



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**

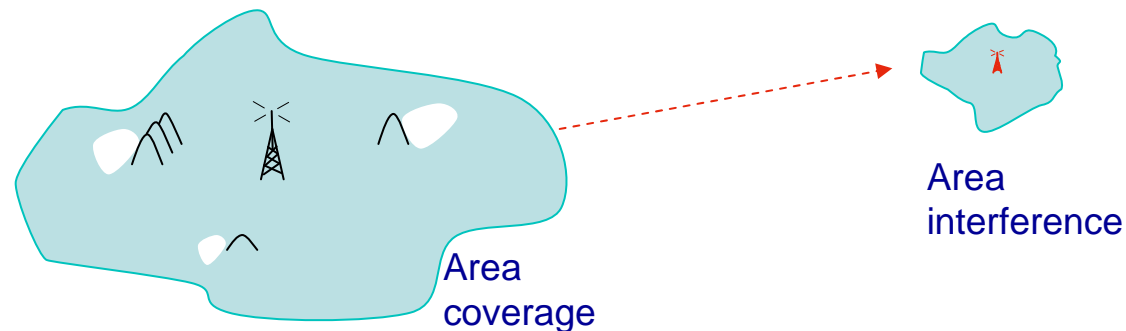
Tel: +44 1235 445492 Fax: +44 1235 446140 Email: m.j.willis@rl.ac.uk



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



Rationale

The new database of measurements has proved very useful but:

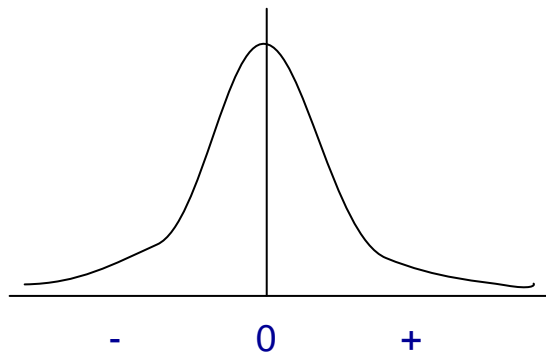
- Inevitably a large measurement database will contain some errors
 - The difference between a prediction and a measurement depends on both model and measurement errors
- Not all of the data within the database is of equal weight. For example:
 - There are some long term measurements and some point samples
 - 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



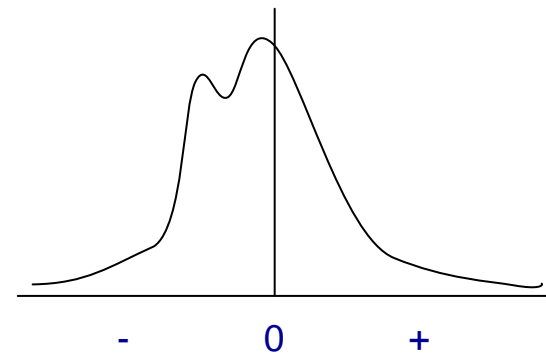
Rationale

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

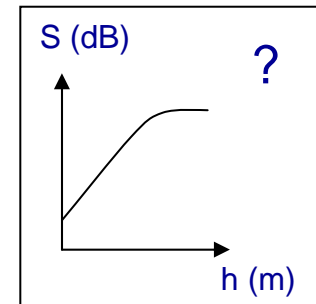
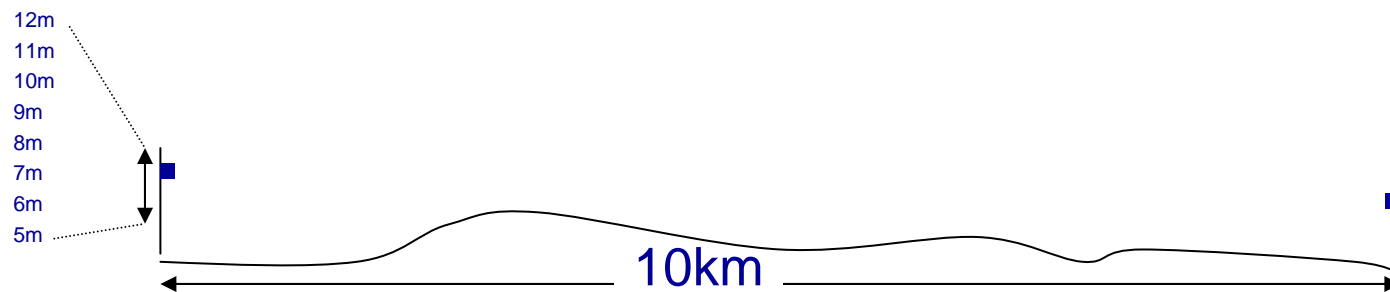


Actual Error Distribution



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



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



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 in 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



Science & Technology
Facilities Council

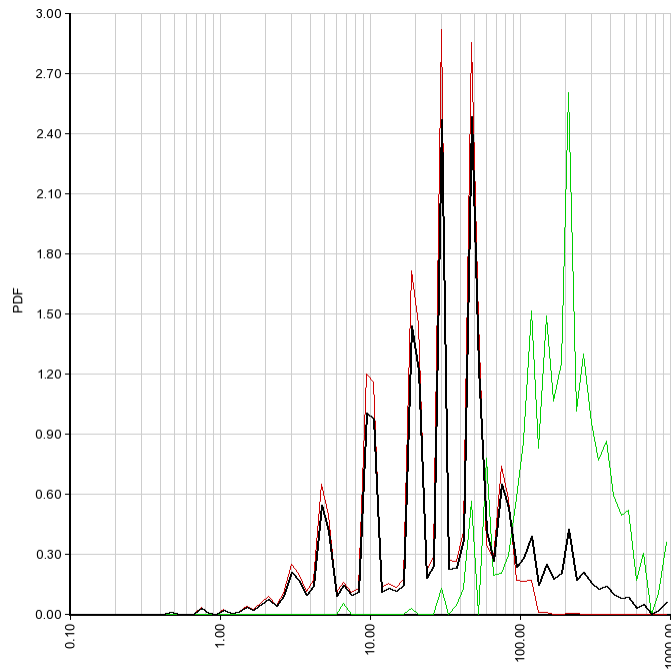
Parameter coverage of the database

(just a few examples)

