CEN/TC 264/WG 43

**Specification of quality objectives**

**for the evaluation of air quality modelling**

**used to supplement fixed measurements**

**for assessment of ambient air quality**

**under the <u>European Air Quality Directive 2008/50/EC</u>.**

# Some notes…revisited @ Tallinn 26/6/2018

**Ari Karppinen**
**Originally @ 17/4/2018 Barcelona**
**CEN TC264-WG43**

# Single module(??) vs. **modelling system**



Modelling systems - FMI
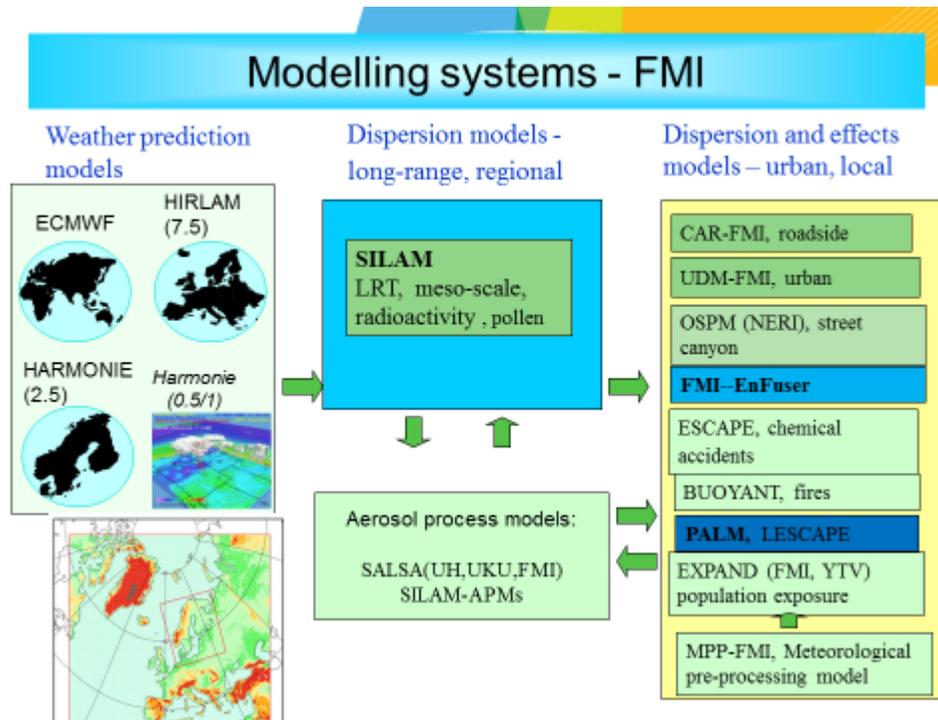
- **Global->regional->local/urban -> street canyon (->CFD/LES (1m scale)**
- **There is not really any question of applicability of a specific application ?!**
- **Evaluation is (now) done for any domain the system is used:**

  **can be : global/European/Finnish/city scale..whatever …normally the number of measurements for the evaluation is not an issue**

**but**

some/all of the model are , naturally, also used for very different purposes; like SILAM ( in the CAMS-ensemble) for AQ-forecasts for Europe  ->

Evaluation is done against background measurements only / the concept of what the measurements stations represent is now important for the real model evaluation (these runs DO NOT try to assess the AQ for "directive" purposes)

? If SILAM-model results alone would be used for assessment in some limited area  - should we strictly forbid (somehow??) that kind of "miss-use" of the model -

or

should our statistical evaluation measures be simply strict/good enough to tell, that SILAM is not good-enough for that specific assessment ?

..and we are not really aiming at giving out a <u>general model evaluation standard</u> .. or are we ??

Answ1: FAIRMODE promised to work on that

+

Answ2: There is also a new initiative for a CEN group which would aim at "real" model standard.

# Black box (or "too simple") models…

**Quite common example:**

- **"Simple (or complicated)  interpolation of measurements"**

    - <span style="color:red">Obviously creates "some" problems to assess with e.g. "leave 1-out" approach..</span> but..  **is this a real issue** ?

