Supporting Information

Identification of Host-Guest Systems in green TADF-based OLEDs with Energy Level Matching Based on Machine-Learning Study

Min-Hsuan, Lee*

Dr. Min-Hsuan, Lee
Rm. 312, Bldg. 11, No. 195, Sec. 4, Chung Hsing Road, Chutung, Hsinchu 310, Taiwan.
E-mail: itriA50587@itri.org.tw

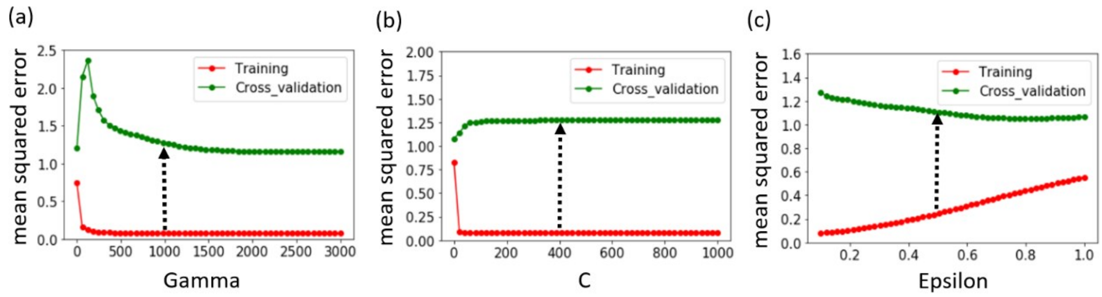

Fig S1. The learning curve of the SVR algorithm in 10-fold cross-validation analysis with different parameter (a) γ (for fixed C = 400 and ε = 0.5), (b) C (for fixed ε = 0.5 and γ = 1000), and (c) ε (for fixed γ = 1000 and C = 400)

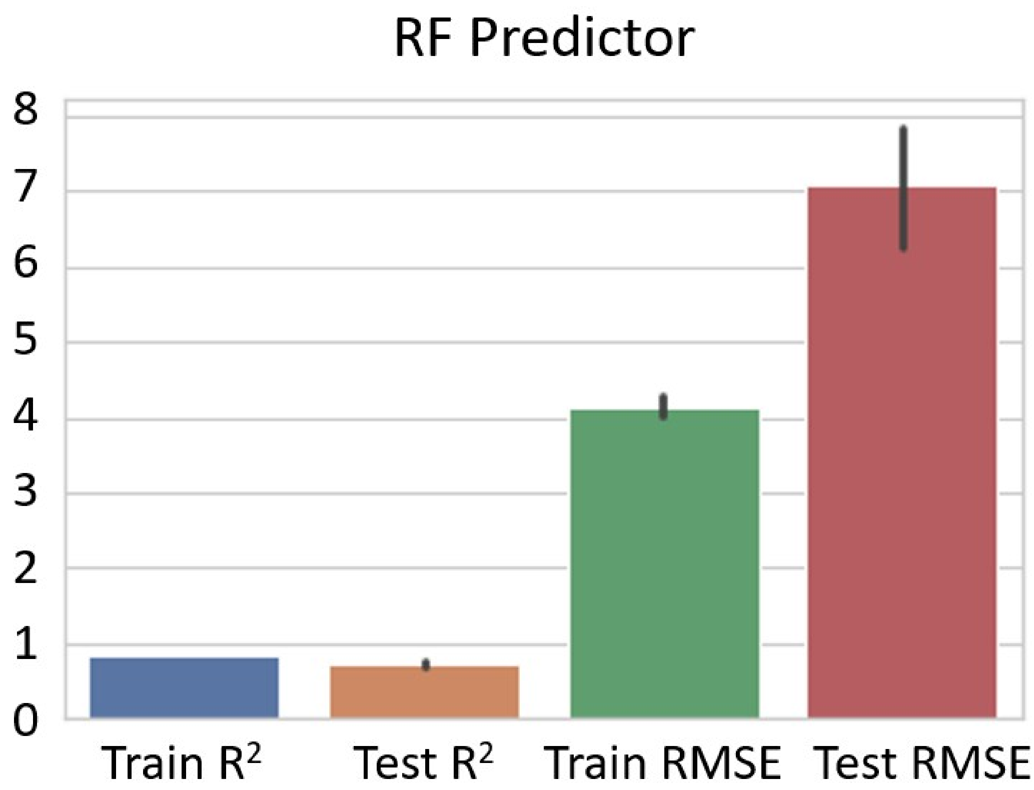

Fig S2. The error bar chart shows the difference in performance evaluation of the RF model

