observación retrospectiva del autor. El autor recuerda directamente haber observado una condición o práctica, pero todavía no se ha recuperado un registro contemporáneo para esta revisión.
3. Interpretación metrológica actual. Una interpretación contemporánea de por qué la condición observada importa para la validez de la medida. Esta interpretación debería evaluarse independientemente frente a fuentes metrológicas modernas antes de la publicación.

Estas clases no deben colapsarse entre sí. Una observación retrospectiva no se trata como prueba documental de la metodología completa de otro investigador, y una debilidad metodológica no demuestra por sí misma que toda conclusión histórica obtenida con ese método fuese incorrecta.

El artículo propone una tesis central:

Un instrumento sofisticado no convierte un procedimiento no controlado en una medición fiable.

De ella se deriva una segunda proposición:

El sistema de medida no es el instrumento. Es el instrumento, el espécimen, la referencia, el entorno, el operador, el protocolo y el tratamiento estadístico actuando conjuntamente.

2. Un número todavía no es una medición

Los instrumentos modernos proporcionan rutinariamente resultados con varios decimales. Esto crea una fuerte impresión visual de precisión. Sin embargo, el número de dígitos mostrados es una propiedad de la cadena de adquisición y cálculo de datos; no es automáticamente una propiedad del conocimiento físico generado por el experimento.

Una separación conceptual útil es:

resolución mostrada $\neq$ repetibilidad $\neq$ reproducibilidad $\neq$ veracidad $\neq$ incertidumbre de medida.

Por tanto, un resultado puede tener una resolución numérica elevada en pantalla y seguir siendo metrológicamente débil. Si recolocar una celda cambia la absorbancia en una magnitud comparable con la absorbancia que se pretende medir, cinco dígitos en la pantalla no convierten esos cinco dígitos en información. Si una respuesta calorimétrica varía cuando un sistema de calefacción próximo se enciende o se apaga, la existencia de un certificado de calibración del instrumento no establece que el entorno completo de medida sea estable. Si un recubrimiento presenta espesor no uniforme, un colorímetro altamente repetible puede seguir produciendo valores que no sean atribuibles exclusivamente a la química de la formulación.

La cuestión práctica no es simplemente si el instrumento produce un número. Es si el efecto de interés puede distinguirse de la variabilidad experimental combinada del método completo.

2.1 El mensurando equivocado puede medirse con precisión: conversión, monómero residual y extractables

Otra fuente de error experimental surge cuando una magnitud medida con precisión se utiliza como proxy de otra diferente sin demostrar la relación entre ambas. Los materiales fotopolimerizados en cuba ofrecen un ejemplo contemporáneo útil.

Un estudio de 2025 sobre Formlabs Clear V4 fotopolimerizado en cuba [11] midió el grado de conversión mediante espectroscopia ATR de infrarrojo medio a partir del cambio de la señal C=C alifática respecto de una referencia interna. Dado que ATR interroga una región superficial poco profunda y no todo el volumen del material, la conversión espectroscópica sigue siendo un proxy dependiente del muestreo para el estado de conversión del objeto impreso completo. El estudio utilizó diez especímenes polimerizados por grupo de conversión e informó de que el tiempo y la temperatura de poscurado afectaban significativamente a la conversión. Entre los grupos reportados, el menor grado de conversión medio fue 80,99% $\pm$ 0,61 y el mayor 97,20% $\pm$ 0,13 [11]. La resina ensayada se describió como basada en UDMA, lo que proporciona un ejemplo directamente relevante de una red multifuncional en la que la insaturación residual no puede equipararse numéricamente con monómero libre residual. Estos datos muestran que el estado químico de un fotopolímero impreso sigue dependiendo fuertemente del proceso incluso después de la etapa nominal de impresión.

El punto metrológico no es que el porcentaje no convertido pueda leerse automáticamente como el mismo porcentaje en masa de monómero libre residual. El grado de conversión es un mensurando basado en enlaces: informa de la fracción de grupos reactivos monitorizados que han reaccionado. El monómero móvil residual, los extractables y los lixiviables son mensurandos basados en composición y transporte. Están relacionados, pero no son intercambiables.

La funcionalidad molecular determina parte de esa relación. En un sistema idealizado compuesto únicamente por monómeros monofuncionales, si el único grupo reactivo de una molécula no reacciona, esa molécula queda sin incorporar; por tanto, una conversión incompleta puede corresponder de forma más directa a una población de monómero residual potencialmente móvil. En una red multifuncional, por el contrario, una molécula puede reaccionar mediante una funcionalidad, incorporarse a la red y conservar otro doble enlace sin reaccionar. Por ello, C=C residual no implica que la misma fracción del material permanezca como monómero libre.

A la inversa, cuando una formulación contiene una proporción importante de diluyente reactivo monofuncional de baja masa molar, una conversión incompleta aumenta la población plausible de moléculas que pueden permanecer libres y potencialmente extraíbles. La fracción realmente extraíble seguirá dependiendo del tamaño molecular, la polaridad, la afinidad con la red, la difusión, el lavado, la historia de poscurado, la geometría y las condiciones de extracción.

La distinción puede escribirse como:

grado de conversión $\neq$ fracción de monómero libre residual $\neq$ extractables/lixiviables $\neq$ exposición biológica.

Estas magnitudes pueden estar causalmente conectadas, pero una no puede sustituirse numéricamente por otra sin un modelo de medida respaldado experimentalmente.

Un proxy preciso no es el mensurando salvo que se haya demostrado la relación entre proxy y mensurando.

Este ejemplo extiende el concepto de error oculto más allá del rendimiento del instrumento. La validez experimental puede fallar no sólo porque una magnitud se mida mal, sino porque se mida bien la magnitud equivocada y después se interprete como si respondiera a otra pregunta.

3. Estudio de caso I: absorbancia UV/visible cerca del suelo práctico de medida

3.1 Método documentado

Durante el trabajo doctoral, las medidas relativas del rendimiento cuántico de fosforescencia requerían densidades ópticas bajas a la longitud de onda de excitación. La tesis indica que se utilizaron valores de absorbancia similares, alrededor de 0,05 a 290 nm en metilciclohexano, para referencia y muestra. También señala que la medida directa cerca de esta absorbancia podía conllevar un error relativo de aproximadamente 20%, "considerando la sensibilidad del detector alrededor de 0,01" [1].

En lugar de apoyarse en la lectura directa de baja absorbancia, los valores se obtuvieron a partir de soluciones más concentradas con absorbancia conocida, alrededor de 1, seguidas de una dilución por un factor conocido. La tesis afirma explícitamente que esto proporcionaba mayor exactitud que la medida directa de la baja absorbancia [1].

3.2 Interpretación metrológica

El punto importante no es que una absorbancia baja nunca pueda medirse directamente. Es que la medida práctica debe realizarse allí donde la relación entre señal útil y error de medida sea adecuada para la conclusión pretendida.