    -> if this "model" passes e.g. "leave-1-out" evaluation , doesn't it actually tell us : <span style="color:orange">A: something about the simplicity of our domain's concentration fields ?</span>

    <span style="color:orange">B: dense enough AQ_measurement network</span>

    or ..<span style="color:red">C: lack of stringency in our statistical indicators ??..</span>

    or <span style="color:green">D: something else ??</span>

# Data assimilation..or not ?

**A "typical" (well documented/acceptable ?) assessment model**

- Uses background measurements combined with local/urban scale models

- **Depending on the location/application, the measured background can explain 90% of the concentration variations (pm2.5 a typical example in many  urban locations)**

- **<span style="color:red">Does the direct evaluation make any sense ?..</span> -> you  should be forced to evaluate only the local contribution ( =measured - background)**

# ..  the (un)normality of  distributions

*...Joost already collected a nice set of exampels showing some real distributions..(some quite normal, some definetively not)*

- It can be (easily) stated/proved, that **if normality is not a correct assumption, then e.g.   correlation coefficient is not a correct statistical measure to evaluate the performance**..

  but..then .. the relation of this fact to our metrics ?:

**A:** we need still to define better measures (not depending on the normality assumption at all ) ?

B: we need to do extensive testing on (all) available data to assess what is the real effect of this issue on  our statistical indicators (my vote for this!)

# …the domains…target areas..etc..
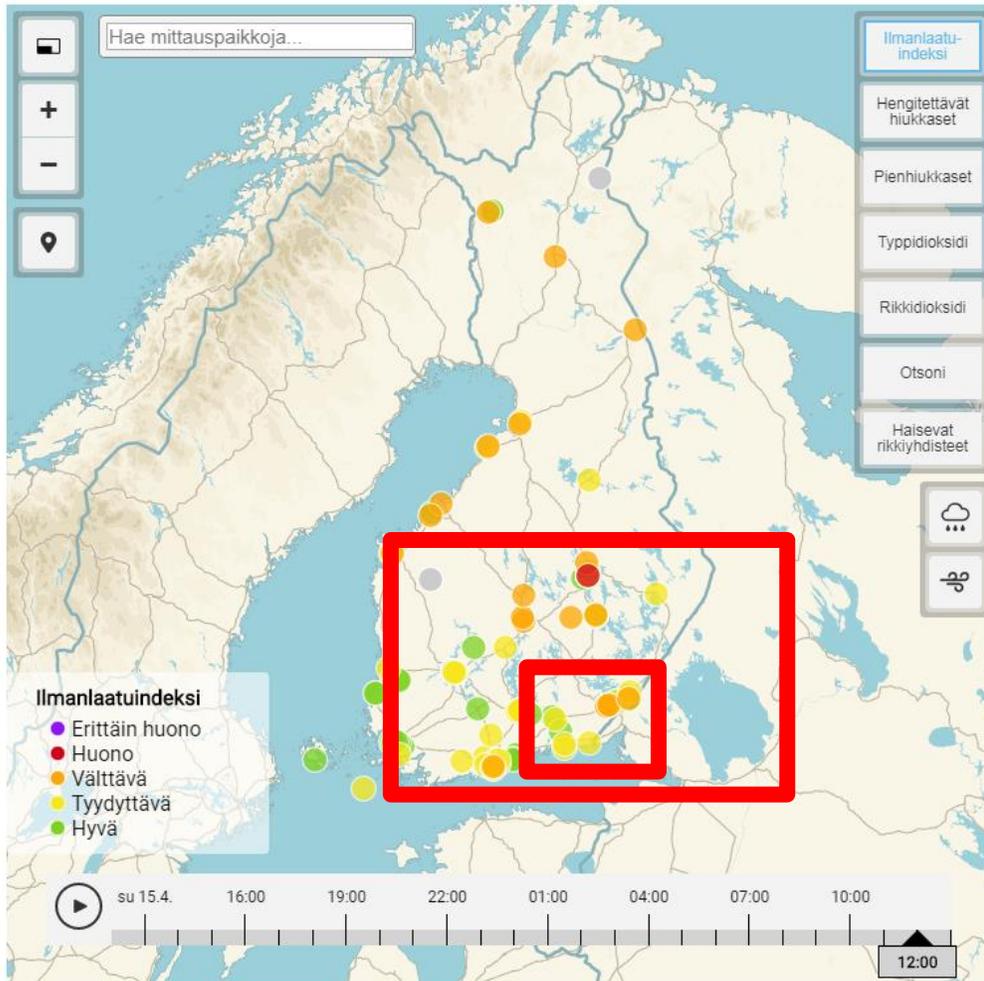


**We evaluate FMI modelling system (hopefully with "green light") for whole Finland for some specific year using ALL available measurements ->**

