TBD*

Ramyaa[1], Kallol Das[2], Suresh Marru

This document is just bulletted text to give an overview
Intro:

- ML has been used successfully in many areas. In sceince the application has mainly been to predict or forecast values that are hard to predict based on scientific models.
- However, ML models can be blackboxes, and even human-understandable models like decisiont trees are treated as black boxes, given the preprocessing and model size
- One consequence of this is that though prediction of a quantity is possible, this does not advance the basic science. For instance, a neural network that predicts rainfall using meteorological features such as temperature and humidity may perform better than models based on atmospherical science, but it does not advance the science.
- Here we report the preliminary work done to use machine leanring to advance atmospheric science - in particular rainfall prediction.

The full project (includes context and future work)

- Existing rainfall models are differenetial equaions based on physics of weather. The models integrate these equations to simulate the weather. Models differ in spatial and temporal resolution, the different atmosphetrical features included, values of parameters etc.
- As all modeling, these models could benefit from increase in details, especially interdependence of features, and different models may differ in the details that would improve them.
- These differences make models better or worse at predicting rainfall, and this may depend on many factors such as geography, time of day, other weather data such as temperature, humidity etc.
- This work is the first step in figuring out the details to improve the models.
-
- We analyze 24 models of rainfall to figure out these details
- Since we cannot guess which details to add and where without a domains expert's invovlement, we begin by exploiting the variations within the models.
- Simply put, we form an aggregate of these models that incorporate models to varying degrees based on different factors
- One way to do this is to create a machine learning model with independent variables being factors that affect the performance of the models, and dependent variable beiing rainfall
- However, if such a machine leanring model were to include in its independent variables meteorological data, there is a risk of simply creating yet another model instead of aggregating existing models, which would not help advance meteorological modeling.
- So, we start by creating aggregate models with non meteorological variables that act as meta variables - here we focus on spatial location.
-
- For every geographical location, we use several Ml algorithms to create an aggregate model (that only uses as independent variables the outputs of the meteorological models).
- ??? we are trying to find patches of spatial locations where some models are used more by the aggregate models. This helps the domain expert to analyse the comanalities in these models that are not present in others, and try to incorporate them in other models.
- RESULT so far: ares where the existing models perform well, the aggregate model improves on it considerably; the areas where the existing models perform poorly, the affregate models are worse
- there are patches of locations where this happens

*This work was supported by Oklahoma University

[1]Ramyaa is with Faculty of Computer Science, New Mexico Tech, 801 Leroy Pl, New Mexico, USA `ramyaa.ramyaa at gmail.com`

[2]Kallol Das is with the Department of Computer Science, New Mexico Tech, 801 Leroy Pl, New Mexico, USA `kalloldash at gmail.com`

APPENDIX

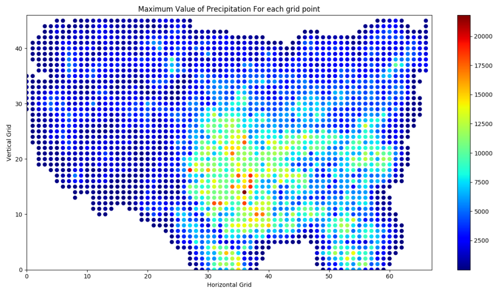

Fig. 1. Maximum Rainfall (1.a.i.)

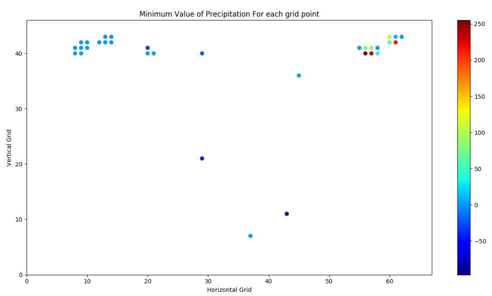

Fig. 2. Minimum Rainfall

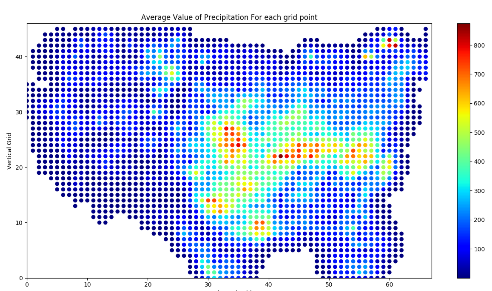

Fig. 3. Average Rainfall (1.a.ii.)

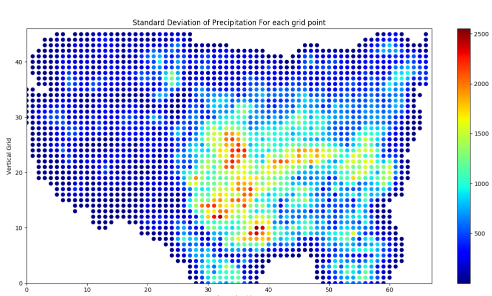

Fig. 4. Standard Deviation (1.a.iii.)

