* 前回

富士河口湖のデータを使って川崎浮島発電所の発電量を予測しようとした

発電量やAMeDASのデータを30分単位に整形して学習を行っていた

AMeDASのデータについて30分単位に変更する際meanで集計していた

見かけ上学習は進むのだが正規化のリスケールをすると値の範囲が不適切になる

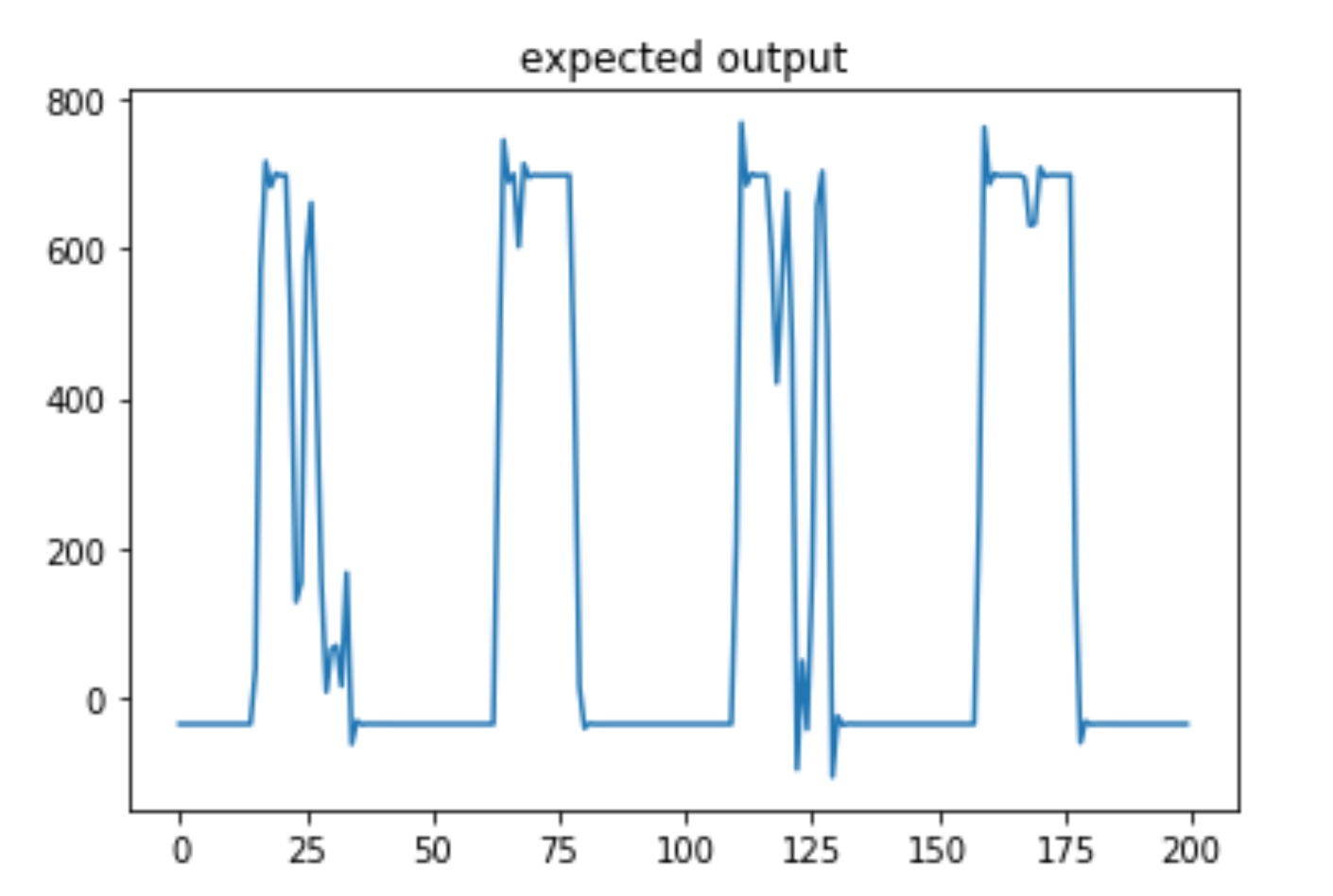
