Science & Technology
Facilities Council

MEASUREMENT DATA FOR IMPROVING ITU-R RECOMMENDATION P.1812

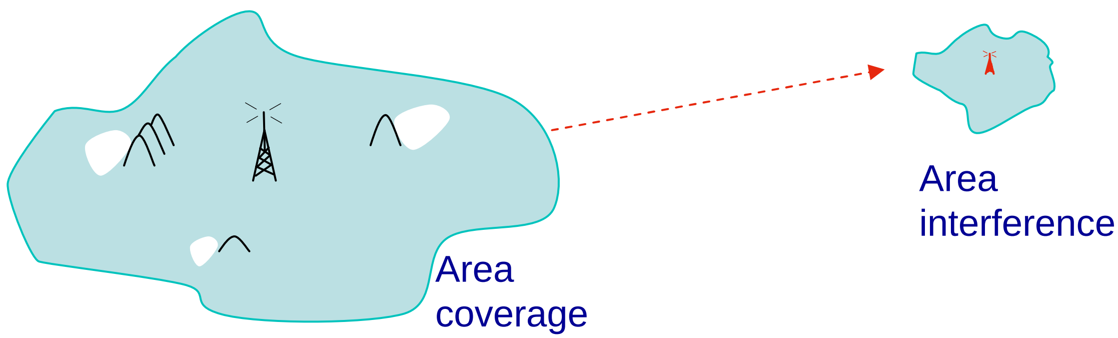
Database of filtered measurement records

Mike Willis, Ken Craig, Nick Thomas
Radio Communications Research Unit, STFC Rutherford Appleton Laboratory, UK
Tel: +44 1235 445492 Fax: +44 1235 446140 Email: m.j.willis@rl.ac.uk

Background

ITU-R Recommendation P.1812

– A path-specific propagation prediction method for point-to-area terrestrial services operating up to 3GHz



– A new recommendation first published in 2007
– The SG3 work plan aims to improve the accuracy of this model

A new database of terrestrial propagation measurements has recently been assembled within ITU-R SG3

– Ideal for testing the aforementioned recommendation

Science & Technology Facilities Council

Rationale

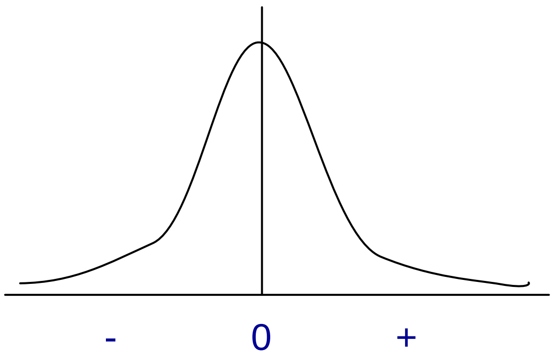
The new database of measurements has proved very useful but:

- Inevitably a large measurement database will contain some errors
  - The difference between a prediction and a measurement depends on both model and measurement errors
- Not all of the data within the database is of equal weight. For example:
  - There are some long term measurements and some point samples
  - There are some height gain measurements which should be used for height gain, but not for path evaluation
- The database is not uniformly distributed over parameter space of the models.
  - Risk of bias if model only matches well to the most numerous measurement set

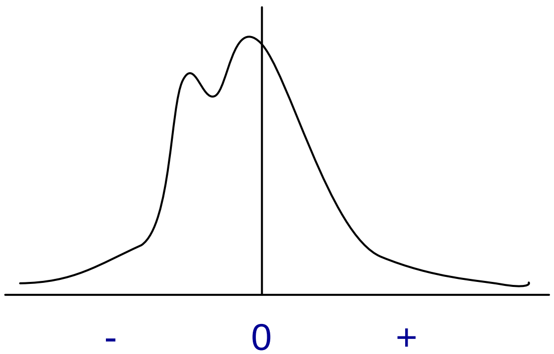
Rationale

In attempting to improve ITU-R P.1812 it became clear differences between models were being obscured by measurement errors

- Some measurements were very clearly in error
  - Line of sight paths with median loss much less than free space (Calibration)
  - Line of sight paths with high excess loss over free space (Clutter)
- The prediction errors were not normally distributed
  - I.E. the errors were not random



Hoped for Error Distribution



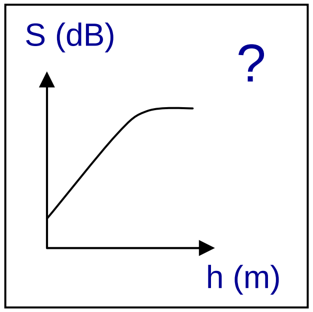
Actual Error Distribution

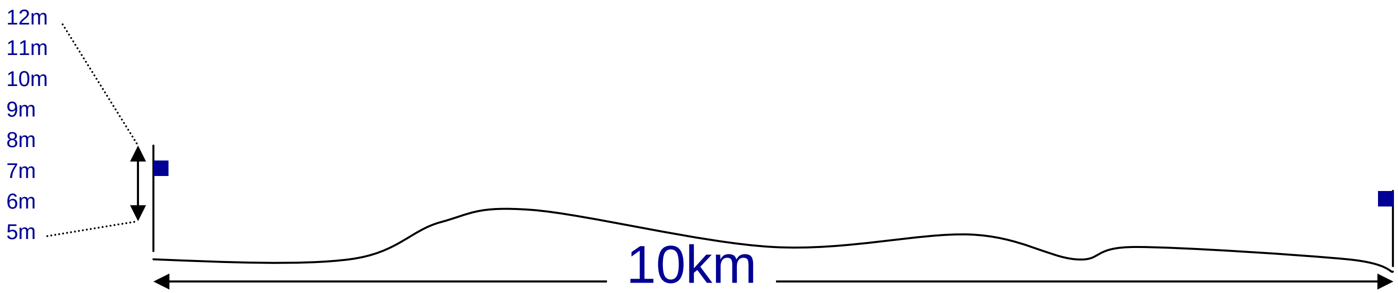
Rationale

Also

- Some datasets were missing some parameters
  - Clutter information along path profile missing
  - Low resolution terrain profiles – especially over sea
  - Missing radioclimatic data and Land/Sea/Coastal information

- The height gain measurements were adding noise to the sampling
  - A lack of associated clutter data meant that many nearly identical paths had large path loss differences





Plan

The measurement database required closer examination and a plan was made:

- Filter the records against a set of rules
  - Flag each record against several criteria
- Which required moving the records into a new database
  - To permit flagging of the data
  - To allow missing data to be filled in where possible
  - To allow new data to be added from new sources
  - To allow model results to be stored alongside the data records
- And to create an extraction application
  - Permit data to be extracted based on database queries of parameter values
  - Plot and display data and results in a readily usable form
  - To help analyse prediction outliers manually

Science & Technology Facilities Council

Data validity

All measurement records were added into the database and then assessed into three categories

- Records thought to be good
- Records with a few minor concerns but otherwise good
  - For example records that were repairable
- Records with major concerns
  - For example
    - Records where the path profile was not monotonic
    - Records with missing vital parameters
    - Records with conflicting information
    - Dummy records designed for model implementation testing

Records falling into the first two are considered suitable for testing

- Records were only flagged, none were thrown away

Flagging 1

Two major level flags were applied

- IsValid
  - Records thought to be good (Flagged as 1)
  - Records with a few minor concerns but otherwise good (Flagged as 1)
    - For example records that were repairable
  - Records with major concerns

    0 Test links

    - 1 TX, RX location concerns
    - 2 Path profile concerns
    - 3 Clutter concerns - depends on our belief in LOS model
    - 4 Duplicate link
    - 5 Calibration concerns - depends on our belief in LOS model

Science & Technology Facilities Council

Flagging 2

- IsLongTerm
  - Records with many time percentages considered long term (1)
  - Records with none or 50% time considered not long term (0)
    - This is the majority of the data