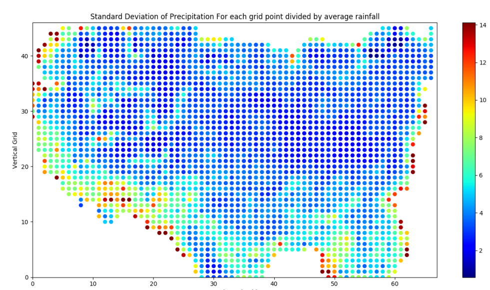

Fig. 5. Standard Deviation divided by average (1.a.iv.)

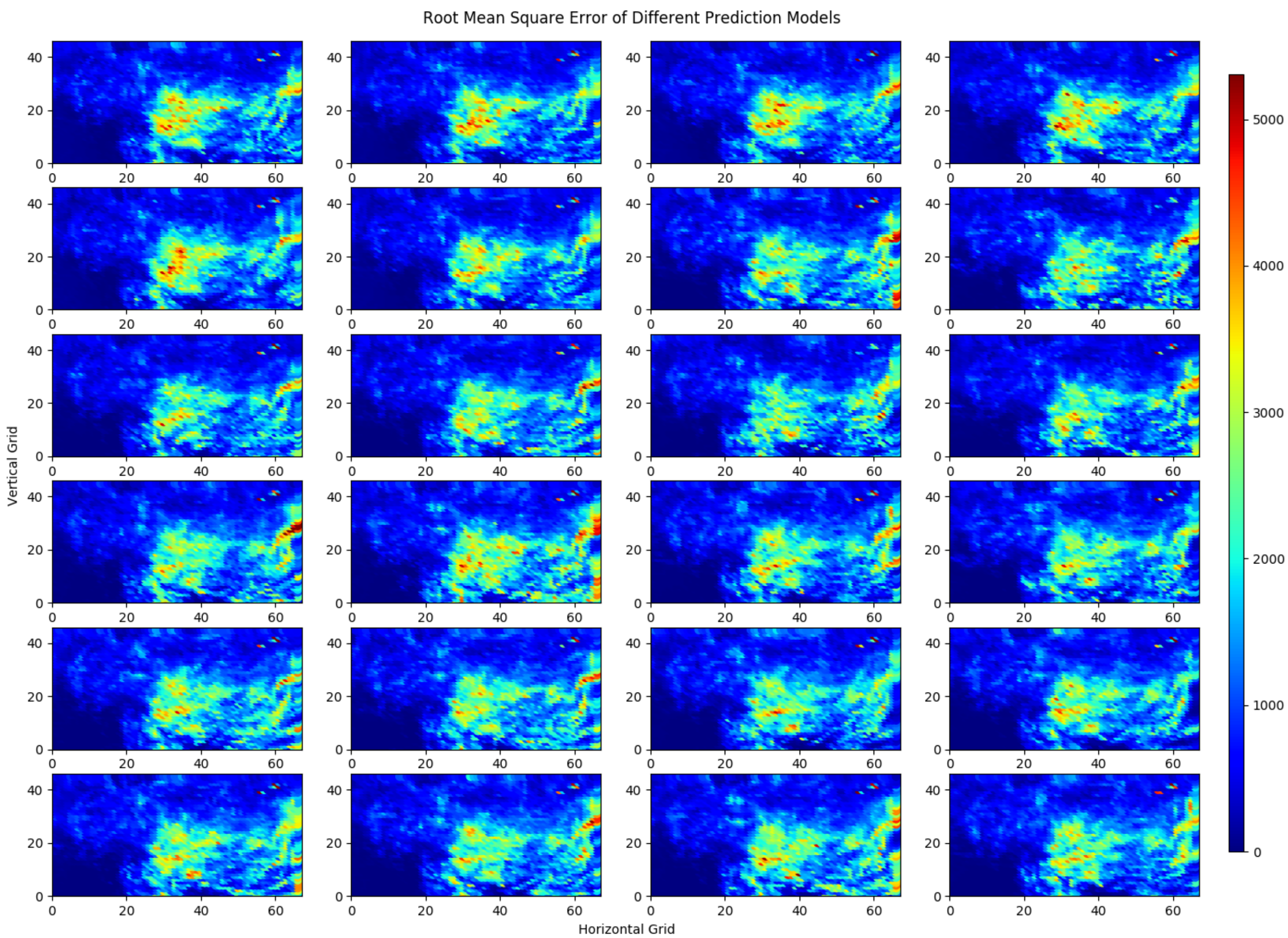

Fig. 6. Errors of Different Prediction Models (1.b.i.)

