Towards a More Realistic Treatment of Uncertainty in Earth and Environmental Sciences: Beyond a Simplified Subdivision into Interval and Random Components

Christian Servin[1,4], Craig Tweedie[2,4], and Aaron Velasco[3,4]

[1]Computational Sciences Program
[2]Environmental Science and Engineering Program
[3]Department of Geological Sciences
[4]Cyber-ShARE Center
University of Texas, El Paso, Texas 79968, USA
christains@utep.edu, ctweedie@utep.edu,
velasco@geo.utep.edu

1. Measurement Uncertainty: Reminder

- Usually, a meas. error $\Delta x \stackrel{\rm def}{=} \widetilde{x} - x$ is subdivided into *random* and *systematic* components $\Delta x = \Delta_s x + \Delta_r x$:
  - the systematic error component $\Delta_s x$ is usually defined as the expected value $\Delta_s x = E[\Delta x]$, while
  - the random error component is usually defined as the difference $\Delta_r x \stackrel{\rm def}{=} \Delta x - \Delta_s x$.
- The random errors $\Delta_r x$ corresponding to different measurements are usually assumed to be independent.
- For $\Delta_s x$, we only know the upper bound Δ_s s.t. $|\Delta_s x| \le \Delta_s$, i.e., that $\Delta_s x$ is in the *interval* $[-\Delta_s, \Delta_s]$.
- Because of this fact, *interval computations* are used for processing the systematic errors.
- $\Delta_r x$ is usually characterized by the corr. probability distribution (usually Gaussian, with known σ).

2. Problem

- Often, the differences $\Delta_r x = \Delta x - \Delta_s x$ corr. to nearby times are strongly correlated.
- For example, meteorological sensors may have daytime or nighttime biases, or winter and summer biases.
- To capture this correlation, environmental scientists proposed a semi-heuristic 3-component model of Δx.
- In this model, the difference $\Delta x - \Delta_s x$ is represented as a combination of:
  - a "truly random" error $\Delta_t x$ (which is independent from one measurement to another), and
  - a new "periodic" component $\Delta_p x$.
- We provide a theoretical explanation for this heuristic three-component model.

3. Analysis of the Problem

- We want to represent measurement error $\Delta x(t)$ as a linear combination of several components.
- We consider the most detailed level of granularity, w/each component determined by finitely many parameters c_i.
- Each component is thus described by a finite-dimensional linear space

$$L = \{c_1 \cdot x_1(t) + \ldots + c_n \cdot x_n(t) : c_1, \ldots, c_n \in \mathbb{R}\}.$$

- In most applications, signals are smooth and bounded, so we assume that $x_i(t)$ is smooth and bounded.
- Finally, for a long series of observations, we can choose a starting point arbitrarily: $t \to t + t_0$.
- It is reasonable to require that this change keeps us within the same component, i.e.,

$$x(t) \in L \Rightarrow x(t + t_0) \in L.$$

4. Definitions and the Main Result

- A function $x(t)$ of one variable is called *bounded* if

$$\exists M\, \forall t\, (|x(t)| \le M).$$

- We say that a class F of functions of one variable is *shift-invariant* if

$$\forall x(t)\, (x(t) \in F \Rightarrow \forall t_0\, (x(t+t_0) \in F)).$$

- By an *error component* we mean a shift-invariant finite-dimensional linear space of functions

$$L = \{c_1 \cdot x_1(t) + \ldots + c_n \cdot x_n(t) : c_i \in \mathbb{R}\}.$$

- **Theorem:** *Every error component is a linear combination of the functions*

$$x(t) = \sin(\omega \cdot t) \ \textit{and}\ x(t) = \cos(\omega \cdot t).$$

5. Proof of the Main Result

- Shift-invariance means that, for some $c_i(t_0)$, we have

$$x_i(t+t_0) = c_{i1}(t_0)\cdot x_1(t) + \ldots + c_{in}(t_0)\cdot x_n(t).$$

- For n different values $t = t_1, \ldots, t = t_n$, we get a system of n linear equations with n unknowns $c_{ij}(t_0)$.
- The Cramer's rule solution to linear equations is a smooth function of all the coeff. & right-hand sides.
- Since all the right-hand sides $x_i(t_j+t_0)$ and coefficients $x_i(t_j)$ are smooth, $c_{ij}(t_0)$ are also smooth.
- Differentiating w.r.t. t_0 and taking $t_0 = 0$, for $c_{ij} \stackrel{\text{def}}{=} \dot{c}_{ij}(0)$, we get