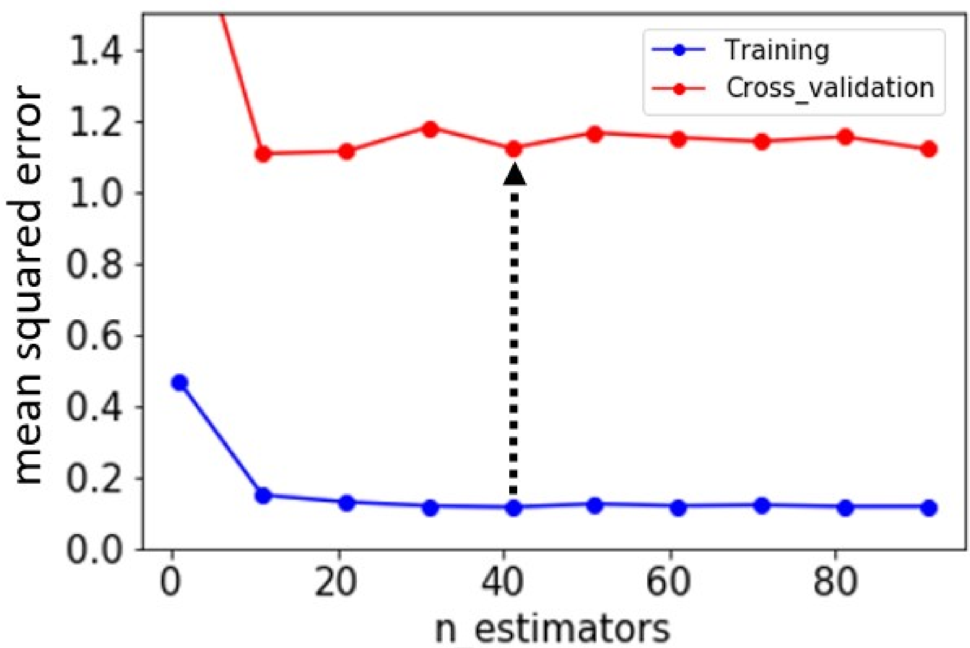

Fig S3. The learning curve of the RF algorithm in 10-fold cross-validation analysis with different parameter n_estimators (for fixed min_samples_split=2, min_samples_leaf=1)