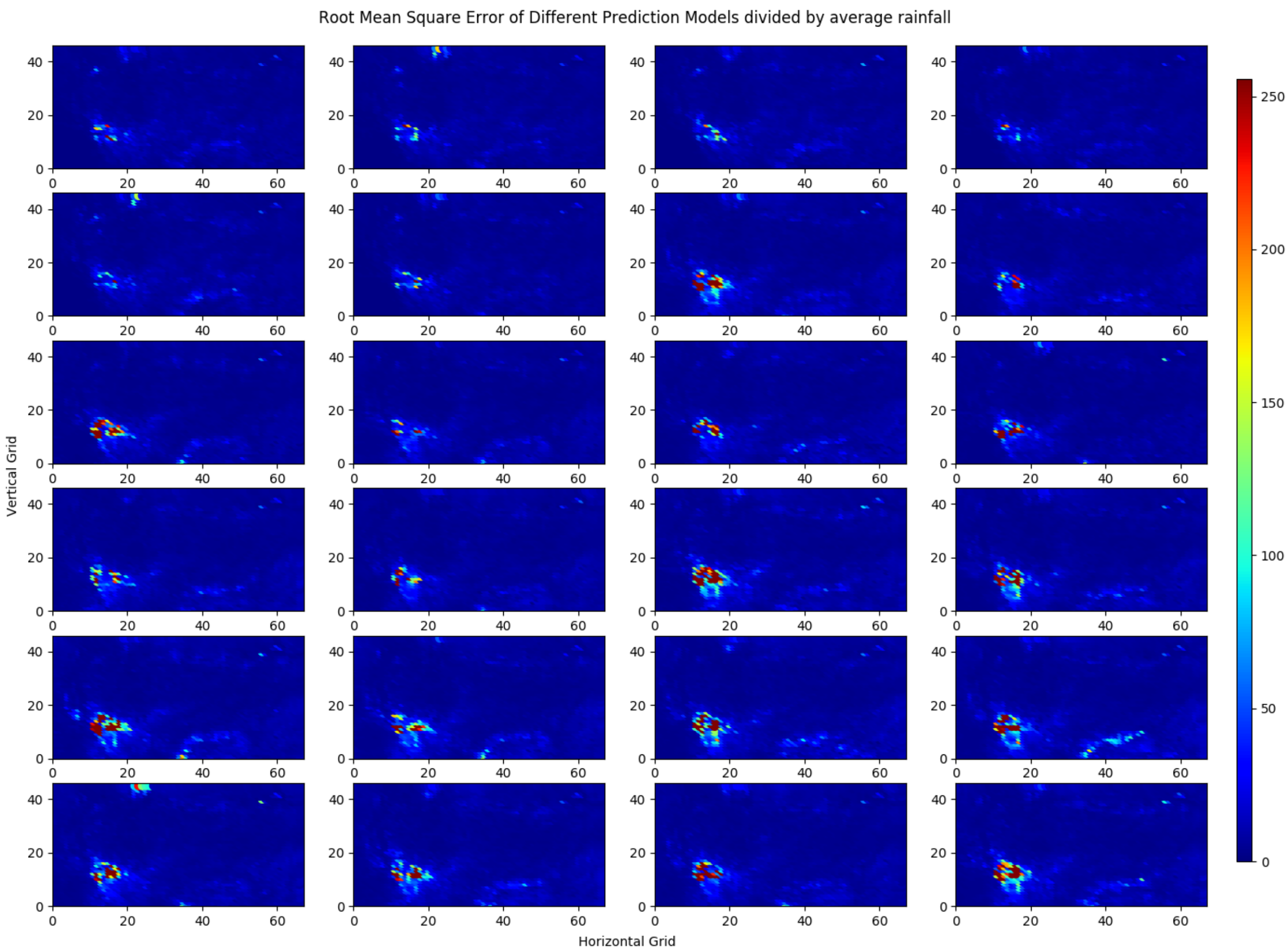

Fig. 7. Errors of Different Prediction Models divided by average (1.b.ii.)