⇒ **I should be able to claim,that the model system is valid also for all smaller domains.. or not ?!**

# …the domains…target areas..etc..
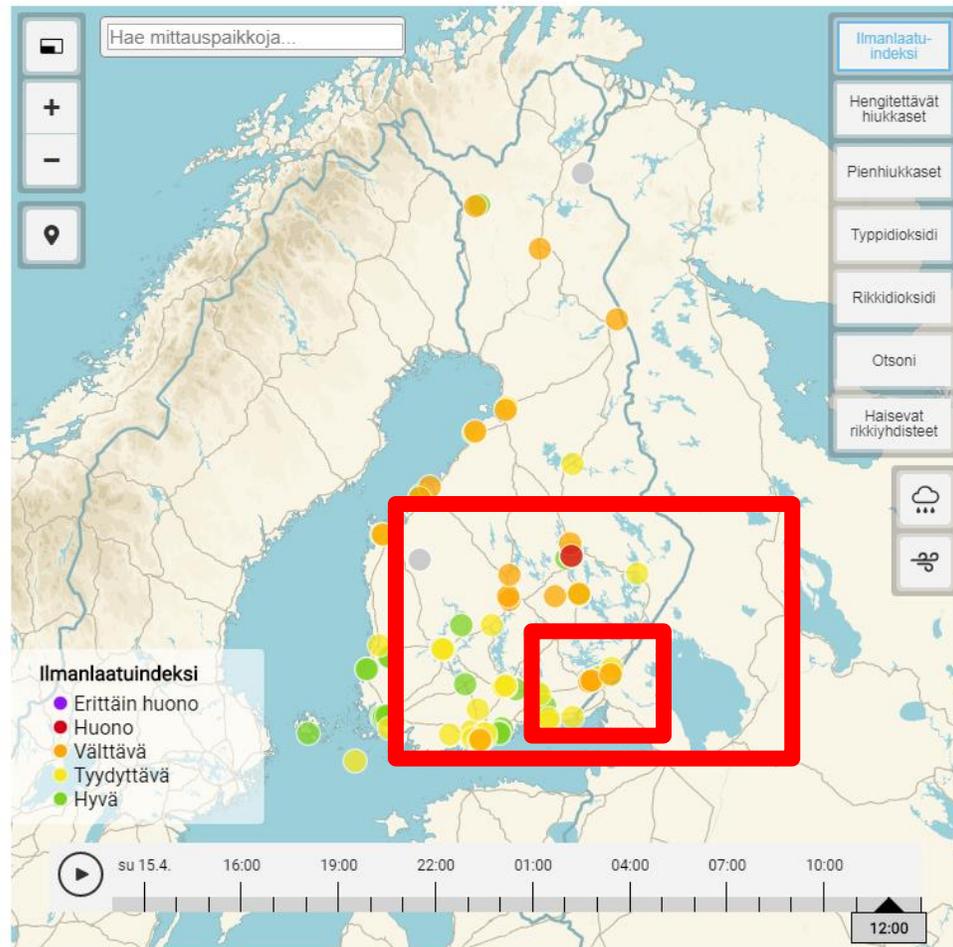


**..is there any (simple) acceptable way of "proving" that model which is acceptable for the whole domain /all stations, would be acceptable also for all subdomains ?**

- eg. 130 stations for whole domain-rule: "calculate also the stats for all 5 station combinations".. only 286243776 combinations to be checked ->If all (or some % ??) those checks pass.. then you can use your model to any subdomain with 5 or more stations.. or can you ? ;)

