

Ideas on Selection and Approval of Models for EPA Regulatory Use

by

John S. Irwin

jsirwinetal@nc.rr.com

EPA Eighth Conference on Air Quality Modeling

September 22-23, 2005

US Environmental Protection Agency, RTP, NC 27711

Where Are We Now

- All candidate models must pass a minimum level of model formulation documentation, user's guides, peer review, and availability.
- EPA has defined the Cox/Tikvart methodology. The BOOT program (developed by Chang and Hanna) implements this procedure. EPA once started to develop software for this, but it never got finalized.
- EPA defined a model evaluation procedure, and this does provide a means for ranking skill, but I do not believe it provides a means for defining whether differences in skill are significant (I may be wrong on the latter point).
- The system described above was devised and implemented by EPA and took a number of years to develop. The results are provided for public review, once EPA is satisfied that the results are sufficient for the action being undertaken (ie, replacing ISC with AERMOD).

What Might Be Changed?

- At best, models predict the statistical properties of what is to be seen “on average”, whereas observations are individual realizations from imperfectly known ensembles.
 - We need to get this message out so that it is commonplace knowledge.
 - We need to devise evaluation methods that account for variability in the observations that models cannot ever explicitly simulate. You can fully explain the distribution of outcomes to occur when a pair of dice are rolled, but you will never predict the exact sequence of outcomes.
- Test procedures should provide a “relative skill” ranking and should test whether differences in skill are significant.
- We need a collection of test procedures that have been developed, tested and found useful. This will take time to develop, and while this occurs we can debate various options for developing a new system for “accepting models for certain regulatory uses.”

Something To Debate

(You likely have at least 3 to 5 years for this debate)

- 1) The minimum acceptable documentation for an acceptable model is ...(contents of user's guides, science formulation reports, sensitivity test reports).
- 2) The minimum acceptable peer review for an acceptable model is ...(requirements and number of unbiased reviewers, documentation of review, response to review).
- 3) EPA will accept for use any model that has completed (1) and (2), which when tested with ASTM Standards A, B and C with data sets 1, 2, ... 10, is found to have skill scores that are better than or at least as good as AERMOD's.

Common Worries

- EPA will be forced to “accept” models that perform badly.
 - How? If a model’s science is sound (peer review) and publicly documented, and performs as well as or better than a model EPA accepts, how can we say it is performing “badly”?
- EPA will lose control.
 - If EPA does not participate in the voluntary standards development process, then yes, in a sense, EPA loses control. On the flip side, the reason I suggest moving the process out to a communal activity is to enlist help from others and to cultivate a standing committee of experts that will promote better test methods in the future.
- EPA’s regulatory process will not allow more than one model to be “acceptable” for the same task, because that would mean there is more than one “acceptable” answer, and people will “shop” for the lowest/highest impact.
 - Leadership is learning to provide answers when new facts become known. The facts are that air quality is not a deterministic science, and peak impacts must be viewed as having a degree of stochastic variability. By publicly stating that the problem contains a stochastic component and a component of science uncertainty, EPA would be laying the foundation for the next wave of advances in modeling theory, development and evaluation. This would also be laying the foundation for the next advances in air quality assessment.