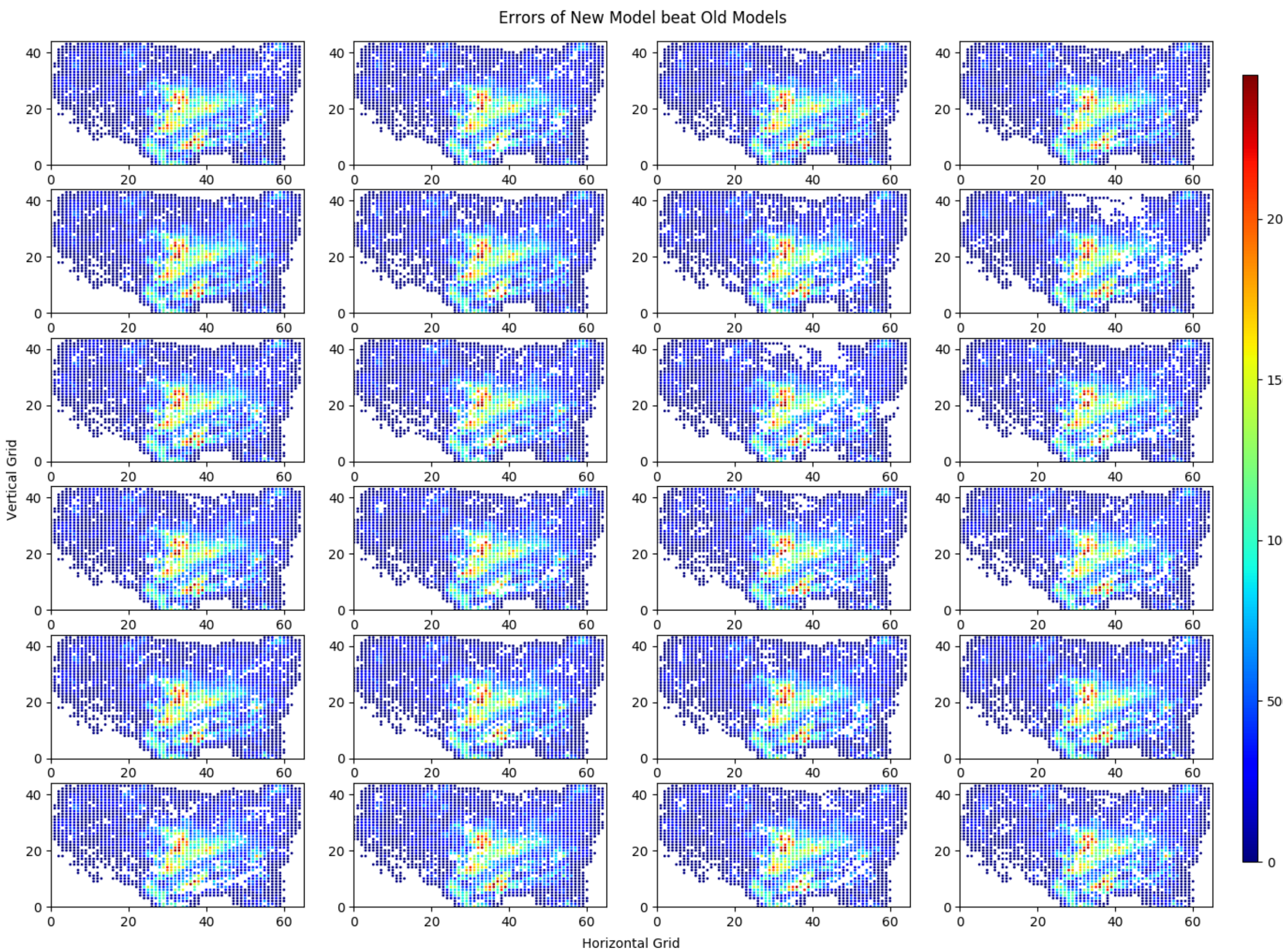

Fig. 8. Error For each model (new is best) (2.a.i.)