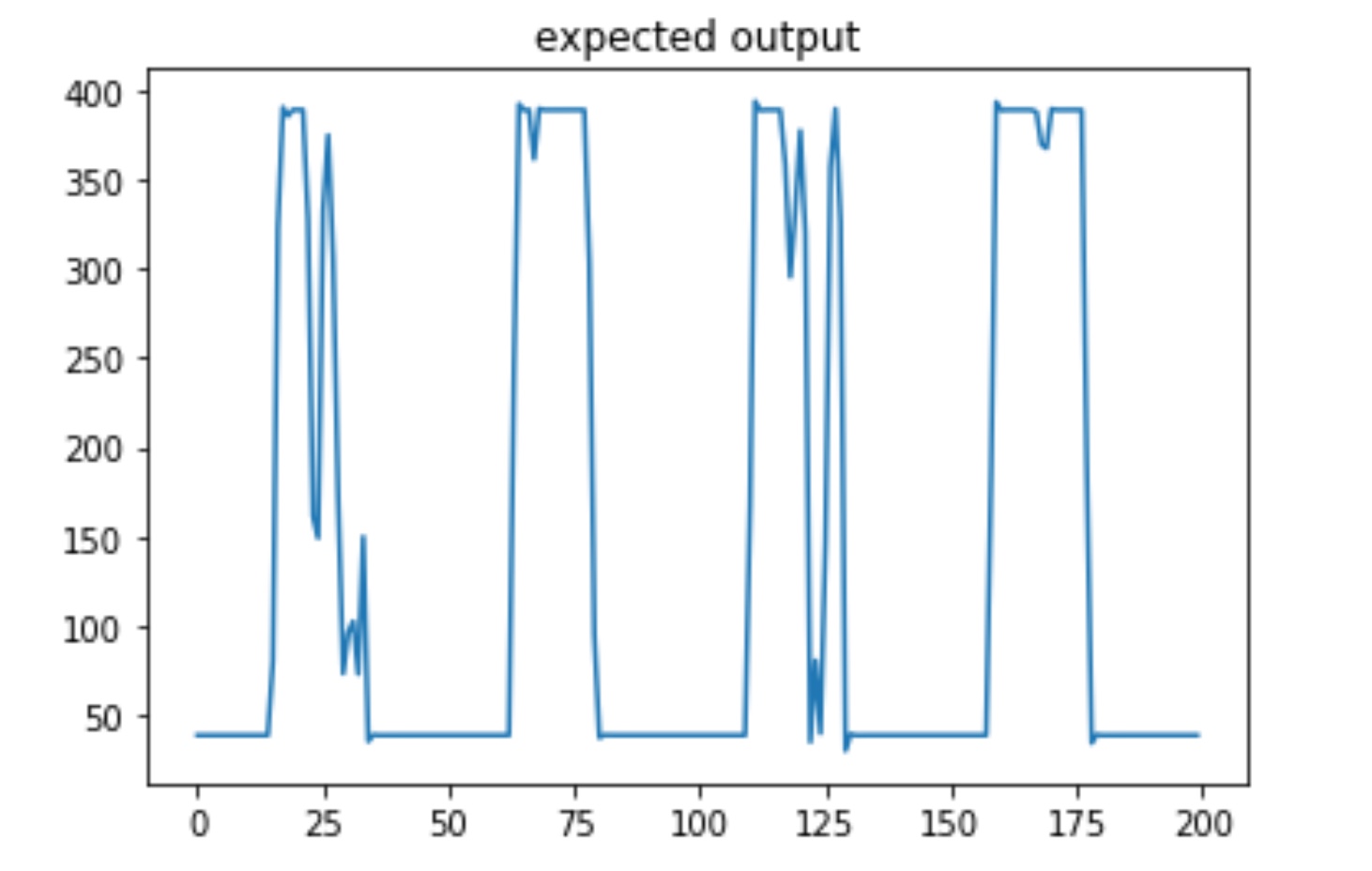
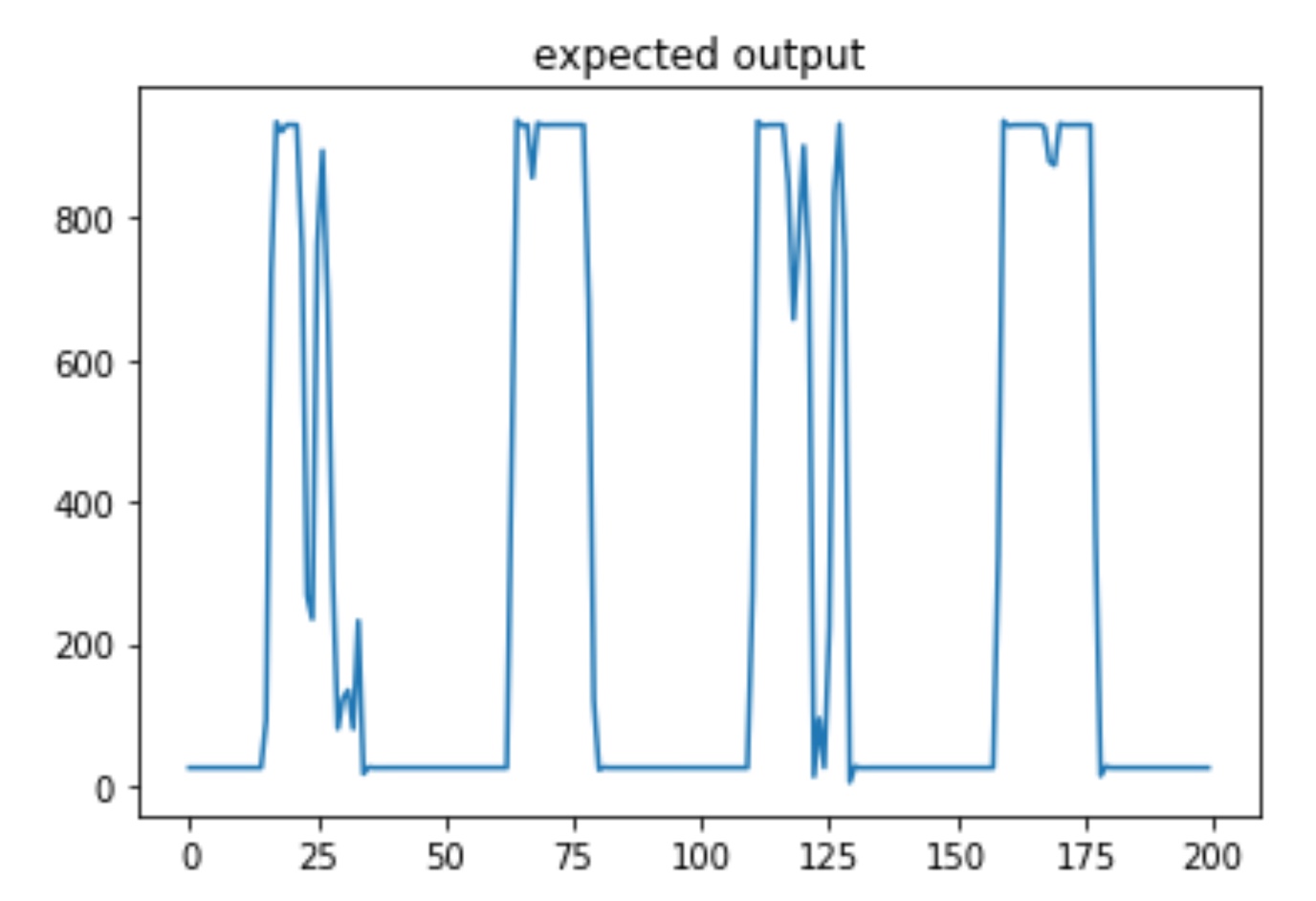
* 変更点