In addition:

An “RxHeightGainGroup” flag was added to identify height gain tests

and

The highest RX point is a record is flagged with “IsTopRXHeightInGroup”

Flag counts

Field and Value		# Links (Measurements)
Total		5832 (35840)
IsValid=0 (test links)		8 (24)
IsValid=-1 (TX, RX location concerns)		38 (130)
IsValid=-2 (profile concerns)		32 (104)
IsValid=-3 (clutter concerns)		341 (3226)
IsValid=-4 (duplicate link)		19 (27)
IsValid=-5 (LOS path loss concerns)		21 (823)
IsValid=1 && IsLongTerm=0 (&& IsWorstMonth=0)	Total	4922 (29061)
	InputsValid=1	4914 (25309)
	InputsValid=1 && IsTopRXHeightInGroup=1	4914 (9639)
IsValid=1 && IsLongTerm=1 && IsWorstMonth=0	Total	428 (2307)
	InputsValid=1	402 (1410)
	InputsValid=1 && IsTopRXHeightInGroup=1	402 (1410)
IsValid=1 && IsWorstMonth=0	InputsValid=1 && IsTopRXHeightInGroup=1	5316 (11049)

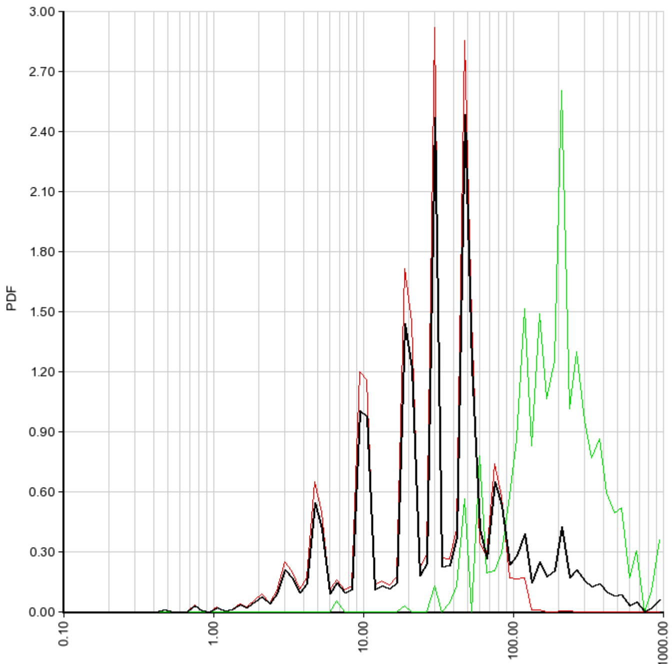
Best set

Parameter coverage of the database

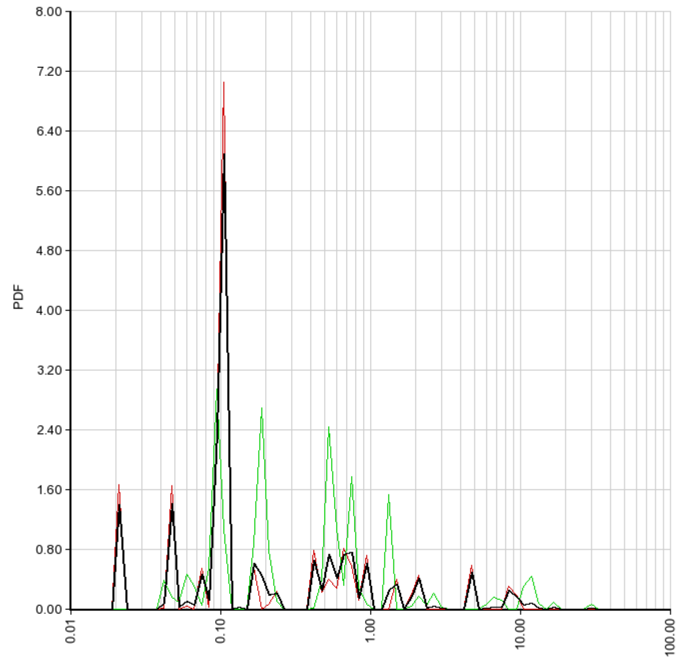
(just a few examples)

Parameter coverage

- How well are longer and shorter paths and higher and lower frequencies represented?



Path Lengths

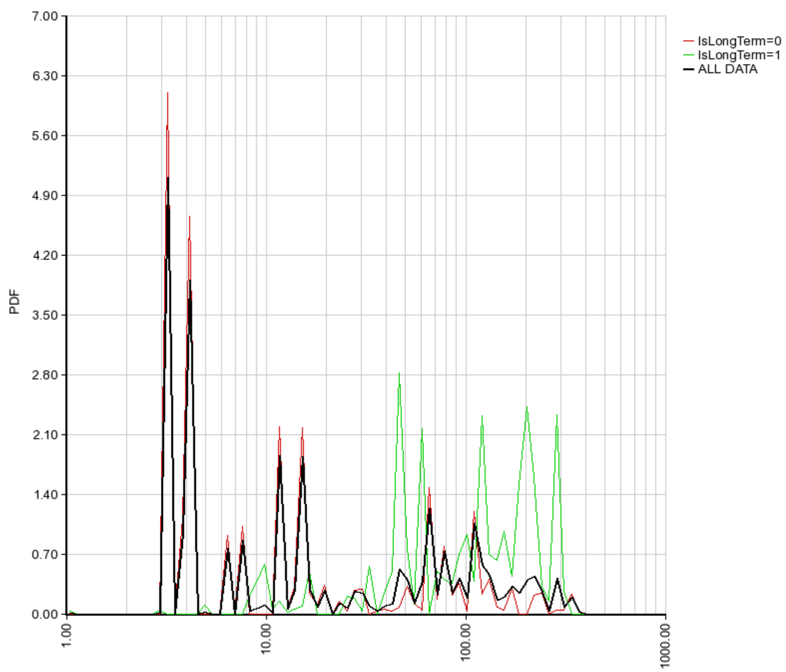


Frequencies

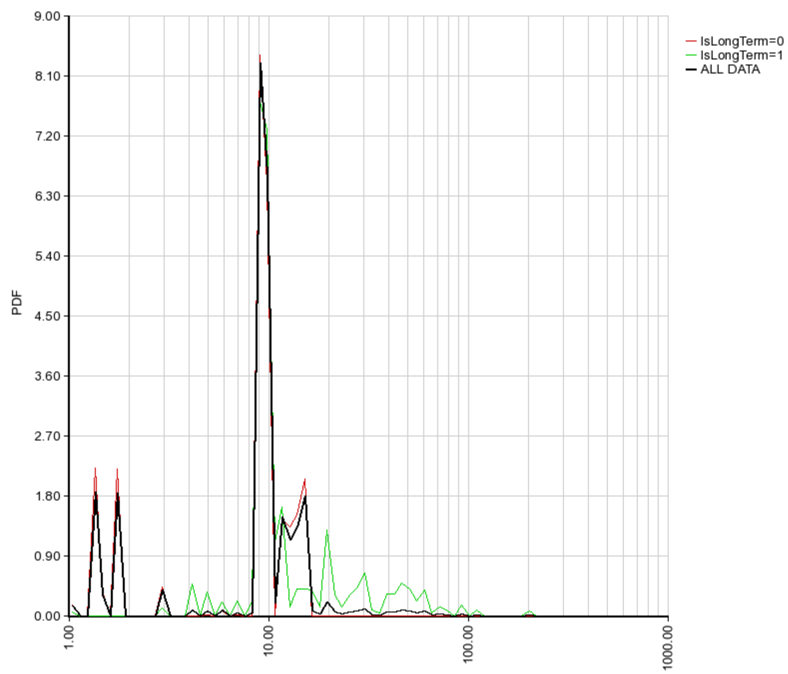
(Green = Long Term data, Red = Short term data, Black = All data)