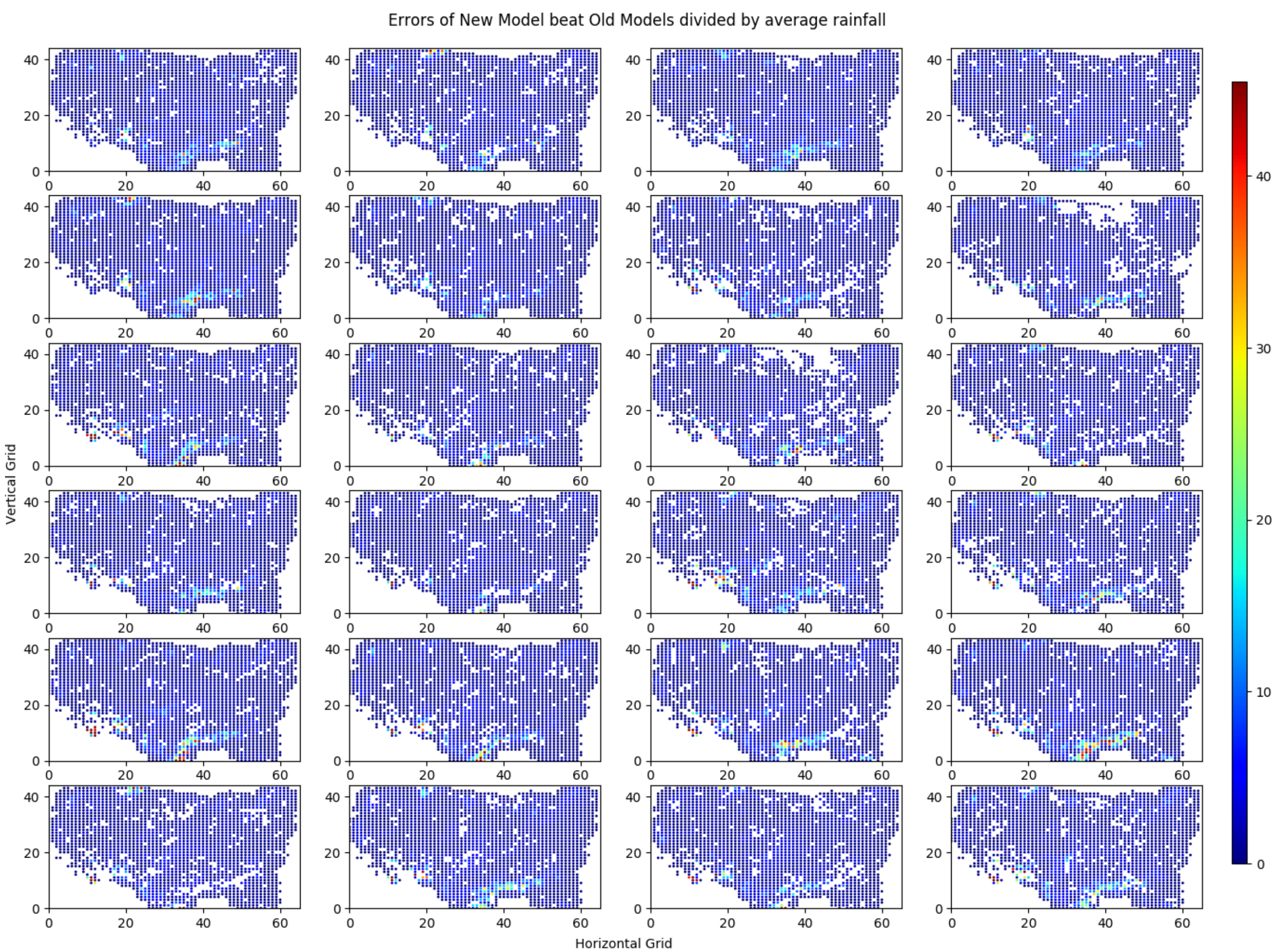

Fig. 9. Error For each model divided by average rainfall (new is best) (2.a.ii.)