Cuando se ha demostrado linealidad de Beer-Lambert en el intervalo de concentración relevante, trasladar la medida primaria a una región de absorbancia más robusta y utilizar una dilución controlada puede reducir el error fotométrico relativo. Ese beneficio debe equilibrarse con la incertidumbre volumétrica introducida por la propia dilución.

Siempre que la ley física subyacente lo permita, mida en una región donde el instrumento funcione de forma fiable, en lugar de hacerlo cerca del suelo práctico de ruido.

En lenguaje metrológico moderno, la expresión de la tesis "sensibilidad del detector alrededor de 0,01" no debería tratarse automáticamente como una especificación completa de incertidumbre; identifica la escala de la limitación instrumental que motivó la estrategia de dilución. Este caso ilustra también por qué el número de cifras significativas debe seguir la incertidumbre del procedimiento y no la precisión numérica de la pantalla del espectrómetro.

4. Estudio de caso II: rendimiento cuántico relativo cuando la referencia se mide en el momento equivocado

4.1 Base documentada

La tesis describe la determinación relativa del rendimiento cuántico de fosforescencia utilizando benzofenona como referencia y establece un valor de referencia asumido de 0,74 en etanol absoluto a 77 K [1]. Las medidas experimentales de fotoiniciadores se realizaron en metilciclohexano, por lo que la transferencia de un valor de referencia entre condiciones de disolvente, la corrección del índice de refracción y el comportamiento del vidrio a baja temperatura son magnitudes que requieren control explícito en una implementación moderna. El método relativo depende de la fracción absorbida, intensidad de excitación, índice de refracción, área integrada de emisión, longitud de paso constante, posicionamiento reproducible de la celda, pérdidas de reflexión equivalentes y otros supuestos ópticos [1,7].

El programa experimental descrito utilizó un espectrómetro de luminiscencia Perkin-Elmer LS-5 y medidas a baja temperatura a 77 K [1].

4.2 Observación retrospectiva del autor: invalidación temporal de la referencia

El problema central observado por el autor no era simplemente que la señal de benzofenona tuviera una repetibilidad finita. El autor recuerda que la respuesta de luminiscencia del Perkin-Elmer cambiaba progresivamente durante una sesión de medida. Una referencia de benzofenona medida mucho antes o mucho después de una muestra desconocida podía representar un estado instrumental diferente de aquel en el que se había medido la muestra.

Dado que el rendimiento cuántico de fosforescencia se determinaba respecto de benzofenona, la separación temporal entre referencia y muestra se propagaba directamente al rendimiento cuántico calculado. Una referencia podía ser químicamente correcta y estar correctamente medida y, aun así, ser metrológicamente inapropiada para una muestra medida horas después si la respuesta instrumental había derivado durante ese intervalo.

El autor recuerda que la señal de benzofenona tendía a aumentar a medida que avanzaba la sesión experimental. En aquel momento se atribuyó a una equilibración térmica progresiva del sistema de medida durante el uso repetido de nitrógeno líquido. El cambio temporal observado es una observación retrospectiva del autor; la causa térmica propuesta sigue siendo una interpretación retrospectiva porque los datos originales de serie temporal no están disponibles aquí.

La corrección práctica adoptada por el autor fue encerrar cada medida desconocida entre medidas de benzofenona realizadas inmediatamente antes y después de la muestra:

BP1 → Sx → BP2

El propósito de esta secuencia no es simplemente promediar dos referencias. Es estimar la respuesta de referencia correspondiente al momento en que se midió la muestra. Si la deriva es aproximadamente lineal durante el corto intervalo de acotación, la respuesta local de referencia al tiempo de medida de la muestra (t_x) puede estimarse por interpolación temporal:

$$R(t_x) = R_1 + (t_x - t_1)/(t_2 - t_1) \cdot (R_2 - R_1).$$

Si la muestra se mide aproximadamente a mitad de camino entre ambas referencias, se reduce a:

$$R(t_x) \approx (R_1 + R_2)/2.$$

Esta corrección sólo es válida en la medida en que la deriva sea suficientemente suave durante el intervalo y la deriva relevante actúe sobre el término de respuesta que se normaliza. Una corrección por razón es apropiada para una deriva predominantemente multiplicativa de la sensibilidad; una deriva aditiva de línea base u offset requiere un modelo de corrección distinto.

4.3 Principio general: una referencia tiene una validez temporal

Una referencia correcta medida en el momento equivocado es una referencia incorrecta para un sistema de medida que deriva.

La validez de una medida de referencia está, por tanto, condicionada no sólo por la identidad, preparación y trazabilidad de la referencia, sino también por su proximidad temporal a la muestra desconocida cuando no se ha demostrado que la respuesta del instrumento sea estacionaria.

Las medidas de referencia deben ajustarse temporalmente a la muestra siempre que el sistema de medida no sea demostrablemente estable a lo largo de la secuencia.

Una única medida de benzofenona realizada al inicio de una sesión larga no puede corregir una muestra medida horas después si la respuesta de luminiscencia ha cambiado durante el intervalo. En tal situación, la razón relevante deja de comparar muestra y referencia bajo el mismo estado instrumental.

4.4 Interacción con la incertidumbre a baja absorbancia y la variabilidad de señal

Este error de referencia temporal se sumaba a otras contribuciones a la incertidumbre. La tesis indica que se utilizaron valores similares de absorbancia baja, alrededor de 0,05 a 290 nm, para referencia y muestra, y que la medida directa cerca de esa absorbancia podía conllevar un error relativo de aproximadamente 20% bajo el supuesto declarado sobre la sensibilidad del detector [1]. Para mejorar la exactitud, la tesis obtuvo esas absorbancias bajas a partir de soluciones más concentradas, de absorbancia conocida alrededor de 1, mediante dilución controlada [1].

La determinación relativa del rendimiento cuántico depende también de las áreas integradas de emisión y de otros factores ópticos. De forma simplificada:

$$\Phi_x = \Phi_R \cdot (\text{Area}_x/\text{Area}_R) \cdot (\text{BR}/\text{B}_x) \cdot (\text{IR}/\text{I}_x) \cdot (\text{n}_x^2/\text{n}_R^2),$$

donde los términos y convenciones exactos dependen de la implementación. La tesis identifica explícitamente la fracción absorbida, la intensidad de excitación, el índice de refracción, el área integrada de emisión y la geometría reproducible entre las dependencias de la determinación relativa [1].

En esta expresión, Area_x y Area_R son las áreas integradas bajo los espectros de emisión corregidos de muestra y referencia; B_x y BR son las fracciones de luz incidente absorbidas por muestra y referencia; I_x e IR son las intensidades de luz de excitación para muestra y referencia a la longitud de onda de excitación. Estas definiciones se indican en el método de la tesis [1, Capítulo V, p. impresa 50; p. PDF 61].

El autor recuerda además que la respuesta de luminiscencia integrada podía variar sustancialmente durante la secuencia experimental. Cuando la referencia no se medía localmente en el tiempo, esta variación no era una incertidumbre secundaria pequeña: podía invalidar el denominador normalizador utilizado para calcular el rendimiento cuántico de la muestra. Por ello, el error resultante se describe mejor como sesgo de referencia dependiente del tiempo que como mero ruido aleatorio.