Parameter coverage

- Height above ground



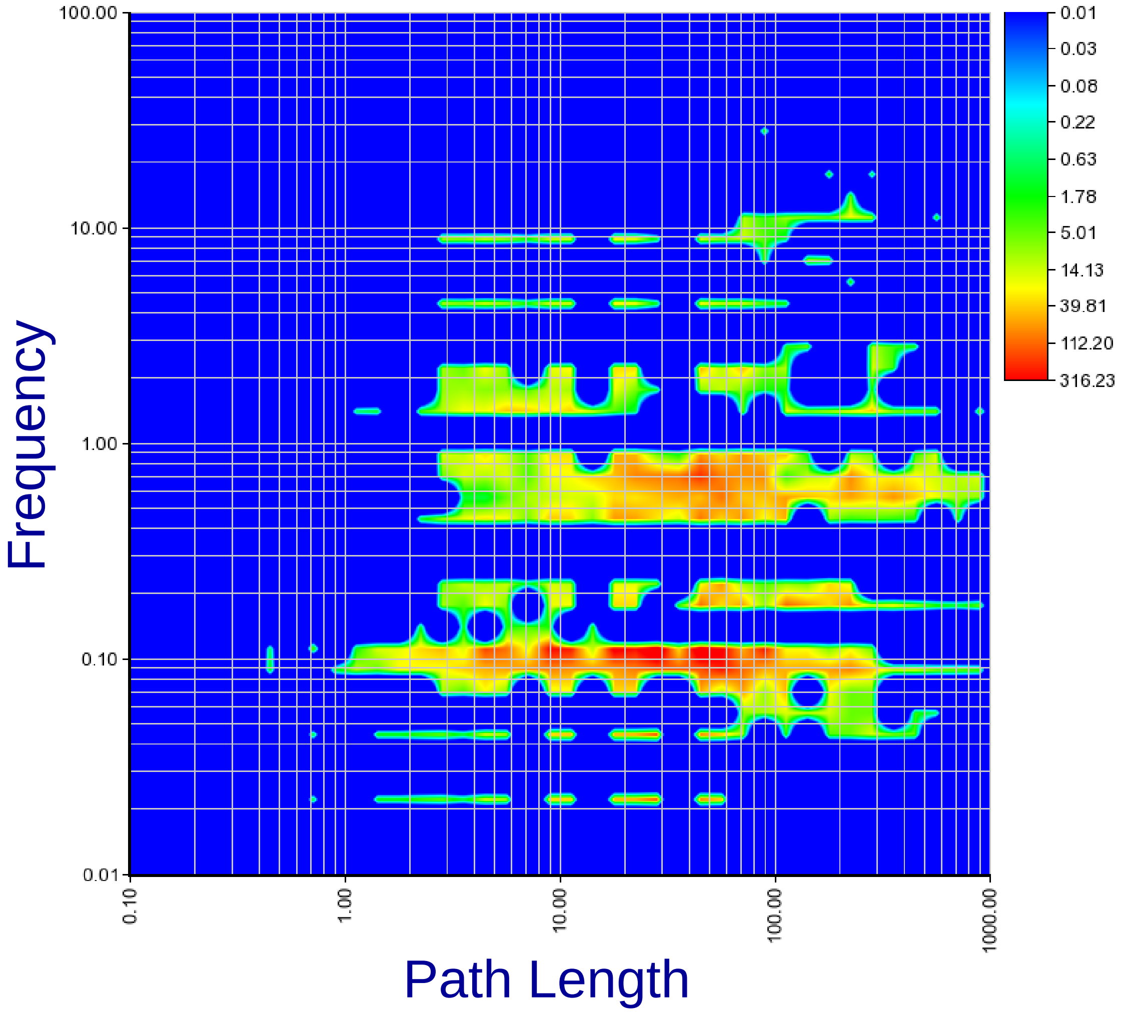
TX



RX

(Green = Long Term data, Red = Short term data, Black = All data)

Joint parameter coverage

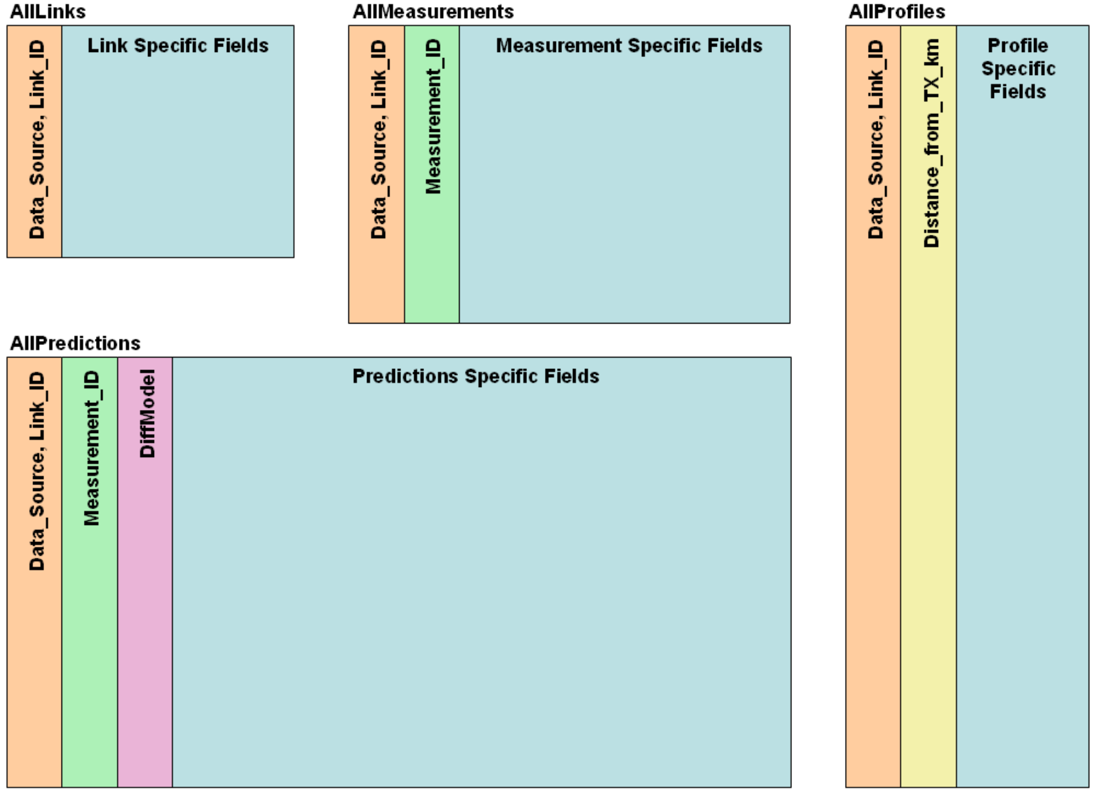


Accessing the data

(and the model results)

