
JURGEN VINJU - CWI & TU/E

Noisy numbers

We use computers to do all of our computations.

The faster and more ubiquitous computer hardware becomes, the more we use it to compute all kinds of numerical output. But what if those computations are wrong?

Unfortunately, many computations produce inaccurate results. In two cases, the computer is the source of the increasing inaccuracy of the output numbers:

(1) The input data is sometimes inaccurate. For example, let us take daily maximum temperatures measured over the last century. These data undoubtedly contain both random errors and structural errors that we do not know about. When we use a statistical procedure to confirm a correlation, then the output of this computation could be wildly different given accidentally different numbers. Many errors are heterogenous, non-constant and non-independent. So even averaging the result of multiple measurements does not help.

(2) Computers do not compute all that accurately. Most of the computations done by computer chips are based on floating points. These floating points have a fixed set of binary digits, leading to rounding errors, and they might differ in significance. As a result, floating-point arithmetic behaves nothing like high school mathematics. Calculations are not always associative or distributive, as is the case with exact numbers. Calculations based on floating points might therefore lead to mistakes that are not immediately obvious. Practically all computers implement the same kind of standardised floating-point arithmetic, and so they all make the same mistakes.

Experts in numerical methods know exactly how to minimise cumulative floating-point errors and the effects of inaccurate inputs. However, we now have a huge programmer population who neither have the time nor the necessary training to do the same. Computations on inaccurate data with inaccurate floating points fail us silently, since they produce numbers that look plausible. What if programming languages would make all that inaccuracy explicit? That at least would make the issue visible. Yet that would come at a price. Perhaps the computation performed would be ten times slower. However, it would be very good to know exactly how noisy the numbers we build our lives around are.