4.5 Cuando la incertidumbre se aproxima al intervalo físico del mensurando

El rendimiento cuántico es una magnitud acotada: para la definición convencional relevante aquí, los valores físicamente admisibles están entre 0 y 1. Un resultado dentro de ese intervalo no es automáticamente informativo sólo porque el cálculo produzca un número plausible.

Si la incertidumbre procedente de la baja absorbancia, la variabilidad del área integrada, la deriva de referencia, el posicionamiento y la inestabilidad óptica crigénica se vuelve comparable con la magnitud del rendimiento cuántico estimado, la discriminación numérica puede colapsar. Un resultado ilustrativo como:

$$\Phi = 0,5 \pm 0,5$$

sería compatible, respetando los límites físicos, con casi todo el intervalo de 0 a 1. Tal resultado no constituiría una estimación bien resuelta de 0,5. El ejemplo numérico es ilustrativo y no se atribuye retrospectivamente a ningún valor publicado específico porque el presupuesto histórico completo de incertidumbre no puede reconstruirse ahora a partir del material conservado.

El punto metrológico esencial es que el fallo no tiene por qué surgir del ruido aleatorio del detector. Puede resultar de combinar una medida de baja absorbancia con una referencia obtenida en un estado instrumental distinto. En un método relativo, un denominador que deriva puede dominar el valor inferido incluso cuando cada lectura individual parece plausible.

Cuando la incertidumbre propagada abarca la mayor parte del intervalo físicamente admisible de un mensurando, el experimento no ha resuelto cuantitativamente ese mensurando.

Para conjuntos de datos históricos, esto no demuestra que todo orden cualitativo fuese erróneo. Significa que las afirmaciones que dependen de separaciones numéricas menores que la incertidumbre no resuelta requieren reevaluación frente a los datos brutos conservados, la estructura de replicación, los controles temporales de referencia y la cualificación del método.

5. Estudio de caso III: la escarcha se convirtió en un componente óptico no controlado

5.1 Observación retrospectiva del autor

Las medidas de fosforescencia a baja temperatura requerían enfriar la celda de muestra con nitrógeno líquido. El autor recuerda que la humedad ambiental producía condensación y escarcha sobre la celda fría. La cantidad y morfología de la escarcha variaban con las condiciones de la sala y con el tiempo.

Una respuesta operativa informal consistía en girar la celda para que el contacto con el soporte eliminara mecánicamente parte de la escarcha. Este procedimiento no definía la cantidad inicial de escarcha, la fracción retirada ni el estado óptico residual después de la manipulación.

5.2 Interpretación metrológica

Una capa variable de escarcha puede cambiar la transmisión, dispersión, reflexión y geometría óptica efectiva. Por tanto, pasa a formar parte del sistema óptico aunque no forme parte del experimento pretendido.

La respuesta robusta de diseño es la prevención, no la corrección manual: aislar la zona óptica criogénica de la humedad ambiental o utilizar una purga apropiada con gas seco para suprimir la condensación.

Si un contaminante ambiental entra en la trayectoria óptica, pasa a formar parte de la medición aunque el protocolo no lo reconozca.

Este caso ilustra también una jerarquía más general desarrollada posteriormente en el artículo: eliminar físicamente una fuente de variabilidad cuando sea posible antes de intentar corregirla numéricamente.

6. Estudio de caso IV: el haz de referencia RTIR generó su propia estructura de línea base

6.1 Configuración original documentada

La espectroscopia infrarroja en tiempo real se utilizó para seguir la fotopolimerización mediante la disminución de la banda del acrilato aproximadamente a 812 cm⁻¹. La tesis describe una muestra reactiva colocada entre placas de cloruro sódico cubiertas con dos piezas de película de polietileno de baja densidad, utilizando un separador de Teflón para proporcionar un espesor de película de 50 micrómetros. Se colocaron dos películas de polietileno en el haz de referencia para restar la absorción del polietileno [1].

La tesis también indica que la línea base para la banda de 812 cm⁻¹ se calculaba trazando una línea desde los hombros del pico [1]. Por tanto, la inestabilidad de esos hombros afecta directamente al área o altura calculada del pico y, en consecuencia, a la conversión derivada.

6.2 Observación retrospectiva del autor

El autor recuerda que los hombros alrededor de la banda de 812 cm⁻¹ oscilaban de forma impredecible cuando las películas de polietileno estaban presentes en el haz de referencia. Un colega explicó entonces el fenómeno en términos de una condición de interferencia óptica dependiente del espesor. En terminología actual, la interpretación más plausible es la presencia de franjas de interferencia de película fina o de Fabry-Perot asociadas a películas paralelas de espesor óptico finito. Esta interpretación física es retrospectiva y no es necesaria para establecer la observación empírica de que retirar las películas del haz de referencia redujo la variación estructurada de la línea base.

Cuando se retiraron las películas de polietileno del haz de referencia, la oscilación de los hombros se redujo considerablemente, permitiendo una línea base más estable y reproducible. Esa retirada sólo es metodológicamente defendible cuando la absorción del polietileno alrededor de la región analítica no produce por sí misma un error material de sustracción; esto debe comprobarse para el grado exacto de película y el espectro. Una geometría de medida posterior simplificó aún más la disposición de la muestra aplicando el material directamente sobre placas de cloruro sódico.

6.3 Principio general

La disposición original de referencia tenía una finalidad conceptual razonable: restar la contribución espectral del polietileno. El fallo fue que el componente de referencia no era ópticamente neutro y podía crear por sí mismo variación espectral estructurada.

Una referencia destinada a eliminar un artefacto puede convertirse ella misma en una fuente de artefacto.

Esto motiva una regla más general:

No corrija matemáticamente lo que primero pueda eliminarse físicamente.

La sustracción numérica sólo es fiable si la magnitud que se sustrae es, a su vez, estable y reproducible.

7. Estudio de caso V: los especímenes no uniformes pueden dominar las medidas de fotoamarilleamiento

7.1 Observación retrospectiva de un problema inicial de preparación

El autor recuerda un experimento inicial de fotoamarilleamiento en el que una resina viscosa se aplicó con una cuchilla. Los recubrimientos resultantes mostraban marcas visibles de la cuchilla y espesor de película no uniforme. Por tanto, los primeros resultados de color estaban afectados por una variable de preparación del espécimen que no formaba parte de la comparación pretendida entre fotoiniciadores.

7.2 Método final documentado

El método doctoral final estaba más controlado. El amarilleamiento de fotoiniciadores comerciales se midió con un colorímetro Datacolour utilizando el sistema CMC. Se aplicaron recubrimientos con 2% de fotoiniciador a un espesor nominal de 75 micrómetros con una barra de hilo Brown K n.º 7 sobre vidrio transparente, se curaron bajo condiciones definidas, se midieron tras 24 h y se replicaron. Los resultados se reportaron como valores medios, y la tesis describió el error relativo como inferior al 5% para las muestras ensayadas [1]. Puesto que la replicación estima precisión, pero no veracidad por sí sola, esa formulación histórica no debe leerse como una demostración moderna de exactitud de medida. El correspondiente estudio de fotoamarilleamiento se publicó en Polymer Degradation and Stability [2].