$$\dot{x}_i(t) = c_{i1}\cdot x_1(t) + \ldots + c_{in}\cdot x_n(t).$$

6. Proof (cont-d)

- *Reminder:* $\dot{x}_i(t) = c_{i1} \cdot x_1(t) + \ldots + c_{in} \cdot x_n(t)$.
- A general solution of such system of equations is a linear combination of functions
$$t^k \cdot \exp(\lambda \cdot t),\ \text{w}/k \in \mathrm{N}, k \geq 0, \lambda = a + \mathrm{i} \cdot \omega \in \mathrm{C}.$$
- Here,
$$\exp(\lambda \cdot t) = \exp(a \cdot t) \cdot \cos(\omega \cdot t) + \mathrm{i} \cdot \exp(a \cdot t) \cdot \sin(\omega \cdot t).$$
- When $a \neq 0$, we get unbounded functions for $t \to \infty$ or $t \to -\infty$.
- So, $a = 0$.
- For $k > 0$, we get unbounded t^k; so, $k = 0$.
- Thus, we indeed have a linear combination of sinusoids.

7. Practical Conclusions

- Let f be the measurements frequency (how many measurements we perform per unit time).
- When $\omega \ll f$, the values $\cos(\omega \cdot t)$ and $\sin(\omega \cdot t)$ practically do not change with time.
- Indeed, the change period is much larger than the usual observation period.
- Thus, we can identify such low-frequency components with *systematic* error component.
- When $\omega \gg f$, the phases of the values $\cos(\omega \cdot t_i)$ and $\cos(\omega \cdot t_{i+1})$ differ a lot.
- For all practical purposes, the resulting values of cosine or sine functions are independent.
- Thus, high-frequency components can be identified with *random* error component.

8. Practical Conclusions (cont-d)

- *Result:* every error component is a linear combination of $\cos(\omega \cdot t)$ and $\sin(\omega \cdot t)$.
- *Notation:* let f be the measurements frequency (how many measurements we perform per unit time).
- *Reminder:*
  - we can identify low-frequency components ($\omega \ll f$) with *systematic* error component;
  - we can identify high-frequency ones ($\omega \gg f$) with *random* error component.
- *Easy to see:* all other error components $\cos(\omega \cdot t)$ and $\sin(\omega \cdot t)$ are periodic.
- *Conclusion:* we have indeed justified to the semi-empirical 3-component model of measurement error.

9. How to Propagate Uncertainty in the Three-Component Model

- We are interested in the quantity

$$y = f(x_1(t_{11}), x_1(t_{12}), \ldots, x_2(t_{21}), x_2(t_{22}), \ldots, x_n(t_{n1}), x_n(t_{n2}), \ldots).$$

- Instead of the actual values $x_i(t_{ij})$, we only know the measurement results $\widetilde{x}_i(t_{ij}) = x_i(t_{ij}) + \Delta x_i(t_{ij})$.
- Measurement errors are usually small, so terms quadratic (and higher) in $\Delta x_i(t_{ij})$ can be safely ignored.
- For example, if the measurement error is 10%, its square is 1% which is much much smaller than 10%.
- If the measurement error is 1%, its square is 0.01% which is much much smaller than 1%.
- Thus, we can safely linearize the dependence of Δy on $\Delta x_i(t_{ij})$.