SQL Database

Data is stored in an SQL database

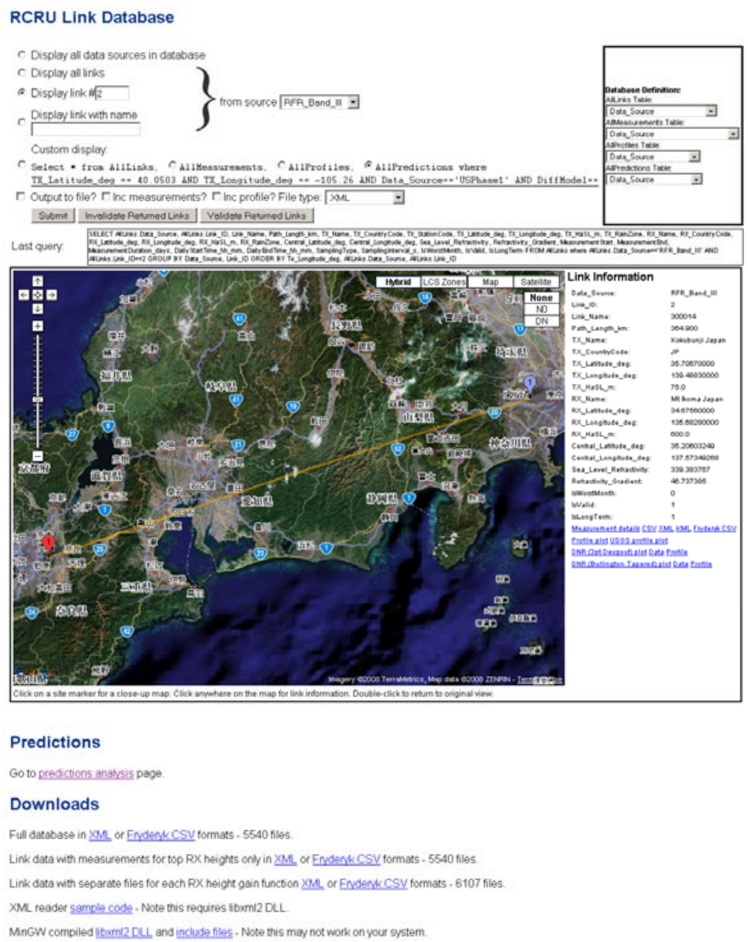


– Can output data in XML, KML, SG3 CSV

Science & Technology Facilities Council

Web access

Link http://www.rcru.rl.ac.uk/njt/linkdatabase/linkdatabase.php



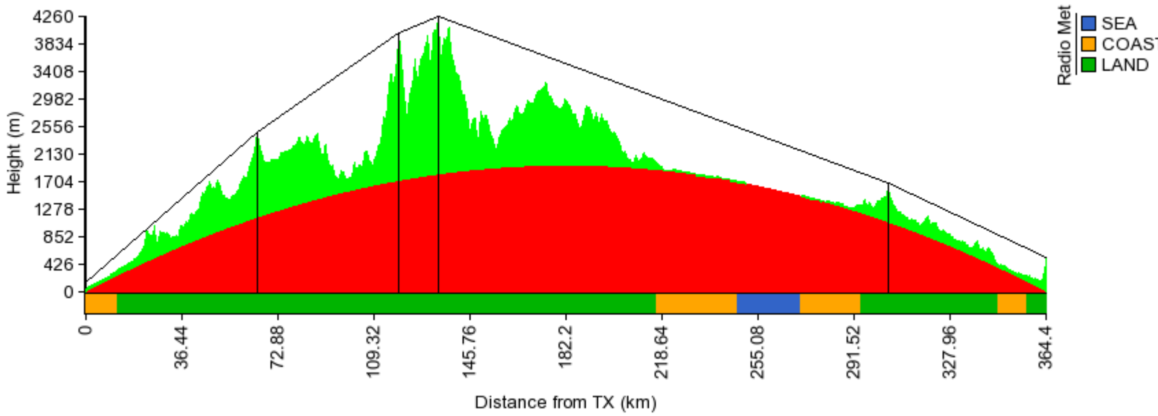
A javascript application facilitates access via the web interface

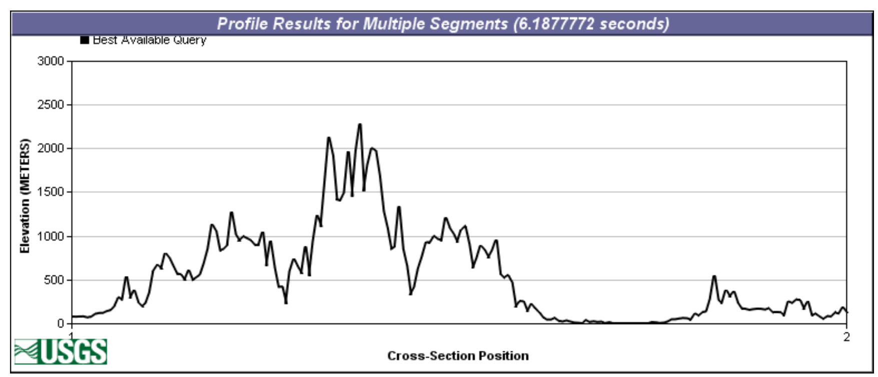
Links in to Google maps

Sample outputs

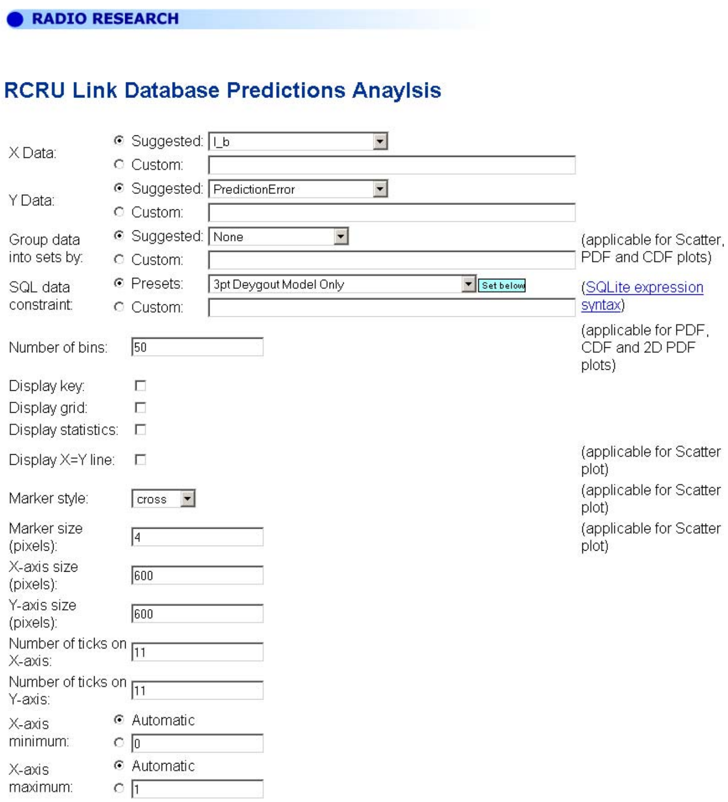
	1	2	3	4	5
Measurement_ID:	0	1	2	3	4
Frequency_GHz:	0.159500	0.159500	0.159500	0.159500	0.159500
TX_AHaG_m:	80.0	80.0	80.0	80.0	80.0
TX_EAHaG_m:					
RX_AHaG_m:	7.0	7.0	7.0	7.0	7.0
Polarization:					
TX_Power_dBm:					
Max_Lb_dB:					
TX_Gain_dBi:					
RX_Gain_dBi:					
RX_AntennaType:					
EIRP_H_dBW:					
EIRP_V_dBW:					
EIRP_Total_dBW:					
HRP_Red_dB:					
TimePercentage:	1.000	10.000	50.000	90.000	99.000
RelativeLossToFS_dB:					
MeasuredFieldStrength_dBuVperm:	-10.9	-17.1	-26.8	-35.2	-42.2
BasicTransmissionLoss_dB:	194.3	200.5	210.2	218.6	225.6
RXHeightGainGroup:	-1	-1	-1	-1	-1
IsTopRXHeightInGroup:	1	1	1	1	1

Profile Points: 366 found
TX-RX Distance (km): 364.4
Terrain Height Range (m): [-9.0-2444.0]
Max Clutter Height (m): 0.0
Clutter Categories: present
Radio Met Categories: present