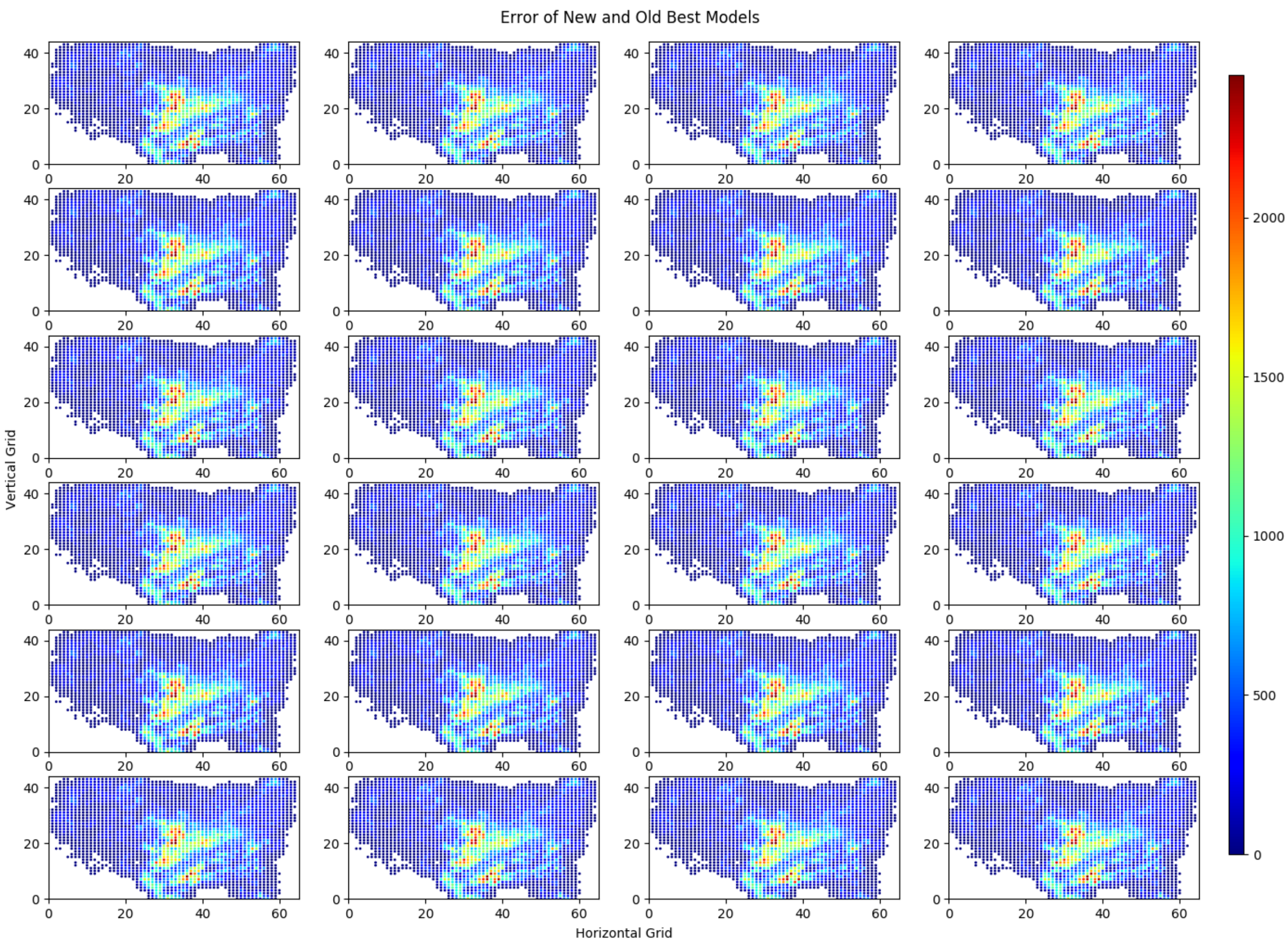

Fig. 10. Error For each model (new or old model be best) (2.b.i.)