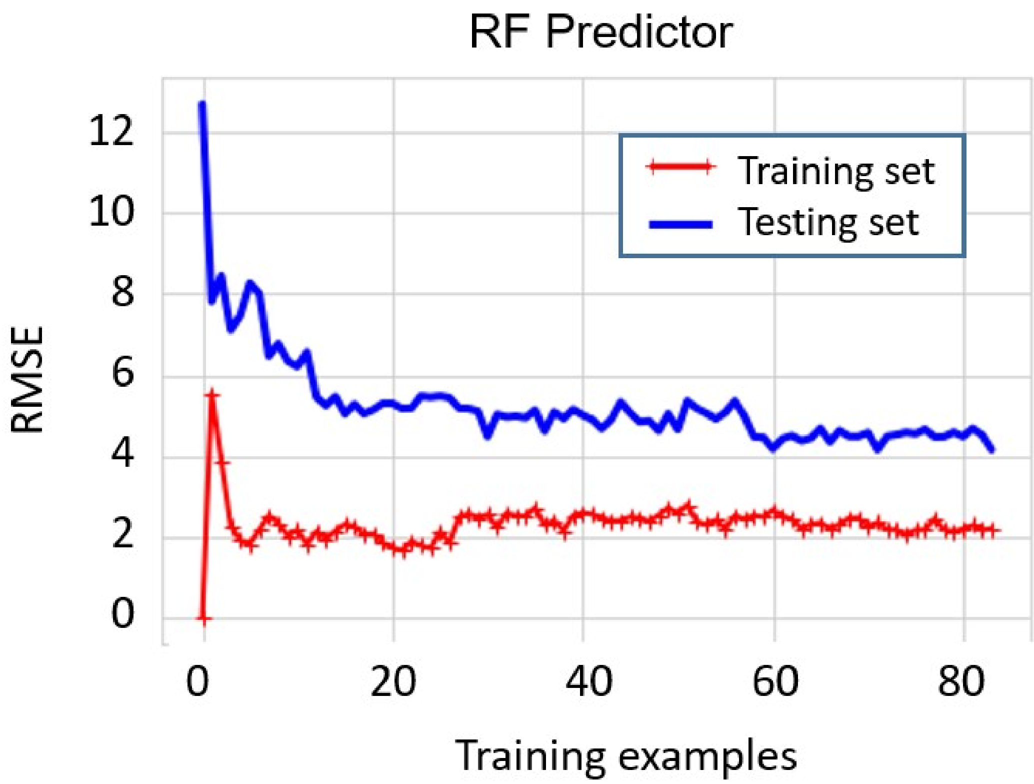

Fig S4. The learning curve for the RF predictor algorithm.

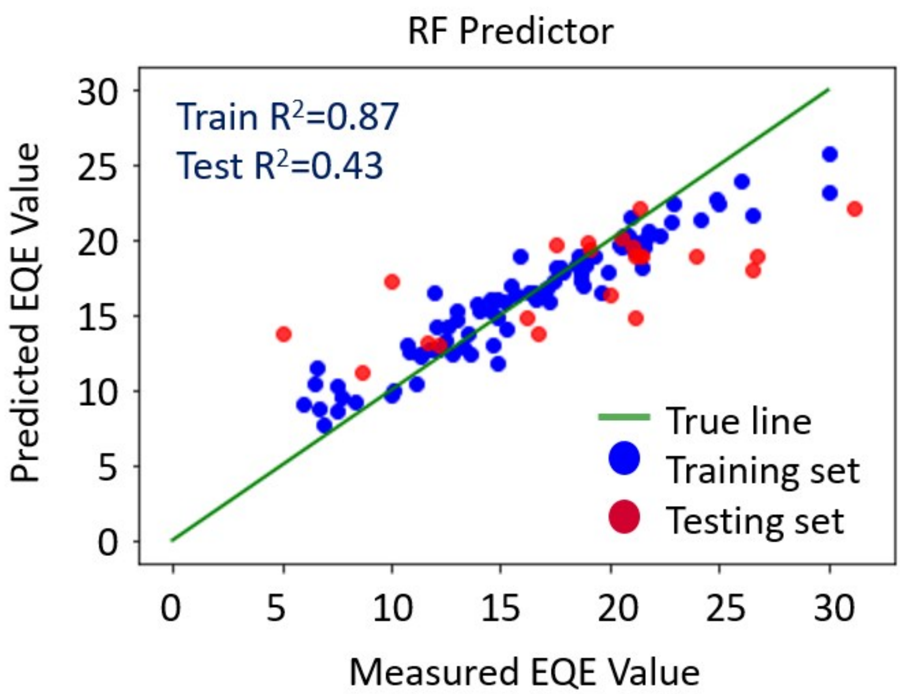

Fig S5. Scatter plots of measured versus predicted EQE values with RF predictor algorithm (the ratio of 8:2 between the training set and testing set)

Table S1. Average R^2 (training set), and R^2 (testing set) from 10 experiments of Random Forest algorithm. The highest R^2 is in experiment #5. The optimized hyperparameters: n_estimators = 40, min_samples_split = 2, min_samples_leaf = 1

Experiment	R^2 (Training set)	R^2 (Testing set)
1	0.85	0.83
2	0.84	0.69
3	0.85	0.70
4	0.85	0.72
5	0.87	0.83
6	0.83	0.71
7	0.86	0.72
8	0.84	0.70
9	0.87	0.72
10	0.84	0.77
Average	0.85	0.74