発電量やAMeDASのデータを10分単位のまま学習を行うことにした。それに伴って読み込まされる時系列の長さを49から144に変更(学習に時間がかかるようになった)

LSTMの入力を[sl(日射時間), max\_tp(10分あたりの最高気温)]に変更

* 進捗

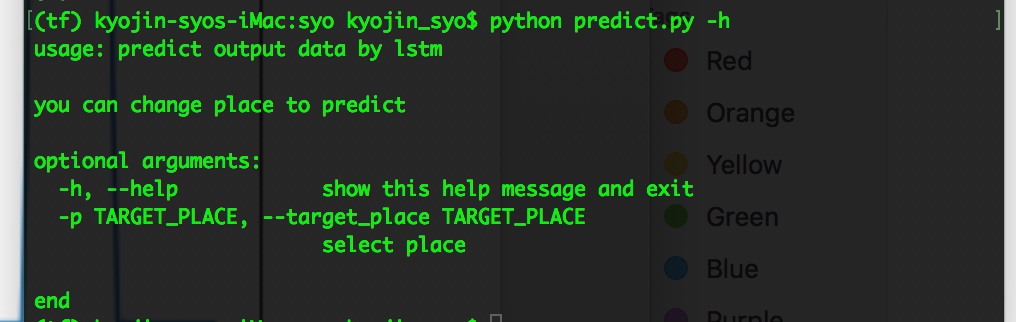
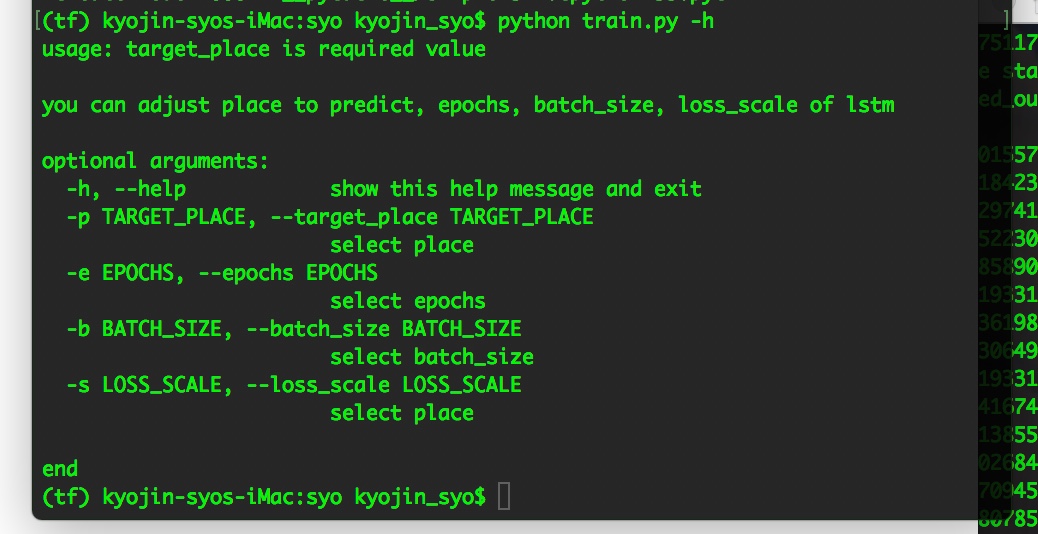
lossを元のスケールに戻して計算するとどうなるか(左からx10, x100, x1000)



いずれも0~3000の元のスケールにはあってない(実験時は河口湖のデータを使ってデータも30分単位で学習予測させていた)

モデルを変更した

Azureで学習を回すためのpythonファイルを作成



batch\_sizeのパラメーター調整を行った(総学習データ210k)

batch\_sizeを大きくすると学習が不安定になった(1000とかにしたらだめだった)

batch\_sizeは(100 ~ 200)くらいがよさそう

標準化と正規化について、今回は発電量の最大最小が決まっているのでmin-max正規化を使う。

ValidationのMAEのスコアが150くらいになったりはするのだがその実態は全ての値を34~36くらいで予測するという

浮島発電所と扇島発電所の相関係数が高かったやつについて

二つの発電量のMAEを撮って見たら130くらいあったので別々のモデルで学習させる日ちようがありそう

* 疑問点

chunk\_sizeが長すぎる？

chunk\_sizeを変えると自動的に使うデータを調整できるようにしたい

そもそもモデルが不適当？

言葉だと説明しづらいので板書とかでやりたい

optimizerが不適当？Adamは学習が進むにつれてだんだん更新しなくなる

MSGDを使って見たがlossをx1000をすると発散する, x1だ-80のとこに行ってこれまた発散

Lossをどう扱うか

普通はlossを定数倍するとかはしないからx1000とかしないべきだが…

評価関数と同じくMAEでlossを計算しているけど一般的なmseでやったほうがいい？