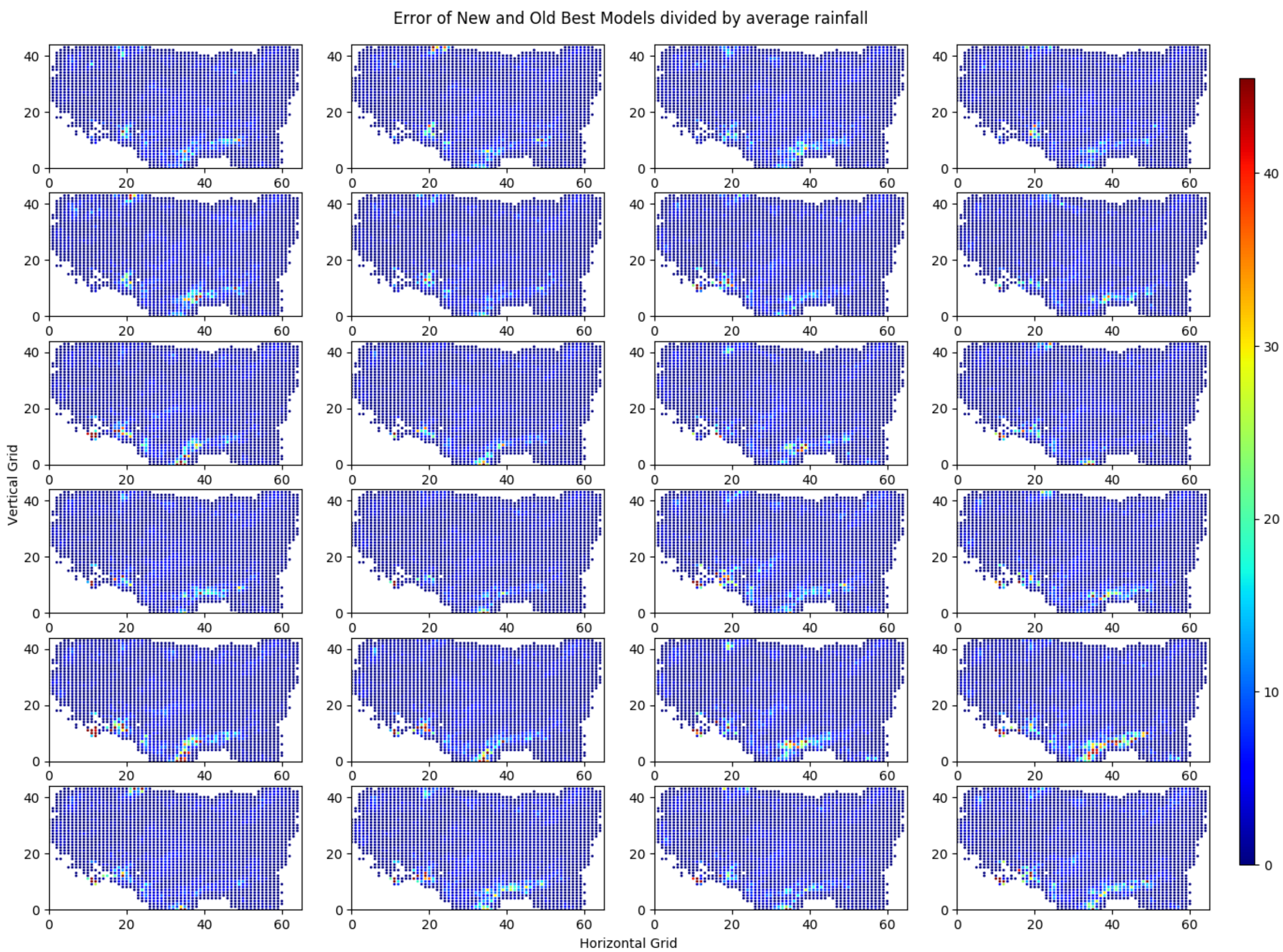

Fig. 11. Error For each model divided by average (new or old model be best) (2.b.ii.)

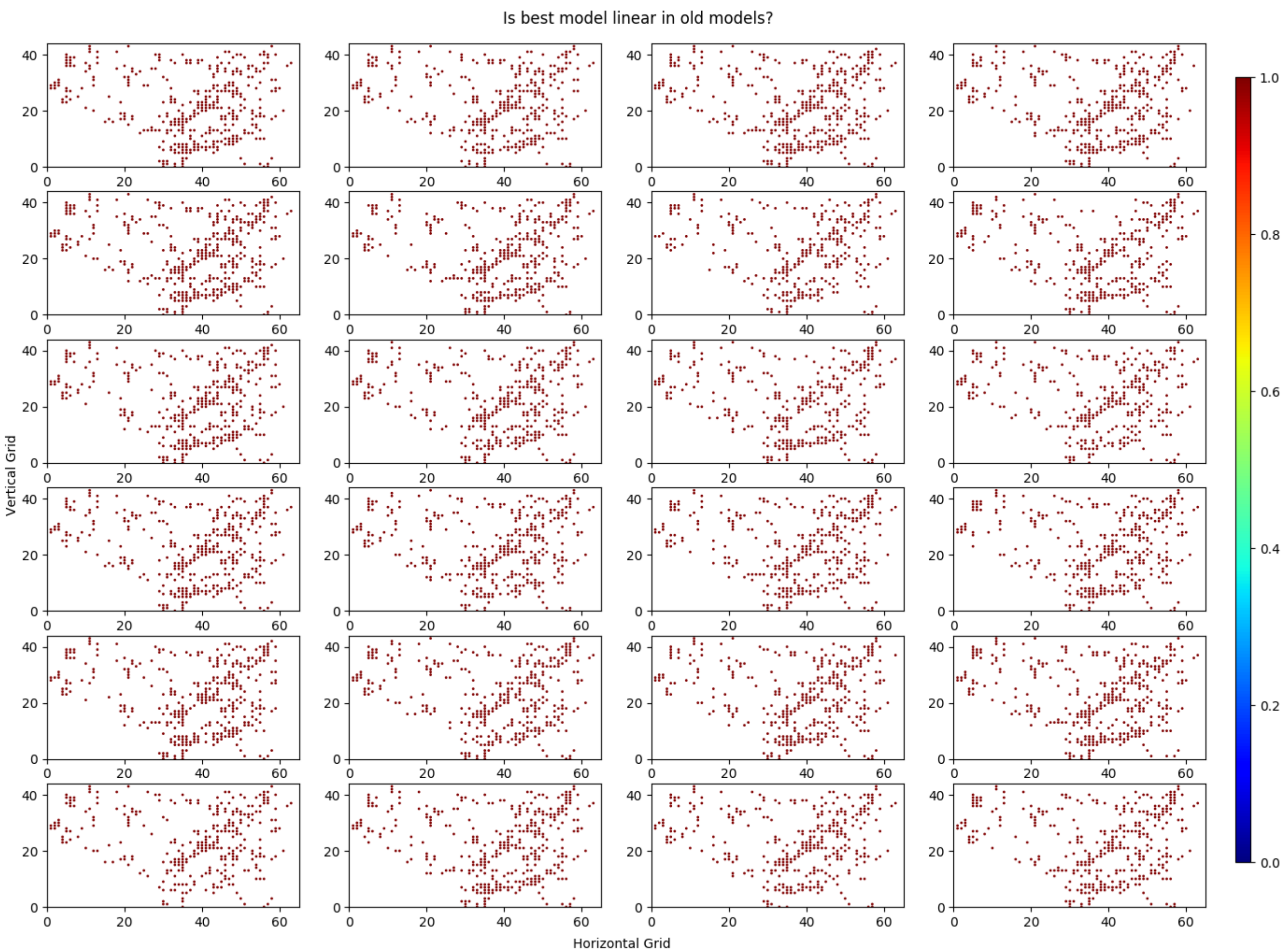

Fig. 12. Plotting the best models which are linear in old models (2.c.)

