Yes, that's been discussed a lot. Short answer seems to be it works pretty well this way, and they haven't found a better way. (So if you're thinking "well that can't work", well ... it does work.) So the question is what might work better without untenable drawbacks?

Any score-based metric kind of takes away our ability to switch/change models at will – remember they have no idea how our predictions are derived, and today's model may not be the same as yesterday. Also anything with scores requires a long history. I suppose they could take historical (staked) performance of the person as evidentiary weight of their competence (which still requires history, but disregards changes to models). So long-participating participants could have some additional internal tweak to their weights, and newbies could be downgraded somewhat or something like that. But those adjustments would be minor – if you based everything on historical performance there would be several major problems with that (attack vector, metamodel cannot adjust quickly, etc). The main mitigation factor to possibly too much influence by a small number of stakers is simply to have as many participants as possible. Truth is there is no way to automatically derive optimal weights, but if you've got something that is working well (and it is), better to think twice before making major changes. Not that they've been afraid of making changes – they did add TC scoring (which was unproven and quite different) to simple corr scoring and made that more important – and now corr scoring has just changed also. So stakers are required to react to these changes.

So yeah, it is tempting to think there was got to be a better (and more complex!) way to do it, but once you try to figure out exactly how what that something would actually work, it becomes quite tricky and reveals its own problems or constraints (they don't want to end up like every other hedge fund – that'd be missing the whole point). Something simple that works very well is hard to beat...