Table S2. All 108 data points from the literature for the machine-learning model

Guest materials	PLQY (%)	Host materials	EQE (%)	DOI
4CzIPN	93.8	CBP	19.3	DOI: 10.1038/nature11687.
4CzIPN	93.8	3CzPFP	31.2	DOI: 10.1021/acsami.5b01220
4CzIPN	93.8	DCzDCN	26.7	DOI:10.1002/adma.201402188
4CzIPN	93.8	BCPO	19.1	DOI.org/10.1016/j.orgel.2014.04.035
4CzIPN	93.8	PPO27	24.2	DOI.org/10.1021/am501301g
4CzIPN	93.8	pCzB-2CN	22.9	DOI: 10.1021/am508259u
4CzIPN	93.8	mCzB-2CN	26	DOI: 10.1021/am508259u
4CzIPN	93.8	oCzB-2CN	20.9	DOI: 10.1021/am508259u
4CzIPN	93.8	SF33	20.6	DOI.org/10.1002/adma.201501376
4CzIPN	93.8	SF34	22.3	DOI.org/10.1002/adma.201501376
4CzIPN	93.8	3TPAPFP	21.2	DOI.org/10.1021/cm403358h
4CzIPN	93.8	4TPAPFP	6.6	DOI.org/10.1021/cm403358h
4CzIPN	93.8	mCPSOB	26.5	DOI.org/10.1016/j.orgel.2014.10.049
4CzIPN	93.8	DBTTP1	18.7	DOI: 10.1039/c5tc01065a
4CzIPN	93.8	DBTTP2	20	DOI: 10.1039/c5tc01065a
4CzIPN	93.8	4CN34BCz	21.8	DOI: 10.1002/adma.201503782
4CzIPN	93.8	CBP	18.8	DOI:10.1039/C4TC02211D
t-4CzIPN	78	SiCz	17.1	DOI: 10.1002/adma.201402188
m-4CzIPN	67	SiCz	19.6	DOI: 10.1002/adma.201402188
3DPA3CN	100	DPEPO	21.4	DOI: 10.1039/c5cc00511f
BFCz-2CN	85	mCP	12.1	DOI: 10.1039/c5cc01940k
BTCz-2CN	85	mCP	11.8	DOI: 10.1039/c5cc01940k
Ac-VPN	86	mCBP	18.9	DOI.org/10.1002/adfm.201505106
Px-VPN	77	mCBP	14.9	DOI.org/10.1002/adfm.201505106
DCN3	39	PPT	13.3	DOI.org/10.1016/j.synthmet.2015.07.008
DHPZ-2BN	35.2	mCBP	6	DOI: 10.1039/c4tc02530j
5CzBN	70	mCBP	16.7	DOI:10.1039/C5MH00258C
5TCzBN	86	mCBP	21.2	DOI:10.1039/C5MH00258C
CPC	91.3	mCP	21.2	DOI: 10.1021/acsami.5b05648
CPC	91.3	26DCZPPY	15.5	DOI: 10.1021/acsami.5b05648
33TCzPN	87	DPEPO	17.9	DOI:10.1039/C5CC07999C

33TCzT Trz	87	DPEPO	25	DOI: 10.1021/acs.jpcc.5b09114
ACRFL CN (ACRF)	67.3	TPSi-F	10.1	DOI: 10.1002/anie.201206289.
ACRFL CN (ACRF)	67.3	PYD2	7.7	DOI.org/10.1016/j.orgel.2014.05.027
ACRFL CN (ACRF)	67.3	CzSi	8.4	DOI.org/10.1016/j.orgel.2014.05.027
ACRFL CN (ACRF)	67.3	DPEPO	12.2	DOI.org/10.1016/j.orgel.2014.05.027
PXZ-TRZ	65.7	CBP	12.5	DOI: 10.1039/c2cc36237f
DMAC-TRZ	83	mCPCN	26.5	DOI:10.1039/C5CC05022G
ATP-PXZ	63	CBP	11.7	DOI: 10.1088/1468-6996/15/3/034202
m-ATP-PXZ	81	mCP	12.6	DOI: 10.1088/1468-6996/15/3/034202
ATP-ACR	49	mCP	7.5	DOI: 10.1088/1468-6996/15/3/034202
m-ATP-ACR	52	mCBP	8.7	DOI: 10.1088/1468-6996/15/3/034202
m-ATP-CDP	77	mCP	7.5	DOI: 10.1088/1468-6996/15/3/034202
4CzCNpy	54.9	mCP	11.3	DOI.org/10.1002/adom.201500016
Ac-HPM	77	DPEPO	20.9	DOI:10.1039/C5T C04057D
Ac-PPM	79	DPEPO	19	DOI:10.1039/C5T C04057D
Ac-MPM	80	DPEPO	20.4	DOI:10.1039/C5T C04057D
DHPZ-2BTZ	33	mCBP	5	DOI:10.1039/C4TC02530J
DHPZ-2BI	67.6	mCBP	12	DOI:10.1039/C4TC02530J
2PXZ-OXD	87	DPEPO	14.9	DOI:10.1039/C3TC30699B
bis-PXZ-TDZ	68.5	DPEPO	10	DOI: 10.1021/jp510751n
Px2BP	70	mCP	10.7	DOI.org/10.1002/anie.201402992
p-Px2BBP	36	mCBP	6.9	DOI.org/10.1002/anie.201402992
AcPmBPX	46	mCBP	10	DOI:10.1039/C4DT03608E
PxPmBP	57	mCBP	11.3	DOI:10.1039/C4DT03608E