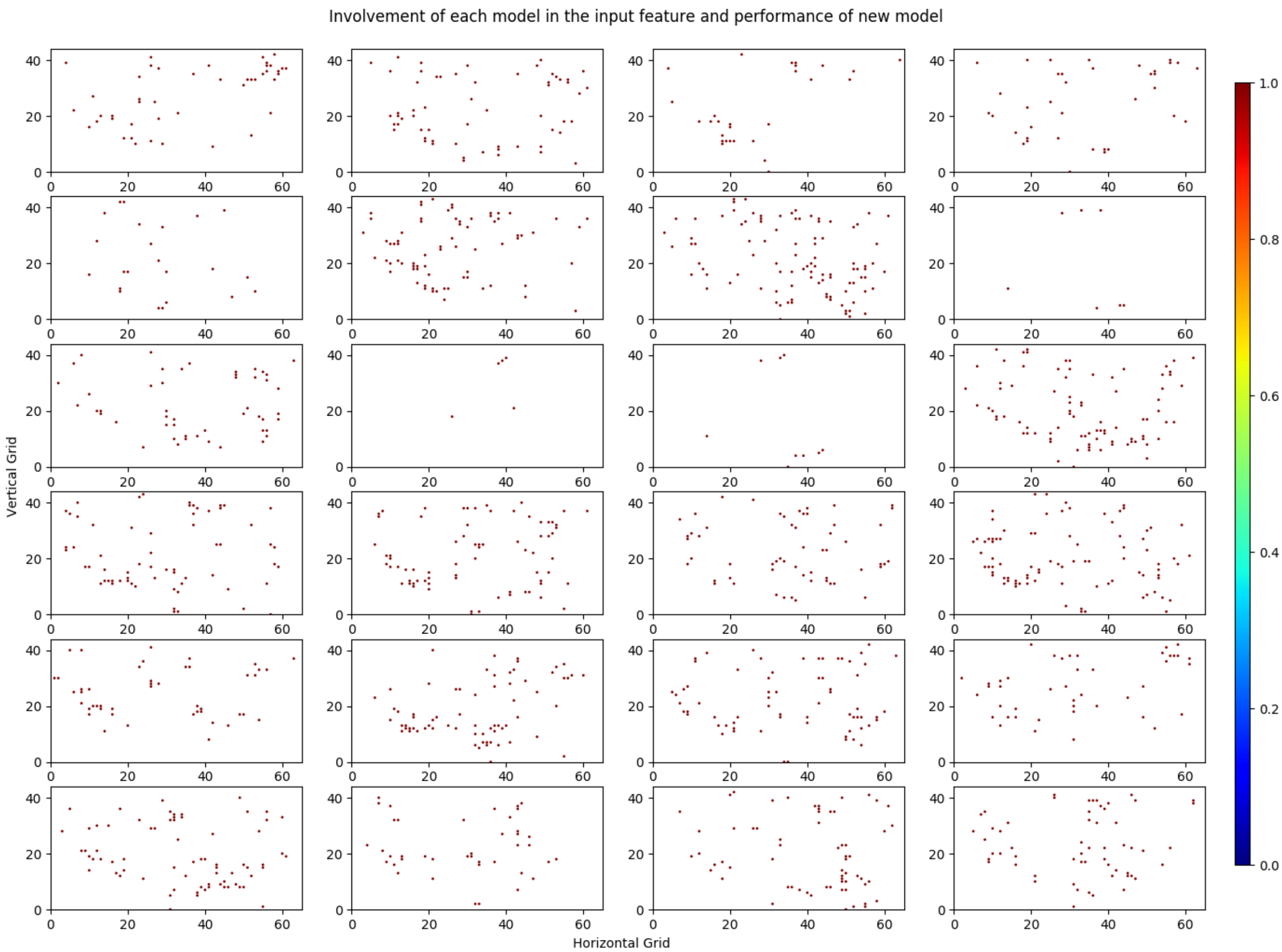

Fig. 13. Involvement of each model in the input feature and performance of new model (3.a.)