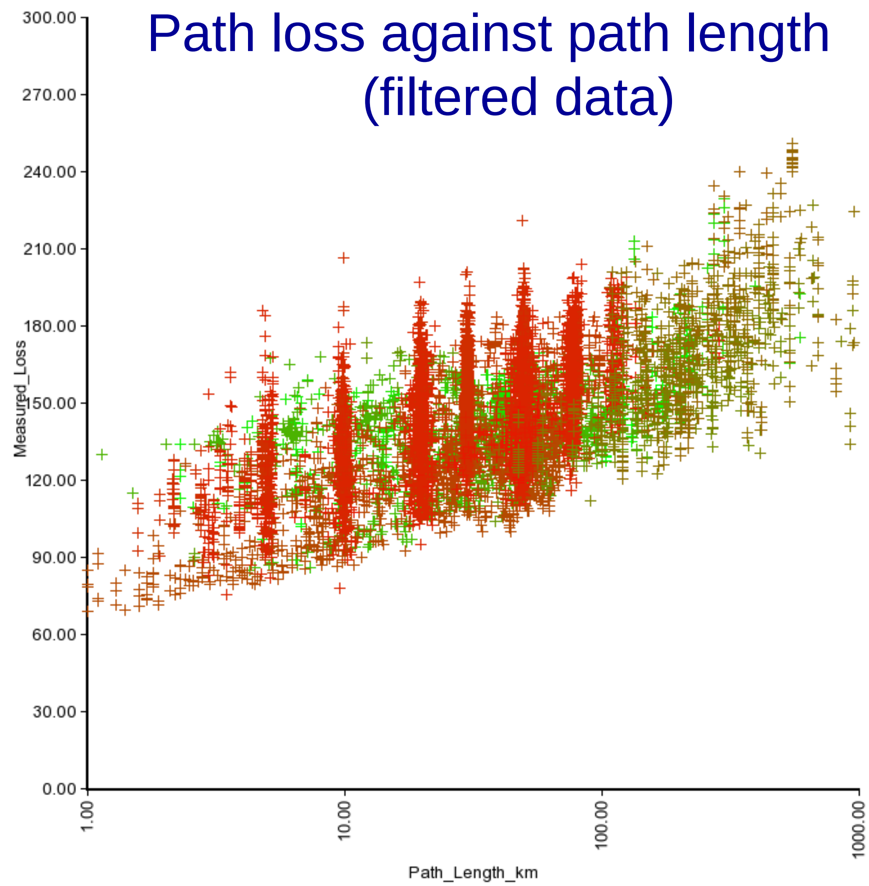




Data analysis



Path loss against path length (filtered data)



http://www.rcru.rl.ac.uk/njt/linkdatabase/linkdatabase_predictions.php

Science & Technology Facilities Council

Displaying model results

Typical prediction analysis data extraction filter:

DiffModel==1 AND IsTopRXHeightInGroup==1 AND IsValid==1 AND IsWorstMonth==0 AND InputsValid==1

Field and Value	Diffraction Model
`DiffModel=1`	3-edge Deygout model as in ITU-R P.1812
`DiffModel=3`	Bullington model with LOS taper
`DiffModel=5`	3-edge Deygout model as in ITU-R P.1812 with some aspects of the US PTP cylindrical edge model
`DiffModel=6`	3-edge Deygout model as in ITU-R P.1812 with Chinese spherical Earth proposal detailed in ITU document 3K/150-E
`DiffModel=7`	Bullington model with LOS taper and Markus Liniger's distance correction [9th order polynomial fit of log(path_length) to 3-edge Deygout mean]
`DiffModel=8`	Bullington model with LOS taper and Markus Liniger's distance correction as additive term
`DiffModel=9`	Bullington model with LOS taper and David Bacon's distance correction [3 point fit to 3-edge Deygout mean]

Science & Technology Facilities Council

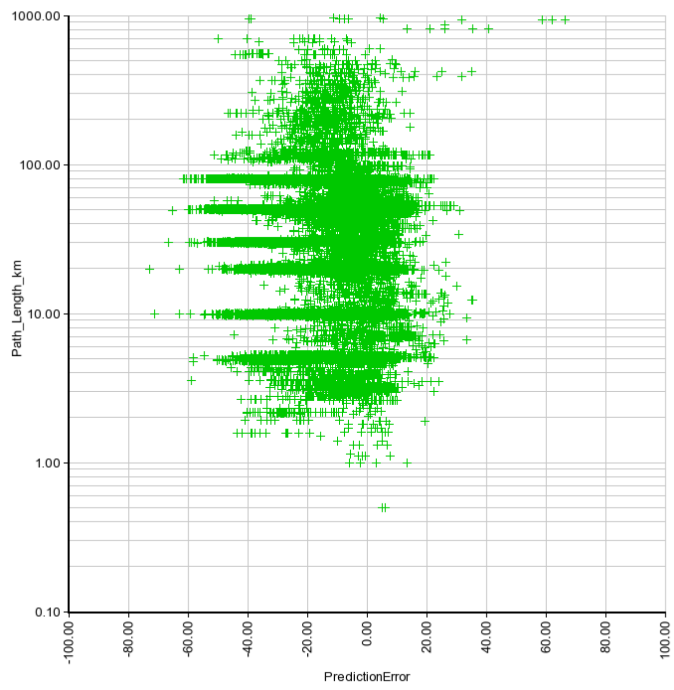
Benefits of new filtered data sets

(by way of examples)