7.3 Interpretación metrológica

Para un recubrimiento transparente o translúcido medido ópticamente, el espesor y la topografía superficial pueden alterar la longitud de trayectoria óptica, la absorción, la reflexión y el estado local de curado. Un colorímetro preciso no puede eliminar la variabilidad que ya se incorporó al espécimen.

Si el espécimen no es reproducible, el instrumento no puede hacer reproducible la medición.

La preparación de la muestra no es, por tanto, una comodidad preliminar. Forma parte del sistema de medida.

8. Estudio de caso VI: acoplamiento térmico ambiental en microcalorimetría de flujo

8.1 Frontera histórica

Durante el mismo periodo institucional, el autor observó una instalación de microcalorimetría de flujo utilizada en un programa de investigación diferente. Este caso se incluye únicamente como observación en primera persona sobre el diseño del sistema de medida. No se vincula aquí a ningún investigador, empresa, tesis o publicación concretos, ni se utiliza para inferir la validez de ningún conjunto de datos histórico.

8.2 Observación retrospectiva del autor

El autor recuerda que el equipo de microcalorimetría de flujo estaba situado cerca de una fuente de calefacción por agua caliente. El entorno térmico local cambiaba cuando el sistema de calefacción del edificio alternaba entre estados de funcionamiento y reposo, y el autor observó una inestabilidad correspondiente en la respuesta calorimétrica. El mecanismo sospechado era un acoplamiento térmico entre el entorno de la sala y un sistema de medida destinado a resolver pequeños efectos térmicos.

El autor propuso separar o controlar de otra forma la perturbación térmica. No se ha recuperado para esta revisión ningún registro contemporáneo de instalación, serie temporal bruta ni presupuesto completo de incertidumbre. Por tanto, el caso sigue siendo una observación retrospectiva, no una constatación documental sobre el método final de otro investigador.

8.3 Interpretación metrológica

La microcalorimetría de flujo se utiliza precisamente porque los cambios térmicos asociados a adsorción o desorción transportan la información de interés. Por ello, un entorno térmico cambiante puede acoplarse directamente al canal de medida.

Cuando tanto el mensurando como la perturbación ambiental se expresan mediante calor, el aislamiento térmico y la estabilidad ambiental forman parte del método de medida. La replicación y el orden de ejecución aleatorizado pueden revelar o reducir la confusión causada por perturbaciones dependientes del tiempo, pero no sustituyen el control de una gran perturbación térmica conocida.

Una medición térmica no debería interpretarse independientemente del entorno térmico en el que se realiza.

9. Estudio de caso VII: vibración y medidas de dureza Vickers

9.1 Observación retrospectiva del autor

Durante el trabajo en un laboratorio industrial de investigación en 1995-1996, el autor recuerda que un equipo de dureza Vickers estaba instalado en una sala pequeña junto con otros equipos capaces de transmitir vibración mecánica. Las medidas mostraban una variabilidad sustancial. Se sospechó que la vibración mecánica era uno de los contribuyentes, y se propuso la separación física entre los equipos generadores de vibración y los sensibles a vibraciones. Otros posibles contribuyentes, entre ellos la preparación del espécimen, la carga y el tiempo de permanencia de la indentación y la lectura óptica de diagonales, no fueron excluidos independientemente en la evidencia conservada.

Este caso se informa como observación retrospectiva. Para la presente revisión no se han recuperado registros contemporáneos de la distribución de la sala, el intervalo de cargas ni las series individuales de medida.

9.2 Interpretación metrológica

Una medida de dureza puede realizarse con un dispositivo calibrado y, aun así, degradarse por el entorno de instalación. Por tanto, el sistema de medida relevante se representa mejor como:

instrumento + soporte + suelo + equipos cercanos + operador + espécimen.

Un instrumento calibrado instalado en un entorno inadecuado no constituye necesariamente un sistema de medida calibrado.

El principio no es específico de la dureza Vickers. Se aplica siempre que vibración, temperatura, humedad, movimiento de aire, ruido eléctrico u otra variable ambiental se acople al mensurando o al sistema de detección.

10. El laboratorio forma parte del instrumento

En las observaciones retrospectivas del autor sobre los casos criogénico, Vickers y microcalorimetría de flujo, el laboratorio circundante entraba en la cadena de medida.

Las variables relevantes pueden incluir:

- vibración y acoplamiento mecánico;
- temperatura y gradientes de temperatura;
- humedad y condensación;
- flujo de aire;
- luz parásita;
- ruido eléctrico y electromagnético;
- contaminación superficial;
- calentamiento, enfriamiento y deriva a largo plazo;
- acondicionamiento de la muestra e historial de manipulación.

La especificación del fabricante suele describir el rendimiento del instrumento bajo condiciones definidas. La medición de laboratorio, sin embargo, se realiza en una sala concreta, con una muestra concreta, una referencia, un dispositivo de sujeción y un operador concretos. Por ello, la cualificación de la medición sigue siendo responsabilidad del método experimental.

Una representación conceptual útil es:

$$M = f(I, S, R, E, O, P, A)$$

donde I es el instrumento, S el espécimen, R la referencia, E el entorno, O el operador, P el protocolo y A el análisis y tratamiento estadístico.

La expresión es conceptual y no una ecuación prescrita de incertidumbre. Su finalidad es evitar que el instrumento se confunda con el sistema completo de medida.

11. Las cifras significativas son una afirmación experimental

Una de las formas más visibles de metrología débil es la falsa precisión numérica.

Considérese un rendimiento cuántico de fosforescencia calculado:

$$\Phi = 0,67123.$$

El cálculo puede devolver legítimamente cinco decimales. Eso no significa que el experimento haya resuelto el rendimiento cuántico hasta cinco decimales.

Si una evaluación de incertidumbre respalda, por ejemplo, una incertidumbre expandida $U = 0,1$ con factor de cobertura $k = 2$, una representación como:

$$\Phi = 0,7 \pm 0,1 \text{ (} U, k = 2 \text{)}$$

comunica la información experimental con mayor honestidad que 0,67123. El ejemplo numérico es ilustrativo; no se atribuye retrospectivamente a ningún valor concreto de la tesis sin un presupuesto de incertidumbre.

La tesis del autor reportó muchos rendimientos cuánticos de fosforescencia con dos decimales, incluidos valores como 0,68, 0,75, 0,61 y 0,12 para los compuestos estudiados; 0,74 aparece también como valor de referencia de benzofenona utilizado en el método [1]. Retrospectivamente, cuando la incertidumbre práctica total era sustancialmente mayor que la última cifra reportada, incluso dos decimales podían implicar una resolución numérica superior a la justificada por el sistema completo de medida.