# the domains...target areas..etc..
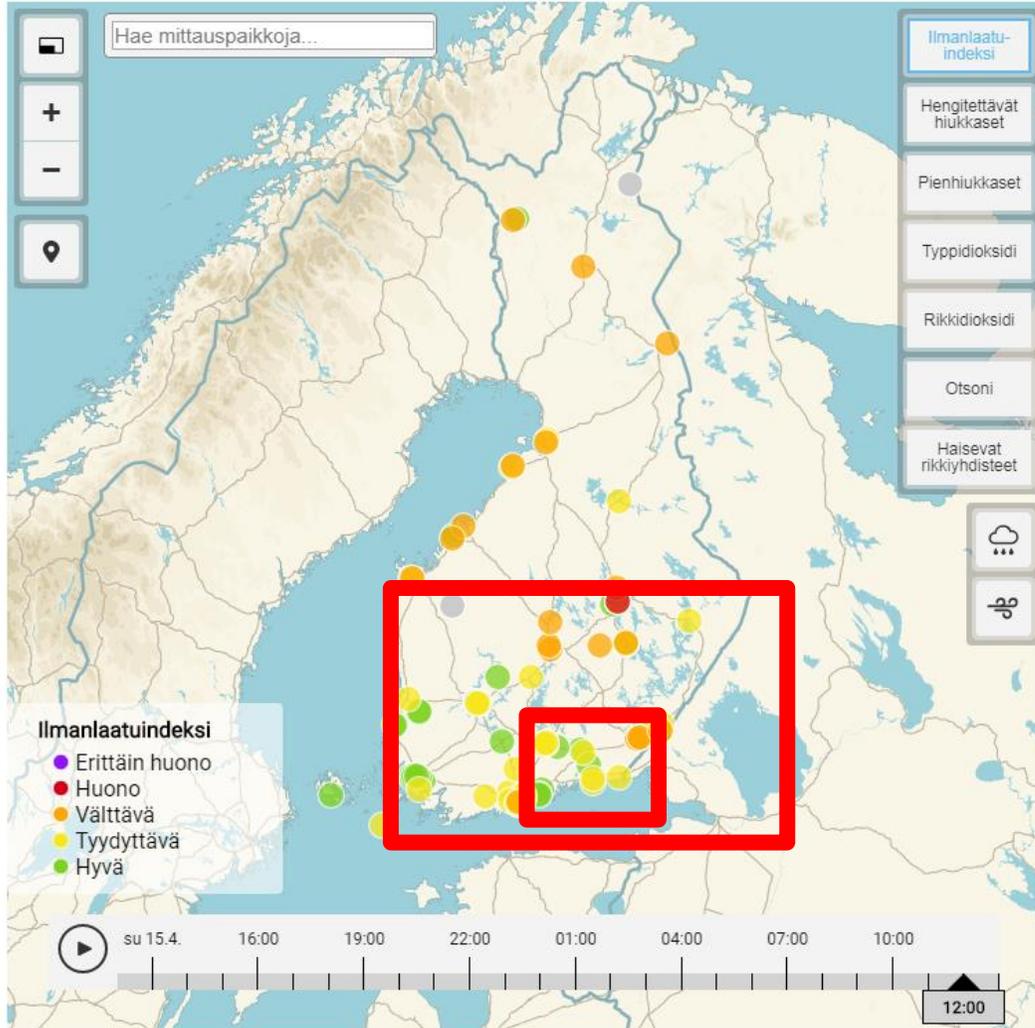


if there is a way :

that would solve practically most of our "# of stations" /"representativity"..etc.. problems

ust evaluate your system in a "big-enough" domain which has "enough" mesurements available.. (ok..slightly simplified..)

Ps. This would also give a clear indication on the type of modelling system you can really use for the assessment

# ..the final slide

- **Project description for supporting work needed for the standard prepartions:**
  - Final (language/detail checks) pending

  - Call for the project(s) will be hopefully  open this autumn!
    - 6 worpackages/ each 40-80 k€
    - Open bidding
  - Next WG meeting will concentrate (again) on revising the Tehnical Document   <= DISCUSSIONS/COMMENTS related to the here might have an effect for these revisions