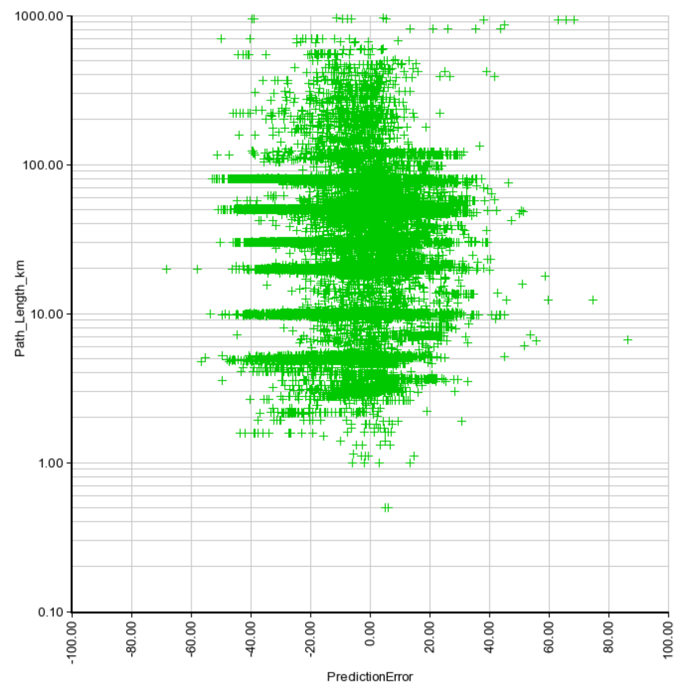
Take two models

Unfiltered data – prediction errors

– Which model is best is hard to tell – neither appears much good



Bullington Taper

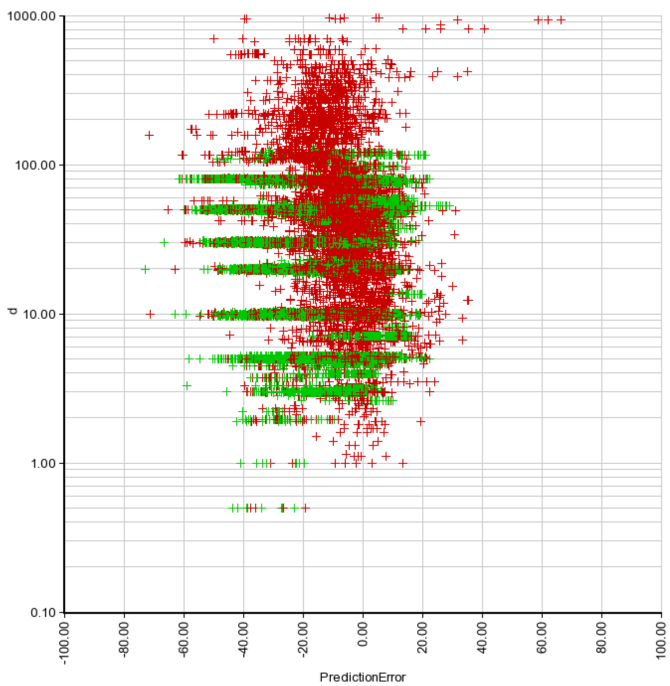


3-Edge Deygout

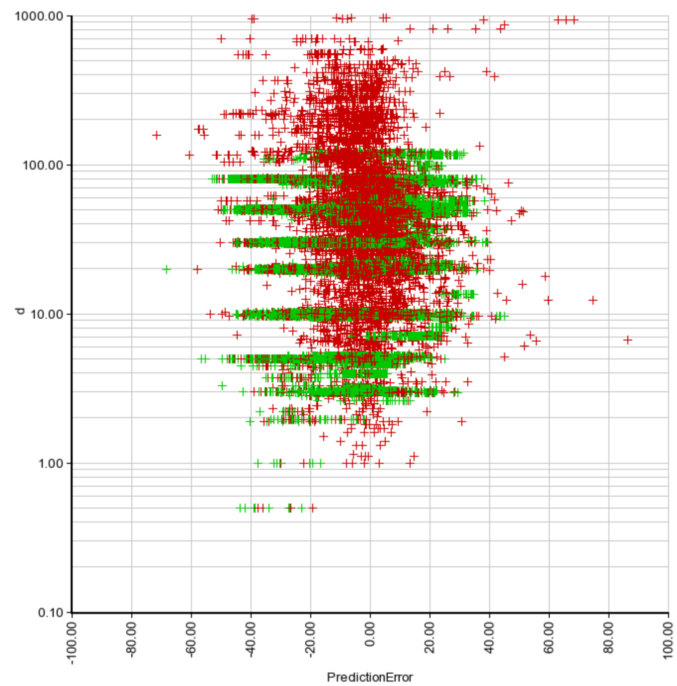
Science & Technology Facilities Council

Take two models

Filtered data – now highlighting the top height measurements, those least likely to suffer terminal clutter



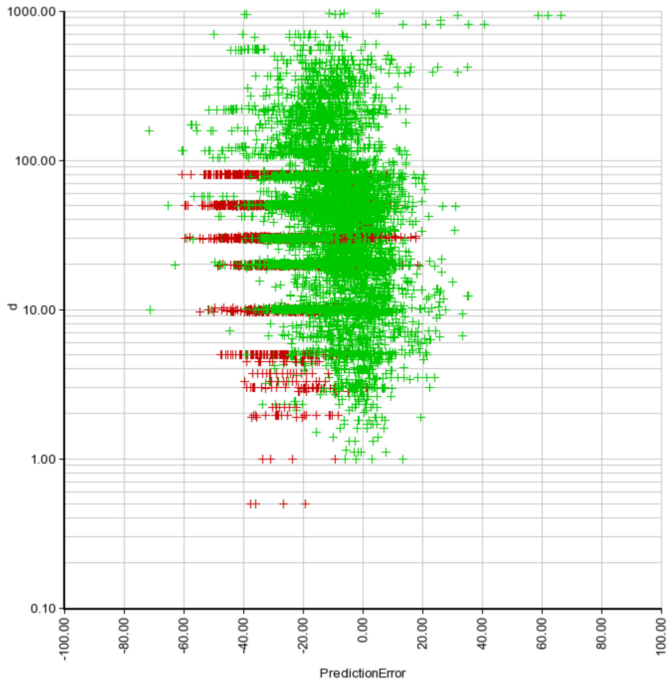
Bullington Taper



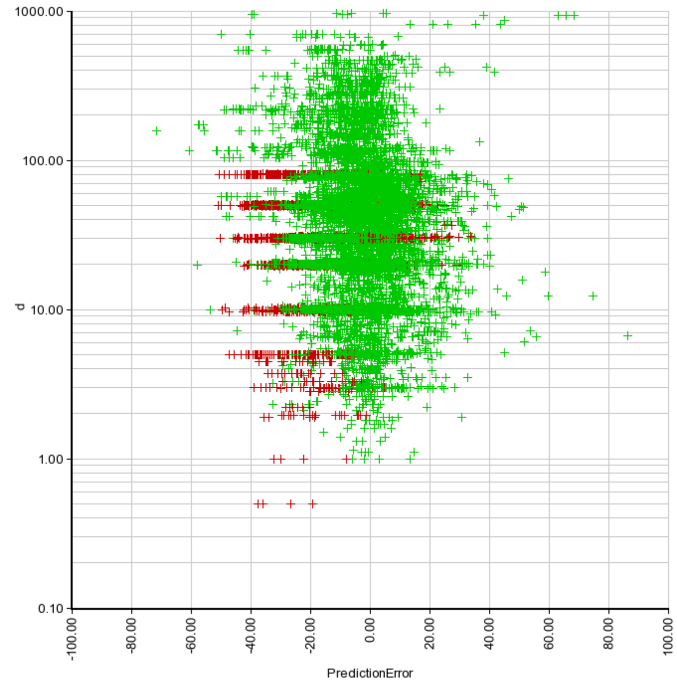
3-Edge Deygout

Take two models

Filtered data – now without height gain measurements but highlighting US Plains Data



Bullington Taper



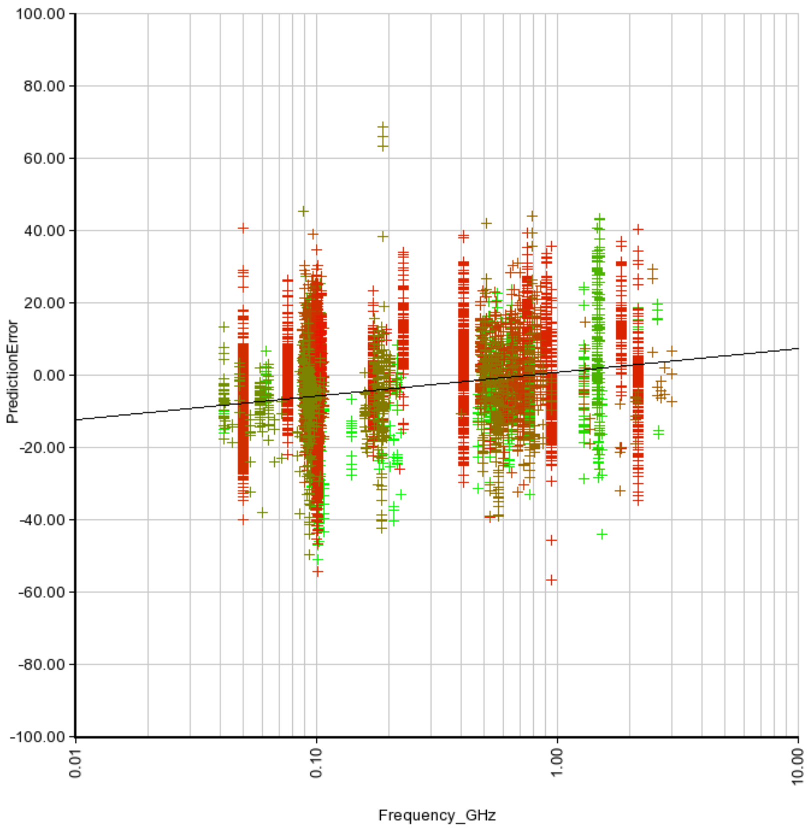
3-Edge Deygout

(we draw no conclusion here, but the differences are becoming clearer)