10. How to Propagate Uncertainty (cont-d)

- *Reminder:* we can safely linearize the dependence of Δy on $\Delta x_i(t_{ij})$, so

$$\Delta y = \sum_i \sum_j C_{ij} \cdot \Delta x_i(t_{ij}), \text{ with } C_{ij} \stackrel{\text{def}}{=} \frac{\partial y}{\partial x_i(t_{ij})}.$$

- In general, $\Delta x_i(t_{ij}) = s_i + r_{ij} + \sum_\ell A_{\ell i} \cdot \cos(\omega_\ell \cdot t_{ij} + \varphi_{\ell i})$.
- Due to linearity, we have $\Delta y = \Delta y_s + \Delta y_r + \sum_\ell \Delta y_{p\ell}$, where

$$\Delta y_s = \sum_i \sum_j C_{ij} \cdot s_i; \quad \Delta y_r = \sum_i \sum_j C_{ij} \cdot r_{ij};$$

$$\Delta y_{p\ell} = \sum_i \sum_j C_{ij} \cdot A_{\ell i} \cdot \cos(\omega_\ell \cdot t_{ij} + \varphi_{\ell i}).$$

- *We know:* how to compute Δy_s and Δy_r.
- *What is needed:* propagation of the periodic component.

11. Propagating Periodic Component: Analysis

- *Reminder:* for each component, we have
$$\Delta y_{p\ell} = \sum_i \sum_j C_{ij} \cdot A_{\ell i} \cdot \cos(\omega_\ell \cdot t_{ij} + \varphi_{\ell i}).$$
- It is reasonable to assume that different phrases $\varphi_{\ell i}$ are independent (and uniformly distributed).
- Thus, by the Central Limit Theorem, the distribution of $\Delta y_{p\ell}$ is close to normal, with 0 mean.
- The variance of $\Delta y_{p\ell}$ is $\dfrac{1}{2} \cdot \sum\limits_i A^2_{\ell i} \cdot (K^2_{ci} + K^2_{si})$.
- Each amplitude $A_{\ell i}$ can take any value from 0 to the known bound $P_{\ell i}$.
- Thus, the variance is bounded by $\dfrac{1}{2} \cdot \sum\limits_i P^2_{\ell i} \cdot (K^2_{ci} + K^2_{si})$.
- So, we arrive at the following algorithm.

12. Propagating Periodic-Induced Component: Algorithm

- First, we apply the algorithm f to the measurement results $\widetilde{x}_i(t_{ij})$ and get the estimate $\widetilde{y}$.
- Then, we select a small value δ and for each sensor i, we do the following:
  - take $x_i^{(ci)}(t_{ij}) = \widetilde{x}_i(t_{ij}) + \delta \cdot \cos(\omega_\ell \cdot t_{ij})$ for all moments j;
  - for other sensors $i' \neq i$, take $x_{i'}^{(ci)}(t_{i'j}) = \widetilde{x}_i(t_{i'j})$;
  - substitute the resulting values $x_{i'}^{(ci)}(t_{i'j})$ into the data processing algorithm f and get the result $y^{(ci)}$;
  - then, take $x_i^{(si)}(t_{ij}) = \widetilde{x}_i(t_{ij}) + \delta \cdot \sin(\omega_\ell \cdot t_{ij})$ for all moments j;
  - for all other $i' \neq i$, take $x_{i'}^{(si)}(t_{i'j}) = \widetilde{x}_i(t_{i'j})$;
  - substitute the resulting values $x_{i'}^{(si)}(t_{i'j})$ into the data processing algorithm f and get the result $y^{(si)}$.

13. Algorithm (cont-d)

- *Reminder:*
  - First, we apply the algorithm f to the measurement results $\widetilde{x}_i(t_{ij})$ and get the estimate $\widetilde{y}$.
  - Then, for each sensor i, we simulate cosine terms and get the results $y^{(ci)}$.
  - Third, for each sensor i, we simulate sine terms and get the results $y^{(si)}$.
- Finally, we estimate the desired bound $\sigma_{p\ell}$ on the standard deviation of $\Delta y_{p\ell}$ as

$$\sigma_{p\ell} = \sqrt{\frac{1}{2} \cdot \sum_i P_{\ell i}^2 \cdot \left(\left(\frac{y^{(ci)} - \widetilde{y}}{\delta} \right)^2 + \left(\frac{y^{(si)} - \widetilde{y}}{\delta} \right)^2 \right)}.$$