X				
PXZ-DPS	90	CBP	17.5	DOI.org/10.1038/nphoton.2014.12
PTSOPO	80	DPEPO	17.7	DOI.org/10.1016/j.orgel.2015.11.019
PTSOPT	14	DPEPO	17	DOI.org/10.1016/j.orgel.2015.11.019
TXO-PhCz	83	mCP	21.5	DOI.org/10.1002/adma.201401393
TXO-TPA	90.2	mCP	18.5	DOI.org/10.1002/adma.201401393
ACRDSO2	71	CBP	17.5	DOI.org/10.1002/adma.201503225
PXZDSO2	62	CBP	15.2	DOI.org/10.1002/adma.201503225
DPO-TXO2	80	CBP	6.5	DOI:10.1039/C5TC03849A.
DPO-TXO2	80	DPEPO	13.5	DOI:10.1039/C5TC03849A.
PXZ-Mes3B	92	CBP	22.8	DOI: 10.1002/ange.201508270
2DAC-Mes3B	100	DPEPO	21.6	DOI: 10.1002/ange.201508270
DAC-Mes3B	87	DPEPO	14	DOI: 10.1002/ange.201508270
4CzIPN	93.8	Sy-Hosted	21.2	DOI: 10.1021/acsami.8b02766
4CzIPN	93.8	Asy-Hosted	16.6	DOI: 10.1021/acsami.8b02766
PPZTPI	73	CBP	20.52	DOI: 10.1039/C7TC05576E
PPZPPI	99	CBP	21.06	DOI: 10.1039/C7TC05576E
CzSOXO	51.2	mCP	13.6	DOI: 10.1021/acsami.5b12559
oPTC	46.6	mCP	19.9	DOI: 10.1021/acsami.6b03954
mPTC	54.9	mCP	17.4	DOI: 10.1021/acsami.6b03954
2DPyMmDTC	59	mCBP	12.8	DOI: 10.1021/jacs.7b03848
4CzIPN	93.8	CBP	18.6	DOI.org/10.1016/j.optmat.2018.10.002
4CzIPN	93.8	o-mCPBI	18.7	DOI: 10.1021/acsami.5b10464
Cuprous Complexes1	33.1	mCP	14.81	DOI: 10.1021/acs.inorgchem.6b01847
Cuprous Complexes2	31.7	mCP	11.17	DOI: 10.1021/acs.inorgchem.6b01847
Cuprous Complexes3	31.5	mCP	6.67	DOI: 10.1021/acs.inorgchem.6b01847
4CzIPN	93.8	CzTrz	15.9	DOI.org/10.1016/j.jiec.2018.09.040