Regression fits

It is possible to make regression fits to the data

– Swiss 9 point distance correction to Bullington vs frequency

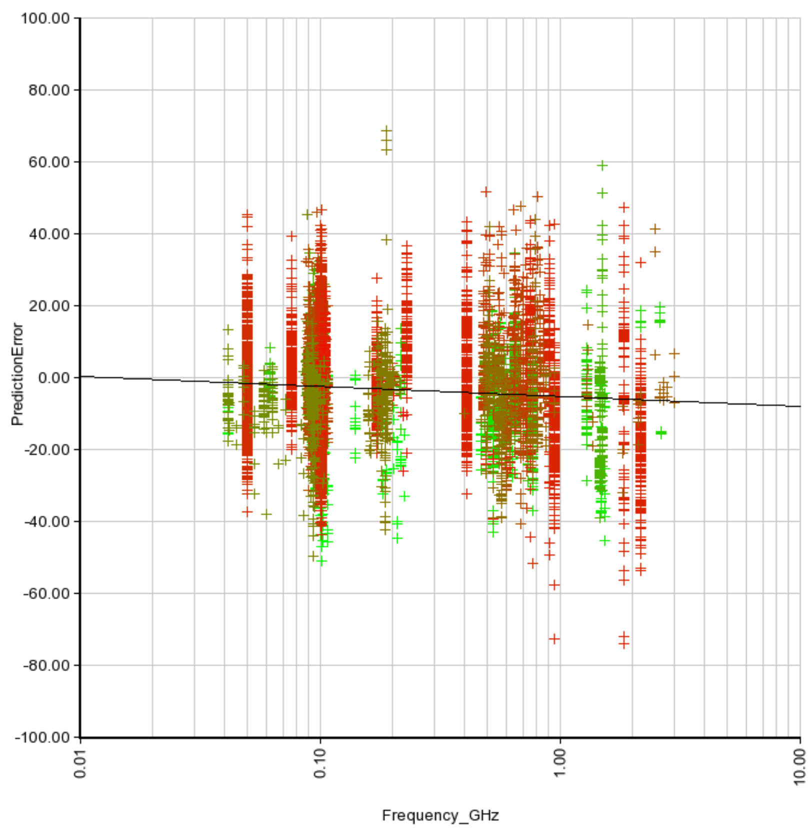


	Points	MeanX	StDvX	MeanY	StDvY	m	b
Data_Source=ABU	108	0.21	0.17	-21.04	14.60	26.04	-0.73
Data_Source=BBC	56	0.18	0.15	-12.36	7.63	-5.29	-16.93
Data_Source=BBCL	25	1.51	0.03	-15.69	10.57	112.28	-35.93
Data_Source=BBCn	252	0.63	0.12	-3.31	8.23	-15.95	-6.63
Data_Source=Cost210	65	1.44	0.56	-1.32	11.91	28.34	-5.00
Data_Source=ERT	9	0.22	0.00	-2.96	5.24	89.40	56.62
Data_Source=HOL	69	0.32	0.29	1.62	7.22	2.40	3.34
Data_Source=IRT	584	0.42	0.27	2.97	7.40	-5.64	-0.04
Data_Source=IRTL	154	1.48	0.02	9.56	15.33	189.19	-22.44
Data_Source=IRTs	63	0.10	0.00	-3.98	7.82	115.61	113.98
Data_Source=ORF	54	0.10	0.00	11.01	9.19	177.86	192.46
Data_Source=RAI	83	0.38	0.21	3.41	7.54	-5.93	0.26
Data_Source=RFR_Band_I	120	0.05	0.01	-8.21	7.07	-22.39	-36.83
Data_Source=RFR_Band_II	242	0.09	0.00	-8.90	11.46	-63.41	-74.17
Data_Source=RFR_Band_III	295	0.18	0.01	-5.18	12.35	99.48	67.95
Data_Source=RFR_Band_IV	250	0.54	0.03	-6.39	12.30	-147.82	-45.83
Data_Source=RFR_Band_V	174	0.76	0.08	-5.79	11.55	-1.28	-5.95
Data_Source=RFR_Band_VI	42	2.01	0.61	-4.72	11.16	12.78	-8.33
Data_Source=S	107	0.63	0.04	-5.32	10.55	58.97	6.40
Data_Source=SUI	1114	0.10	0.00	-1.14	9.63	6.49	5.46
Data_Source=Swiss	405	0.66	0.13	2.25	9.26	6.04	3.41
Data_Source=TDF	64	0.61	0.10	1.93	11.22	-48.74	-8.80
Data_Source=USPhase1	4921	0.09	0.02	-8.76	10.95	-1.94	-10.81
Data_Source=USPhase2	1642	0.72	0.68	1.47	12.86	1.01	1.83
Data_Source=YLE	100	0.11	0.00	5.72	6.70	-1.60	4.15
Data_Source=YLEs	51	0.21	0.17	-3.51	11.04	-9.69	-11.07
ALL DATA	11049	0.31	0.43	-4.43	12.18	6.56	0.64

Regression fits

It is possible to make regression fits to the data

- Alternative distance correction to Bullington vs frequency



	Points	MeanX	StDvX	MeanY	StDvY	m	b
Data_Source=ABU	108	0.21	0.17	-21.43	14.79	22.07	-4.22
Data_Source=BBC	56	0.18	0.15	-8.97	7.67	-11.21	-18.65
Data_Source=BBCL	25	1.51	0.03	-25.05	13.38	559.60	-125.91
Data_Source=BBCn	252	0.63	0.12	-11.17	9.33	-2.09	-11.60
Data_Source=Cost210	65	1.44	0.56	-1.80	12.15	29.29	-5.61
Data_Source=ERT	9	0.22	0.00	-1.40	8.87	149.75	98.38
Data_Source=HOL	69	0.32	0.29	-2.87	8.96	-9.51	-9.67
Data_Source=IRT	584	0.42	0.27	1.82	8.56	-9.63	-3.31
Data_Source=IRTL	154	1.48	0.02	-10.75	18.53	1417.26	-250.44
Data_Source=IRTs	63	0.10	0.00	-3.07	8.45	117.73	117.05
Data_Source=ORF	54	0.10	0.00	14.12	8.83	104.25	120.48
Data_Source=RAI	83	0.38	0.21	0.41	6.53	-7.07	-3.35
Data_Source=RFR_Band_I	120	0.05	0.01	-7.51	7.00	-16.74	-28.91
Data_Source=RFR_Band_II	242	0.09	0.00	-8.11	11.78	-67.64	-77.72
Data_Source=RFR_Band_III	295	0.18	0.01	-5.08	12.21	70.38	46.66
Data_Source=RFR_Band_IV	250	0.54	0.03	-6.27	12.30	-147.68	-45.66
Data_Source=RFR_Band_V	174	0.76	0.08	-5.53	11.43	12.77	-3.98
Data_Source=RFR_Band_VI	42	2.01	0.61	-4.09	12.58	13.68	-7.96
Data_Source=S	107	0.63	0.04	-14.80	11.90	164.17	17.83
Data_Source=SUI	1114	0.10	0.00	-1.22	10.13	-5.20	-6.51
Data_Source=Swiss	405	0.66	0.13	5.38	13.59	3.88	6.12
Data_Source=TDF	64	0.61	0.10	2.14	19.71	-155.57	-32.13
Data_Source=USPhase1	4921	0.09	0.02	-3.40	13.08	-3.02	-6.59
Data_Source=USPhase2	1642	0.72	0.68	-2.73	16.17	-10.72	-6.58
Data_Source=YLE	100	0.11	0.00	5.80	7.01	-2.13	3.71
Data_Source=YLEs	51	0.21	0.17	-7.59	11.38	-26.59	-28.32
ALL DATA	11049	0.31	0.43	-3.19	13.50	-2.87	-5.42