Esto no es una afirmación retrospectiva de que las tendencias reportadas fueran inválidas. Es una distinción entre resolución numérica calculada y resolución metrológica defendible. La distinción se vuelve especialmente grave para un mensurando acotado. Si el intervalo de incertidumbre ocupa una fracción grande del intervalo completo de rendimiento cuántico de 0 a 1, los decimales adicionales no añaden simplemente precisión cosmética: pueden sugerir visualmente una discriminación que el experimento no logró.

Las cifras significativas deberían describir el experimento, no la calculadora.

Una convención práctica de reporte consiste en indicar el tipo de incertidumbre comunicada y, cuando proceda, el factor de cobertura o la probabilidad de cobertura. La incertidumbre se redondea habitualmente a una o dos cifras significativas y el valor reportado a la posición decimal correspondiente, de acuerdo con la guía GUM [5].

12. De números aislados a evidencia experimental

12.1 Media y desviación estándar

Una medición única proporciona un valor pero no una estimación empírica de la variabilidad del método bajo condiciones repetidas.

Para mediciones replicadas independientes x_i , la media muestral es:

$$\bar{x} = (1/n) \sum x_i,$$

y la desviación estándar muestral es:

$$s = \sqrt{[\sum (x_i - \bar{x})^2 / (n - 1)]}.$$

La media y la desviación estándar responden a preguntas diferentes. La media estima una respuesta central bajo las condiciones ensayadas; la desviación estándar cuantifica la dispersión observada entre réplicas.

El procedimiento doctoral RTIR indica explícitamente que se realizó replicación para estimar errores puros y prevenir errores graves. Los resultados se expresaron como valores medios y desviaciones estándar, y las gráficas de reactividad se promediaron [1]. Un estudio cinético publicado y relacionado con el mismo programa de investigación proporciona contexto adicional a nivel de publicación [3]. En las medidas de color, la tesis indica que se midieron muestras replicadas para obtener una estimación de "accuracy" y que los resultados se expresaron como valores medios [1]. En terminología metrológica moderna, la replicación por sí sola estima un componente de precisión; no establece veracidad ni exactitud sin una referencia apropiada o una evaluación del sesgo [6].

La misma tesis señala que son deseables al menos cinco medidas replicadas de dureza pendular porque las imperfecciones locales de la película pueden afectar lecturas individuales [1].

Sin replicación, la dispersión experimental es desconocida; sin dispersión experimental, la precisión numérica no puede justificarse.

12.2 Las lecturas repetidas no son automáticamente réplicas independientes

La replicación debe definirse al nivel pertinente para la incertidumbre que se pretende estimar.

Cinco lecturas repetidas del instrumento sin retirar el espécimen pueden estimar la repetibilidad de lectura a corto plazo, pero no el error de reposicionamiento. Cinco puntos sobre un mismo recubrimiento pueden caracterizar heterogeneidad superficial local, pero no la preparación del recubrimiento entre lotes. Cinco especímenes preparados independientemente pueden capturar un componente más amplio de variación experimental.

Por tanto, las publicaciones deberían indicar qué se ha replicado: lecturas, posiciones, especímenes, preparaciones, días experimentales, operadores o combinaciones de estos niveles.

12.3 Una media sin dispersión está incompleta

Supóngase que dos condiciones tienen cada una $n = 5$ réplicas independientes y proporcionan medias de 10,2 y 10,8. La diferencia numérica es 0,6.

Si las desviaciones estándar correspondientes son 0,1 y 0,1, el error estándar de la diferencia es pequeño en relación con 0,6 bajo un modelo simple de muestras independientes. Si las desviaciones estándar son 1,8 y 2,1, el error estándar de la diferencia es mucho mayor. La inferencia formal depende del tamaño muestral, la estructura de varianzas y el diseño, no sólo de las desviaciones estándar.

Los números 10,2 y 10,8 no han cambiado. La interpretación sí.

La magnitud de una diferencia debe juzgarse en relación con la variabilidad del método que la produjo.

12.4 Error puro y falta de ajuste

El capítulo de diseño experimental de la tesis indica que las observaciones replicadas permiten estimar el error puro y, por tanto, realizar una prueba de falta de ajuste. También identifica Root MSE como estimador del error experimental [1]. Más precisamente, el Root MSE residual puede representar la desviación estándar común del error sólo cuando el modelo ajustado es adecuado; de lo contrario, el término residual puede contener tanto error puro como falta de ajuste.

Esto es más que un detalle estadístico. Sin una estimación empírica del error, puede resultar difícil distinguir un modelo deficiente de un sistema de medida ruidoso. La replicación proporciona la línea base frente a la cual puede juzgarse la inadecuación del modelo.

13. Un factor cada vez frente al Diseño Factorial de Experimentos

13.1 Por qué OFAT resulta atractivo

La experimentación de un factor cada vez (OFAT) es intuitiva: seleccionar una condición de referencia, cambiar el factor A, identificar un nivel aparentemente favorable, mantenerlo fijo, cambiar después el factor B y continuar secuencialmente. Los experimentos diseñados pueden extraer información sobre interacciones que estos recorridos OFAT no estiman de forma eficiente [8].

Para sistemas simples, monótonos y con interacciones despreciables, OFAT puede ser útil para una exploración preliminar. Su limitación es estructural: no cartografía eficientemente las interacciones entre factores.

13.2 La interacción es información

Para dos factores, un modelo sencillo puede escribirse como:

$$Y = \beta_0 + \beta_A A + \beta_B B + \beta_{AB} AB + \varepsilon.$$

El término AB representa interacción: el efecto de cambiar A depende del nivel de B.

Un recorrido OFAT puede perder esta dependencia porque explora sólo determinadas líneas a través del espacio experimental. Un diseño factorial combina deliberadamente niveles de los factores para que puedan estimarse efectos principales e interacciones dentro de los supuestos del diseño.

Para tres factores a dos niveles A, B y C, un diseño factorial completo contiene:

$$2^3 = 8$$

combinaciones de factores. Sujeto a una estructura adecuada de error y replicación, esas ejecuciones pueden proporcionar información sobre los tres efectos principales, las tres interacciones de dos factores y la interacción de tres factores.

OFAT pregunta qué hace un factor mientras los demás permanecen congelados. El diseño factorial pregunta cómo responde el sistema cuando los factores actúan conjuntamente.

13.3 Un ejemplo conceptual sencillo

Imagínese que la respuesta de fotopolimerización depende de la temperatura y de la concentración de fotoiniciador. Una secuencia OFAT podría examinar primero la temperatura con 2% de fotoiniciador, elegir 30 °C y estudiar después la concentración únicamente a 30 °C. Si el efecto de la concentración de fotoiniciador cambia fuertemente con la temperatura, el óptimo puede encontrarse en una combinación que el recorrido OFAT nunca ensaya.

El fenómeno perdido no es ruido experimental. Es una interacción que el diseño no estaba estructurado para resolver.

Una interacción no es ruido; es información que un diseño OFAT puede ser estructuralmente incapaz de detectar.

14. Diseño factorial, superficies de respuesta y diseño de mezclas son herramientas diferentes

DOE no es una única disposición experimental. El diseño debe corresponder a la estructura de los factores.