DPA-o-Trz	80	mCP	17.2	DOI: 10.3938/jkps.72.873
MPA-o-Trz	80	mCP	16.3	DOI: 10.3938/jkps.72.873
PXZPDO	68	CBP	18.8	DOI: 10.1021/jacs.8b04795
PXZDMePDO	54	CBP	12.2	DOI: 10.1021/jacs.8b04795
DMACPDO	86	CBP	23.9	DOI: 10.1021/jacs.8b04795
DMACDMePDO	64	CBP	14.6	DOI: 10.1021/jacs.8b04795
4CzIPN	93.8	Cz-S	10.8	DOI.org/10.1016/j.dyepig.2018.12.001
4CzIPN	93.8	Cz-SO2	15.3	DOI.org/10.1016/j.dyepig.2018.12.001
4CzIPN	93.8	2Cz-S	16.5	DOI.org/10.1016/j.dyepig.2018.12.001
4CzIPN	93.8	2Cz-SO2	14.5	DOI.org/10.1016/j.dyepig.2018.12.001
IndCzpTr-1	73.9	mCBP	14.5	DOI: 10.1039/C8TC01419A
IndCzpTr-2	66.2	mCBP	30	DOI: 10.1039/C8TC01419A
SBDBQ-DMAC	74	CBP	13	DOI: 10.1039/C7SC04669C
22bpmAc	38	DPEPO	15.7	DOI: 10.1021/acs.chemmater.8b00006
25bpmAc	60	DPEPO	20.5	DOI: 10.1021/acs.chemmater.8b00006
55bpmAc	57	DPEPO	24.9	DOI: 10.1021/acs.chemmater.8b00006
4CzIPN	93.8	CBP	21.5	DOI.org/10.1016/j.orgel.2014.10.049
4CzIPN	93.8	mCPSOB	13	DOI.org/10.1016/j.orgel.2014.10.049
4CzIPN	93.8	DMAC-DPS	16.2	DOI.org/10.1016/j.orgel.2016.11.010
4CzIPN	93.8	PPT	12.1	DOI:10.1039/C8CC03425G
4CzIPN	93.8	Cz-PO	20.5	DOI: 10.1021/acsami.6b13002
4CzIPN	93.8	Cz-PS	21.7	DOI: 10.1021/acsami.6b13002
4CzIPN	93.8	Cz-Trz	21.4	DOI: 10.1021/acsami.6b13002
BP-phIDID	56.9	CBP	13.9	DOI: 10.1021/acsami.7b13158
Tria-phIDID	69.8	CBP	20.8	DOI: 10.1021/acsami.7b13158

IndCzpTr-2	71.9	mCBP	30	DOI: 10.1039/C8TC01419A
4CzIPN	93.8	CCNCz	17.1	DOI.org/10.1016/j.dyepig.2019.01.024

Machine-leaning algorithm functions used in this work within the sckit-learn

library:

Random Forest Regression:

- **Importing the libraries**

```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
```

- **Importing the dataset**

```
df = pd.read_csv("TADF OLEDs experimental data.csv")
X = df.iloc[:, :-1]
y = df[['EQE']]
X_train, X_test, y_train, y_test = train_test_split
(X,y,test_size=0.10,random_state=10)
```

- **Fitting the Random Forest Regression Model to the dataset**

```
from sklearn.ensemble import RandomForestRegressor
rfr = RandomForestRegressor (
        bootstrap=True, criterion='mse', max_depth=None,
    max_features='auto', max_leaf_nodes=None,
    min_impurity_decrease=0.0, min_impurity_split=None,
    min_samples_leaf=1, min_samples_split=2,
    min_weight_fraction_leaf=0.0, n_estimators=40, n_jobs=-1,
    oob_score=False, random_state=None, verbose=0, warm_start=False)
rfr.fit (X_train,y_train.values.ravel())
```

- **Predicting the results**

```
rfr_x_predict = rfr.predict(X_train)
rfr_y_predict = rfr.predict(X_test)
```

Support Vector Regression

- **Importing the libraries**

```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
```

- **Importing the dataset**

```
df = pd.read_csv("TADF OLEDs experimental data.csv")
X = df.iloc[:, :-1]
y = df[['EQE']]
X_train, X_test, y_train, y_test = train_test_split
```

(X,y,test_size=0.10,random_state=10)

- **Fitting the Support Vector Regression Model to the dataset**

from sklearn import svm

svr_rbf = svm.SVR(C=40, epsilon=0.1, gamma=3000, kernel='rbf', cache_size=200, coef0=0.0, degree=3, max_iter=-1, shrinking=True, tol=0.001, verbose=False)

svr_rbf.fit(X_train,y_train.values.ravel())

- **Predicting the results**

svr_rbf_x_predict = svr_rbf.predict(X_train)

svr_rbf_y_predict = svr_rbf.predict(X_test)

Python code of validation curve:

degree = np.arange(1,100,10)

train_loss, test_loss = validation_curve(

RandomForestRegressor(min_samples_split=2, min_samples_leaf=1),

preprocessing.scale(X),preprocessing.scale(y),

'n_estimators', degree, **cv=10**,scoring = 'neg_mean_squared_error')