Which is better between Model-based RL and Dyna

In RL based Model, the general way to generate a policy is Environmental interaction —— Model learning —— Direct planning —— Greedification. The detail is to build a simulation model, and let it exploring. After the simulation on model, it shall generate a experience in order to train the model. Then we have to choose to implement a distribution model or sample model based on different problem. After then model learning, there should be a simulated experience updates the value and to generate the policy, this process which called planning.

However, the path to a policy in Dyna has different workflow. It is from Environmental interaction —— Model learning —— Simulation —— Direct RL methods —— Greedification. The model learn from experience and simulate another “real experience”, then it apply direct RL methods to get a value function. The implementation of direct RL is to directly improve the value function and policy, and it is not going to affected by bad models. Since that, the Dyna-Q algorithm is more better than RL based model algorithm because it is simpler and not affected by bad models.