14.1 Diseños factoriales

Los diseños factoriales son apropiados cuando los factores pueden variarse independientemente dentro de rangos definidos. Algunos ejemplos pueden ser tiempo de exposición, temperatura, irradiancia, concentración de aditivo o duración de poscurado, siempre que los factores elegidos sean físicamente independientes dentro del espacio de diseño especificado.

14.2 Métodos de superficie de respuesta

Cuando se espera curvatura y el objetivo incluye optimización dentro de una región continua, un factorial de dos niveles no puede estimar por sí solo términos cuadráticos puros. Los puntos centrales pueden revelar curvatura, mientras que diseños de superficie de respuesta de segundo orden, como diseños compuestos centrales o Box-Behnken, pueden estimar comportamiento cuadrático [9].

14.3 Diseños de mezclas

Las proporciones de formulación requieren una geometría distinta porque las fracciones de los componentes están restringidas:

$$\sum x_i = 1.$$

Aumentar un componente obliga a disminuir al menos otro. Los factores, por tanto, no son independientes en el sentido factorial convencional.

El trabajo del autor de 1999 sobre mezclas eutécticas de fotoiniciadores utilizó explícitamente Diseño de Experimentos y diseños de mezclas para optimizar la reactividad y la estabilidad en masa de mezclas que contenían benzofenona, benzil dimetil cetil, isopropil tioxantona y etildimetilaminobenzoato [1,4]. La tesis documenta modelización por regresión, evaluación basada en ANOVA, análisis de residuos y errores estándar de parámetros, mientras que el estudio publicado presenta la optimización mediante diseño de mezclas y el compromiso reactividad-estabilidad [1,4].

El estudio mostró un compromiso importante: proporciones relativamente altas de etildimetilaminobenzoato (EPD) se asociaban a mayor reactividad pero menor estabilidad, mientras que la benzofenona contribuía fuertemente a la estabilidad de la mezcla. Por tanto, se utilizó optimización de respuestas múltiples para identificar composiciones que equilibraran más de un objetivo de rendimiento [1,4].

Este ejemplo ilustra por qué la optimización de formulaciones está mal representada por ajustes independientes de un factor cada vez. El espacio experimental está restringido, los componentes interactúan y la respuesta útil puede ser una región de compromiso y no el máximo aislado de una sola respuesta. La geometría y el análisis de diseño de mezclas se tratan específicamente en la literatura de experimentos con mezclas [10].

15. DOE no puede rescatar un método de medida inválido

DOE aumenta la información extraída de un experimento sólo si la respuesta que se modeliza representa adecuadamente el mensurando pretendido.

La aleatorización y el bloqueo son protecciones centrales frente a la variación molesta. Un orden aleatorizado de ejecución ayuda a evitar que la deriva temporal o los ciclos ambientales se alineen con los niveles de los factores, mientras que el bloqueo puede separar fuentes de variación conocidas, como día, lote u operador, de los efectos de los factores de interés [9]. Estas herramientas reducen la confusión, pero no hacen desaparecer la perturbación; una gran variación no controlada sigue inflando la varianza del error y reduciendo la potencia.

Si los valores de respuesta están contaminados por deriva, espesor variable del espécimen, acoplamiento térmico, vibración o una referencia inestable, y esas perturbaciones se alinean con los niveles de los factores -por ejemplo, porque no se aleatorizó el orden de ejecución-, un modelo estadístico puede atribuir estructura molesta a los factores experimentales. A la inversa, un sesgo sistemático constante común a todas las ejecuciones, un valor de referencia incorrecto o una respuesta que no mida la propiedad pretendida no pueden repararse mediante aleatorización o bloqueo.

Por tanto, una secuencia robusta es:

cualificar el método de medida → estimar repetibilidad → definir replicación → diseñar el experimento, incluyendo aleatorización/bloqueo cuando proceda → ajustar y validar el modelo → optimizar.

DOE aumenta la información de un proceso de medida cualificado; no convierte una respuesta inválida en un mensurando válido.

La estadística no puede fabricar información que el experimento nunca generó.

16. La herencia de protocolos como modo de fallo

Los procedimientos de laboratorio se transmiten a menudo mediante demostración y práctica local, además de mediante métodos formales escritos. Esa transmisión puede ser eficiente, pero un procedimiento puede estabilizarse socialmente antes de que todos los supuestos relevantes hayan sido cuestionados desde el punto de vista metrológico.

No todos los casos de esta revisión deberían describirse como procedimientos heredados. Algunos son métodos documentados, otros son prácticas tempranas del propio autor y otros son observaciones retrospectivas de condiciones de laboratorio. No obstante, ilustran tipos de procedimiento que pueden persistir cuando sus supuestos no se ponen a prueba deliberadamente. Entre los ejemplos se encuentran asumir que una referencia permanece estable durante una secuencia larga, retirar manualmente la condensación en lugar de prevenirla, añadir un componente al haz de referencia sin comprobar si crea estructura espectral, aceptar

preparación no uniforme de especímenes o instalar equipos sensibles sin cualificar el entorno local.

El problema no es que esas decisiones sean irracionales. Cada una puede tener una justificación operativa. El problema es que una justificación plausible no equivale a una validez de medida demostrada.

La reproducibilidad de un procedimiento no es lo mismo que la validez de un procedimiento.

Un método puede reproducirse consistentemente y conservar al mismo tiempo la misma influencia no controlada.

17. El presupuesto oculto de error

Un presupuesto convencional de error contiene fuentes de incertidumbre que han sido reconocidas y estimadas. Un presupuesto oculto de error está formado por fuentes que todavía no se han reconocido como parte de la medición.

Conceptualmente:

$Y_{medido} = f(Y_{verdadero}, EI, ES, ER, EE, EO, ET, EA),$

donde los términos de error representan contribuciones del instrumento, espécimen, referencia, entorno, operador, tiempo y análisis.

Estas contribuciones no son necesariamente independientes ni aleatorias. Una capa de escarcha puede cambiar simultáneamente transmisión y dispersión. Una variación de espesor de película puede alterar tanto la densidad óptica como el estado de curado. La temperatura puede afectar tanto a la electrónica del instrumento como al espécimen. El reposicionamiento puede alterar simultáneamente la alineación y la línea base.

Por ello, el error más peligroso no es necesariamente el mayor componente conocido. Puede ser el componente que el experimentador todavía no ha identificado.

Esto sugiere una estrategia práctica: perturbar deliberadamente el sistema de medida antes de confiar en él.

Algunos ejemplos son:

- retirar y reinsertar repetidamente la misma celda;
- repetir referencias a lo largo del día;
- medir antes y después de la equilibración térmica del instrumento;
- encender y apagar equipos sospechosos de producir vibración;
- ensayar diferentes condiciones de temperatura o humedad de la sala cuando sea seguro y pertinente;
- retirar un elemento óptico sospechoso de producir interferencia;
- hacer que un segundo operador prepare especímenes de forma independiente;
- repetir el proceso completo de preparación del espécimen en lugar de repetir únicamente la lectura instrumental.