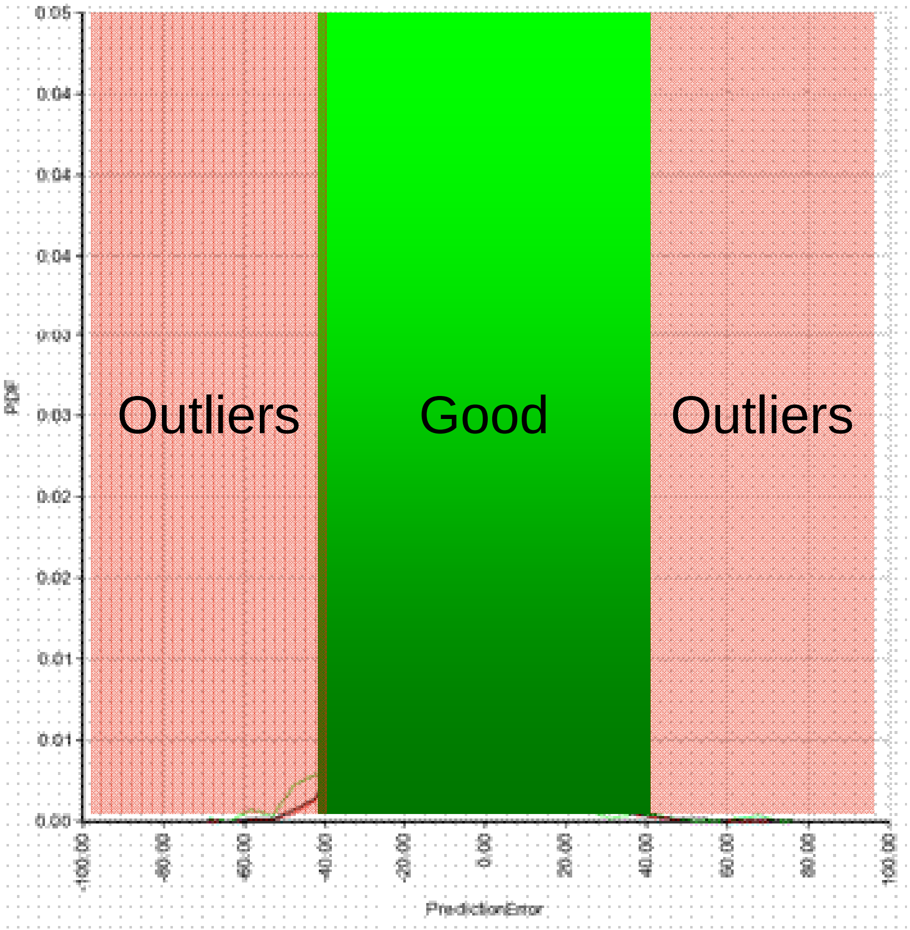
- Better at higher frequencies

Data outliers

(From P.1812)

Outliers

Even with filtered data some measurements are still poorly predicted by P.1812



– Manually looked at all data more than 40 dB out

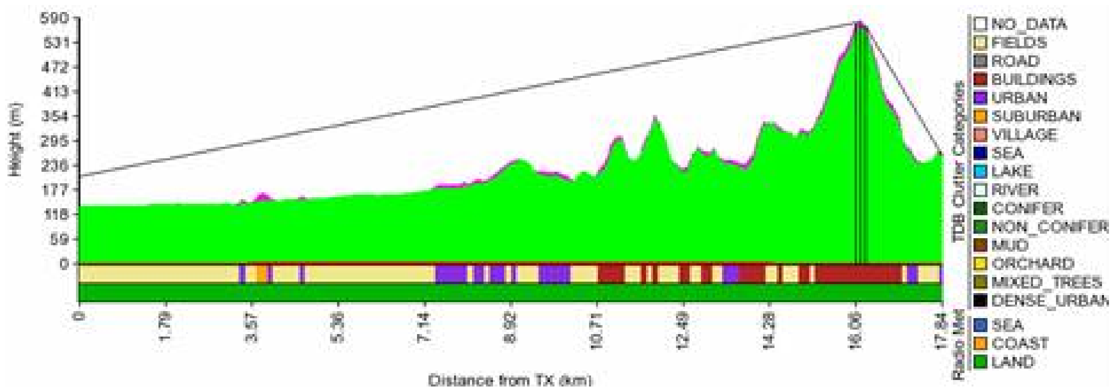
Location accuracy

Some locations are not exact

- Is the supplied profile right?
- We hope so but in this terrain it is very important, especially with low antennas at the receiver

Location accuracy

Here we find a 45 dB over predicted loss



Allegedly, the receiver is 1.5m above ground, just over the brow of a hill, in the middle of a forest a long way from the road.

This is unlikely – moving the receiver to the top of the hill and re-calculating the profile gave a model loss within 6 dB of the measurement.

Unknown clutter

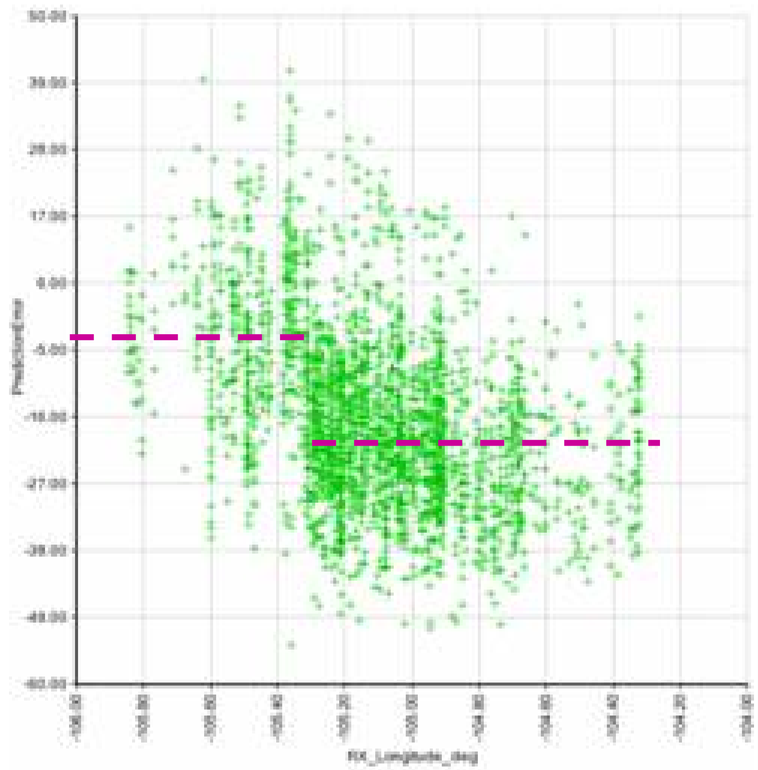
We noticed something odd with US Phase1 records

- Many of these were made at 100MHz using a 68metre tower in Boulder, Colorado.

Prediction error is a function of Longitude.

Paths into the mountains immediately to the West are predicted well.

Paths in plains to the East have 20 dB more mean loss than the model predicts.



At 50% time, this can not be a problem with the ducting h_m model

We believe it is clutter loss

Unrepresented climates

Paths in the Gulf region are not well modelled

- Ducting occurs for more than 50% of the time

This path measurement is 60dB higher than the prediction

Science & Technology Facilities Council

What to do?

- Outliers are an issue - We need to note they exist
  - If we eliminate all measurements that disagree with the model the model is bound to agree with the remaining data.
- In some cases we can eliminate the measurement:
  - Should we not test against Gulf area measurements until we have a ducting model for that climate?
  - We can eliminate paths where we know there is clutter contamination
    - but what about those we don't know about?
- For all models, the mean prediction error for some datasets appears abnormal.
    - Should we eliminate these data sets?
    - Can we equalise the means for evaluating model fit?
- Should we apply a weighting function to distinguish long term and short term measurements

Finally

The models so far for all valid data and top heights against path length