Model Maturity Level (Draft started in 2007)
Robert Ferraro, NASA JPL primary author
Richard Rood, University of Michigan, revisionist

1.1.1 Model Maturity Level (Draft)

Models exist in various stages of research and development. A system model may consist of a collection of subsystem models of various levels of fidelity and maturity. A model may also span a number of levels of maturity, depending on how it is applied. For example, a ¼ degree ocean global circulation model may be in operational use in a data assimilation system, but the same model may have no validity when applied to a 10 year climate forecast, or may be untested for assimilation at 1/10 degree resolution. Therefore, when assessing a model’s maturity, the application and resolution limits must be specified as part of the assessment.

The Model Maturity Level designations should apply equally well to engineering models and science models. For models of systems (science or engineering), the maturity level of the system as a whole is assessed. The system is assessed independent of the maturity of its component models. (There may conceivable be cases where subsystem models are at a lower maturity level than the overall system assessment.)

In order to characterize the maturity state of a modeling application, the following assessment system is proposed.

**MML 5 - Model is operational in a decision support or analysis system**

Model validity and prediction skill are well characterized for the parameterizations and resolutions in use for the operational decision support or analysis system

**MML 4 - Model forecast skill is validated and measured**

The model has been tuned and tested with a range of initial conditions and boundary conditions. The prediction skill has been measured and shown to be valid within the range of initial and boundary conditions for the intended resolution and parameterizations. Non-physical artifacts no longer exist, or exist only outside the operating range of validated model behavior.

**MML 3 - Model reproduces observable phenomena**

Using actual initial conditions and boundary conditions, the model generates the target phenomena. The modeled phenomena compare favorably to actual observations. Non-physical artifacts may still exist that do not affect the behavior of the target phenomena.

**MML 2 - Model is implemented and demonstrated to be stable**
The model has been fully implemented. Test cases have been completed which demonstrate that representative instances of the target phenomenon are reproduced by the model and that the model is stable over the expected initial conditions and boundary conditions.

**MML 1 - Conceptual design for a new model is complete**

The target phenomenon to be modeled is identified. The resolution requirements and parameterizations are defined, and the target prediction skill and limitations are specified. Any new algorithms to be incorporated into the model have been implemented and tested.

1.1.2 Some definitions and comments for discussion

‘Prediction’ is the evolution of a model beyond its initial conditions. This may be forward or backward in time or space.

‘Prediction Skill’ is the measure of the deviation of a prediction from the actual evolution of an observed phenomenon. It is typically a function of distance (time) beyond the initial model conditions. Often represented as a percent error or percent confidence. E.g., hurricane track error at

'Forecasting' is integrating a model FORWARD in time, starting today or so. E.g., weather forecasting.

'Hindcasting' is integrating a model FORWARD IN TIME, but starting at some time in the past for a period of time that is entirely in the past. It is used to validate a model (compare to observations) or to fill in gaps in observational data.

'Backward Integration' is integrating a model BACKWARD in time. E.g., to find where an air parcel or water parcel came from.
Comments received

It's pretty clear where one can fit in from what I can tell. I think the note that a model can exist at various levels based on application is exactly right, but we will have to see how comfortable people feel about that classification. If it is argued, I'd say it's all part of the "maturity" aspect. One might feel more comfortable knowing that when a model is created it sits at a level regardless of resolution and boundary conditions, but I think that view is not accurate.

I think the jump from one level of the hierarchy to another is not linear, but that is okay. Generally it seems tougher to go from one level of the scale up to the next, but that seems appropriate (not unlike the Capability Maturity Model CMM stuff).

I agree with you about hindcasting (but I could be wrong). One of the things we often do in stratospheric research is "back trajectories". For example, measurements made on an aircraft, and trace back the history of that air parcel for some period of time (week or so) to see where it came from. I think some ocean models also do that. I don't know what the general term is. backcasting? reversecasting? Or what it implies for the maturity of a model.

Actually,

I reviewed these definitions and find them reasonable. The only thing that I might add is that there is always a continually feedback loop between stages 2,3, and 4 (i.e., new validation and skill improvements are ongoing). Every few years these improvements propagate to the operational MML-5 (e.g., improvements to NWP fields). The point here is that the model is never "done." (I guess you can say that about all software).