Un método robusto debería permanecer dentro de límites de aceptación predefinidos bajo perturbaciones esperables durante el uso normal.

18. El principio de fallo complementario

Los controles experimentales se introducen habitualmente para suprimir un modo de fallo conocido. Sin embargo, cada control puede introducir un supuesto nuevo.

Las películas de polietileno en el haz de referencia RTIR ilustran este punto. Su finalidad era restar la absorción del polietileno, pero las propias películas podían generar estructura de interferencia que comprometía los hombros utilizados para construir la línea base de 812 cm⁻¹.

Por tanto, la pregunta apropiada no es sólo:

¿Qué fallo evita este control?

sino también:

¿Qué fallo nuevo puede introducir este control?

Esta revisión propone el siguiente principio de fallo complementario:

Para cada control experimental introducido con el fin de suprimir un modo de fallo, compruebe explícitamente si el control crea un segundo modo de fallo.

El principio se aplica a referencias, blancos, dispositivos de fijación, algoritmos de corrección, controles ambientales e incluso al preprocesamiento estadístico.

19. Implicaciones retrospectivas para datos históricos

La identificación de una variable no controlada plantea una pregunta difícil: ¿qué debería concluirse sobre datos antiguos obtenidos bajo el mismo procedimiento o uno similar?

La respuesta no puede ser global. Una debilidad metodológica no justifica afirmar que todos los datos históricos son falsos.

Deben distinguirse al menos tres casos.

19.1 Conclusión cualitativa robusta

El efecto observado es mucho mayor que la variabilidad experimental plausible. El valor numérico exacto puede ser incierto, pero la conclusión cualitativa sigue respaldada.

19.2 Precisión numérica cuestionable

La tendencia puede ser real, pero el número de cifras significativas o la magnitud de la diferencia reportada supera lo que el método puede sostener.

19.3 Conclusión potencialmente vulnerable

El efecto afirmado es del mismo orden que la variabilidad no controlada. Puede ser necesaria una nueva medida bajo un protocolo cualificado antes de considerar cuantitativamente segura la conclusión.

Este marco es intencionadamente conservador. Separa la crítica de un método de las alegaciones sobre intención. La reevaluación histórica debería basarse en protocolos conservados, datos brutos, estructura de replicación y tamaño del efecto cuando esos materiales sigan disponibles.

20. Un marco práctico de cualificación del sistema de medida

Antes de utilizar un método cuantitativo para comparar formulaciones, materiales o tratamientos, deberían responderse las siguientes preguntas.

20.1 Instrumento

- ¿Está el instrumento calibrado para el intervalo de uso previsto?
- ¿Cuál es su resolución práctica, no sólo nominal?
- ¿Deriva la respuesta con el calentamiento, el enfriamiento o el tiempo?

20.2 Espécimen

- ¿Es reproducible la preparación del espécimen?
- ¿Están controlados el espesor, la geometría, el estado superficial y el acondicionamiento?
- ¿Se utilizan especímenes independientes cuando importa la variabilidad entre especímenes?

20.3 Referencia y blanco

- ¿Es la referencia más estable que la señal que pretende corregir?
- ¿Cambia la respuesta de referencia al reposicionarla?
- ¿Puede la propia referencia introducir interferencia u otro artefacto?

20.4 Entorno

- ¿Pueden vibración, temperatura, humedad, condensación, flujo de aire, luz o ruido eléctrico afectar a la medición?
- ¿Están definidos límites de aceptación para las variables ambientales pertinentes?

20.5 Operador

- ¿Depende el resultado de una manipulación subjetiva?
- ¿La retirada y inserción modifican materialmente el resultado?
- ¿Se ha evaluado la reproducibilidad entre operadores cuando sea relevante?

20.6 Tiempo

- ¿Deriva el instrumento o el espécimen durante la secuencia de medida?
- ¿La calibración o el acotamiento de referencias son suficientemente frecuentes para detectar esa deriva?

20.7 Replicación

- ¿Qué constituye una réplica independiente?
- ¿Es suficiente el número de réplicas para estimar la variabilidad relevante para la afirmación?

20.8 Cifras significativas

- ¿Coincide el número de dígitos reportados con la incertidumbre observada?
- ¿Se reportan estadísticas de incertidumbre o dispersión junto a la media?

20.9 Diseño experimental

- ¿Pueden interactuar múltiples factores?
- ¿Está aleatorizado el orden de ejecución cuando una variación molesta dependiente del tiempo podría confundirse con los efectos de los factores?
- ¿Deben bloquearse variables molestas conocidas como día, lote, instrumento u operador?
- ¿Es suficiente OFAT para la pregunta estrecha planteada o se necesita DOE factorial o de superficie de respuesta?
- ¿Son las variables componentes de una mezcla sujetos a una restricción de suma, lo que requiere métodos de diseño de mezclas?

20.10 Validación

- ¿Se examinan los residuos del modelo y la falta de ajuste?

- ¿Se comprueba experimentalmente el óptimo predicho en lugar de aceptarlo sólo a partir del modelo?
- ¿Puede reproducirse el resultado en un día distinto, con otra preparación de espécimen u otro operador cuando la afirmación requiera ese nivel de robustez?

21. Discusión

Los casos individuales descritos abarcan espectroscopia, fotoquímica, fotopolimerización, colorimetría, calorimetría y ensayos mecánicos. Sugieren que el instrumento nominal no tiene por qué ser la fuente dominante de incertidumbre; ninguno de los casos históricos presentados proporciona una descomposición cuantitativa completa de todos los componentes de incertidumbre.

En su lugar, la contribución limitante podía surgir del posicionamiento de la celda, la condensación, la deriva de la referencia, la interferencia de película fina, la preparación del espécimen, la temperatura de la sala, la vibración mecánica o una replicación insuficiente.

Esto produce una asimetría recurrente en el trabajo experimental. Cuanto más sofisticado es el instrumento, más fácil resulta suponer que la parte difícil del experimento ya ha sido resuelta por el fabricante. Sin embargo, los fabricantes cualifican los instrumentos bajo condiciones definidas. Los investigadores deben cualificar las mediciones bajo las condiciones en las que realmente los utilizan.

La respuesta constructiva no es el pesimismo metodológico. Es diseñar el experimento para que la incertidumbre se vuelva visible.

La replicación expone la dispersión. Las medias y desviaciones estándar evitan que lecturas aisladas se presenten como magnitudes estables. Las cifras significativas pueden ajustarse a la escala resuelta experimentalmente. Los diseños factoriales revelan interacciones que OFAT puede perder. Los diseños de mezclas respetan la dependencia matemática de las proporciones de formulación. El análisis de residuos, la estimación del error puro y las pruebas de falta de ajuste distinguen estructura de modelo de dispersión experimental.

Sin embargo, estas herramientas forman una secuencia, no sustitutos entre sí. La replicación no elimina el sesgo sistemático. DOE no repara un entorno instrumental inestable. ANOVA no puede separar una fuente de error perfectamente confundida con un factor experimental. Por tanto, la aleatorización y el bloqueo forman parte del diseño del sistema de medida, no son decoración estadística opcional. Un modelo altamente significativo puede seguir describiendo una respuesta sistemáticamente sesgada.

La arquitectura experimental más sólida comienza, por tanto, antes de recoger el primer conjunto de datos: identificar fuentes plausibles de error, perturbar deliberadamente el sistema, simplificar la medida física cuando sea posible, cuantificar la repetibilidad y sólo entonces utilizar diseño estadístico para extraer eficientemente los efectos de los factores.

22. Conclusiones

Los casos examinados en esta revisión respaldan un conjunto de conclusiones metodológicas generales.

1. Una salida numérica no es automáticamente una medición válida.
2. La calibración de un instrumento no valida el sistema completo de medida.
3. Una referencia puede aumentar la incertidumbre si es menos estable que la señal que pretende corregir.
4. Las variables ambientales pasan a formar parte del instrumento cuando se acoplan al canal de medida.
5. La preparación de la muestra forma parte del sistema de medida.
6. La replicación es necesaria para estimar la dispersión experimental al nivel relevante para la afirmación.
7. Las medias deberían acompañarse normalmente de una medida apropiada de dispersión.
8. Las cifras significativas deberían reflejar la incertidumbre experimental y no la salida de la calculadora.
9. OFAT puede perder interacciones que DOE factorial está específicamente diseñado para estimar.
10. DOE de mezclas es necesario cuando las proporciones de formulación están restringidas y no pueden variar independientemente.
11. DOE puede aumentar la información procedente de medidas fiables, pero no puede hacer válida una medición no controlada.
12. Los procedimientos de laboratorio, estén escritos o transmitidos mediante práctica local, deberían tratarse como métodos que deben someterse a prueba y no como supuestos que deban aceptarse sin cualificación.

La conclusión más amplia es:

La fiabilidad de un experimento no está limitada por la sofisticación de su mejor componente, sino por el comportamiento no controlado de su componente relevante más débil.

Por tanto, un resultado experimental defendible requiere la cualificación del sistema completo de medida: instrumento, espécimen, referencia, entorno, operador, protocolo y estadística.

Referencias

- [1] Seguro, J. Photochemical Characteristics and Design of Synergistic Photoinitiator Combinations for UV Curable Inks and Coatings. Doctoral thesis, Manchester Metropolitan University, Department of Chemistry and Materials, in collaboration with Swale Process, Manchester, UK, 1999.
- [2] Seguro, J.; Allen, N. S.; Edge, M.; McMahon, A.; Wilson, S. Photoyellowing and discolouration of UV cured acrylated clear coatings systems: influence of photoinitiator type. *Polymer Degradation and Stability* 64(1) (1999) 39-48. DOI: 10.1016/S0141-3910(98)00169-4.
- [3] Allen, N. S.; Seguro, J.; Edge, M.; McMahon, A.; Roberts, I. A comparative kinetic study of UV-curable resins: the relationship between structure and rheology. *Surface Coatings International* 82(6) (1999) 285-292. DOI: 10.1007/BF02720124.

- [4] Seguro, J.; Allen, N. S.; Edge, M.; McMahon, A. Design of eutectic photoinitiator blends for UV/visible curable acrylated printing inks and coatings. *Progress in Organic Coatings* 37(1-2) (1999) 23-37. DOI: 10.1016/S0300-9440(99)00052-1.
- [5] Joint Committee for Guides in Metrology. Evaluation of measurement data - Guide to the expression of uncertainty in measurement. JCGM 100:2008. DOI: 10.59161/JCGM100-2008E.
- [6] Joint Committee for Guides in Metrology. International Vocabulary of Metrology - Basic and general concepts and associated terms (VIM), 3rd ed. JCGM 200:2012. DOI: 10.59161/JCGM200-2012.
- [7] Brouwer, A. M. Standards for photoluminescence quantum yield measurements in solution (IUPAC Technical Report). *Pure and Applied Chemistry* 83(12) (2011) 2213-2228. DOI: 10.1351/PAC-REP-10-09-31.
- [8] Czitrom, V. One-Factor-at-a-Time versus Designed Experiments. *The American Statistician* 53(2) (1999) 126-131. DOI: 10.1080/00031305.1999.10474445.
- [9] 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, Gaithersburg, MD, 2002. Relevant DOE material includes §5.3, "Choosing an experimental design"; §5.3.3.3 on full factorial designs; §5.3.3.3.3 on blocking of full factorial designs; and the process-modelling/response-surface sections. Official NIST record: <https://www.nist.gov/publications/handbook-151-nistsematech-e-handbook-statistical-methods> (consultado el 20 de septiembre de 2026).
- [10] Cornell, J. A. *Experiments with Mixtures: Designs, Models, and the Analysis of Mixture Data*, 3rd ed. Wiley, 2002. DOI: 10.1002/9781118204221.
- [11] Cheadle, A. M. G.; Maier, E.; Palin, W. M.; Tomson, P. L.; Poologasundarampillai, G.; Hadis, M. A. The impact of modifying 3D printing parameters on mechanical strength and physical properties in vat photopolymerisation. *Scientific Reports* 15 (2025) 12592. DOI: 10.1038/s41598-025-97294-8.

Frontera de conocimiento

Las afirmaciones históricas documentales de este manuscrito se limitan a la tesis de 1999 y a las publicaciones citadas. Los recuerdos en primera persona están explícitamente etiquetados y no se utilizan como evidencia documental sobre el método o conjunto de datos completo de otra persona. La interpretación metrológica moderna se fundamenta en las fuentes GUM, VIM, IUPAC y DOE citadas, pero no se ha aplicado retrospectivamente como un presupuesto cuantitativo de incertidumbre a los conjuntos de datos históricos.

Limitaciones

No se reanalizó ninguna serie temporal histórica de datos brutos para esta revisión. No se reconstruyó ningún presupuesto completo de incertidumbre para las medidas de la década de 1990. Los casos de microcalorimetría de flujo y Vickers siguen siendo observaciones retrospectivas. El mecanismo de interferencia de película fina en RTIR es una interpretación actual coherente con el comportamiento observado de la línea base, no un diagnóstico contemporáneo conservado. Por tanto, los datos históricos sólo deberían reevaluarse al nivel respaldado por los métodos, registros brutos, estructura de replicación y tamaños de efecto que se conserven.

Control de fuente

La versión 1.0 adopta el identificador independiente HEB-001 y conserva sin alteración científica el contenido aprobado independientemente en v0.7. HEB-001 es un artículo técnico autónomo fuera de la secuencia ER-001 a ER-500. Su temática es complementaria a los Engineering Reviews relacionados identificados en el control de propiedad y no sustituye, renumera ni absorbe ningún tema del roadmap ER. La maquetación final, la contrafirma del autor y las acciones de repositorio siguen siendo pasos controlados separados.

© 2026 Juan Seguro. Todos los derechos reservados.