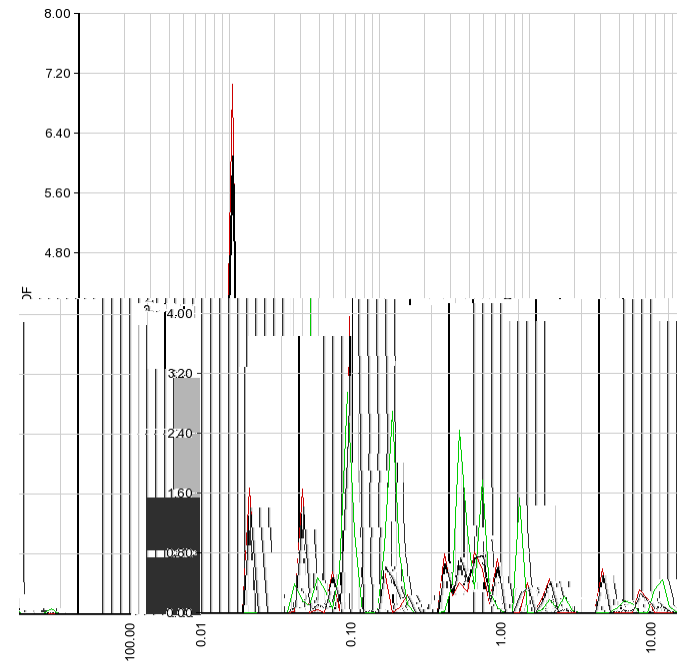


Parameter coverage

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



Path Lengths



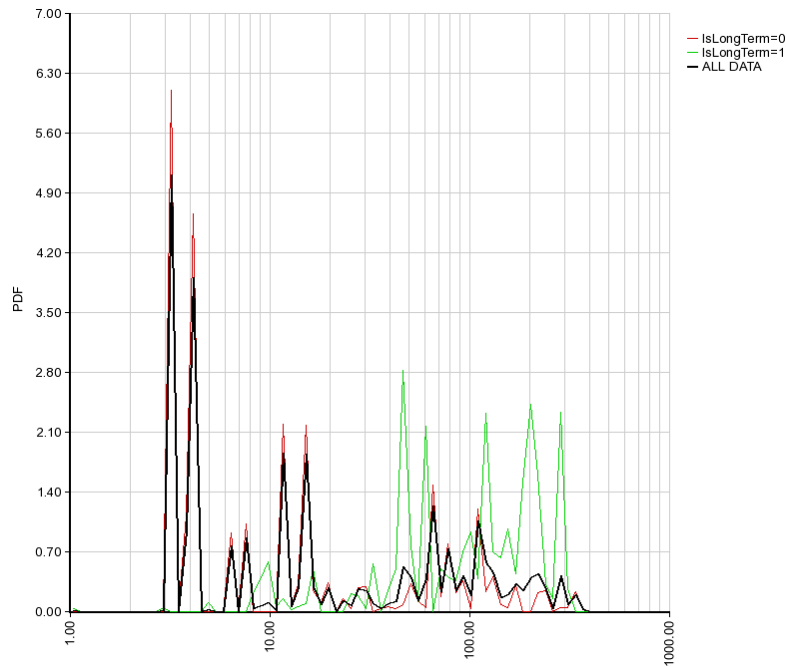
Frequencies

(Green = Long Term data, Red = Short term data, Black = All data) ¹²

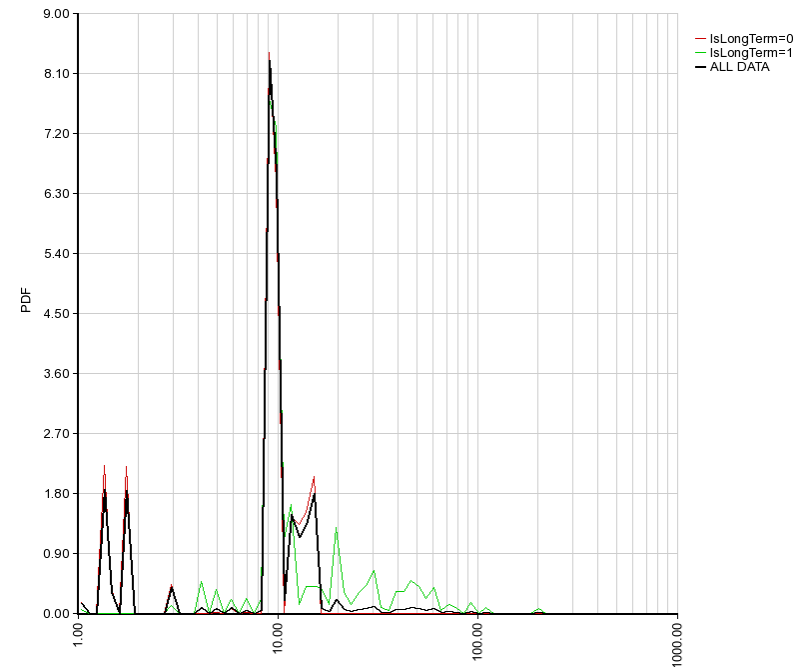


Parameter coverage

- Height above ground



TX

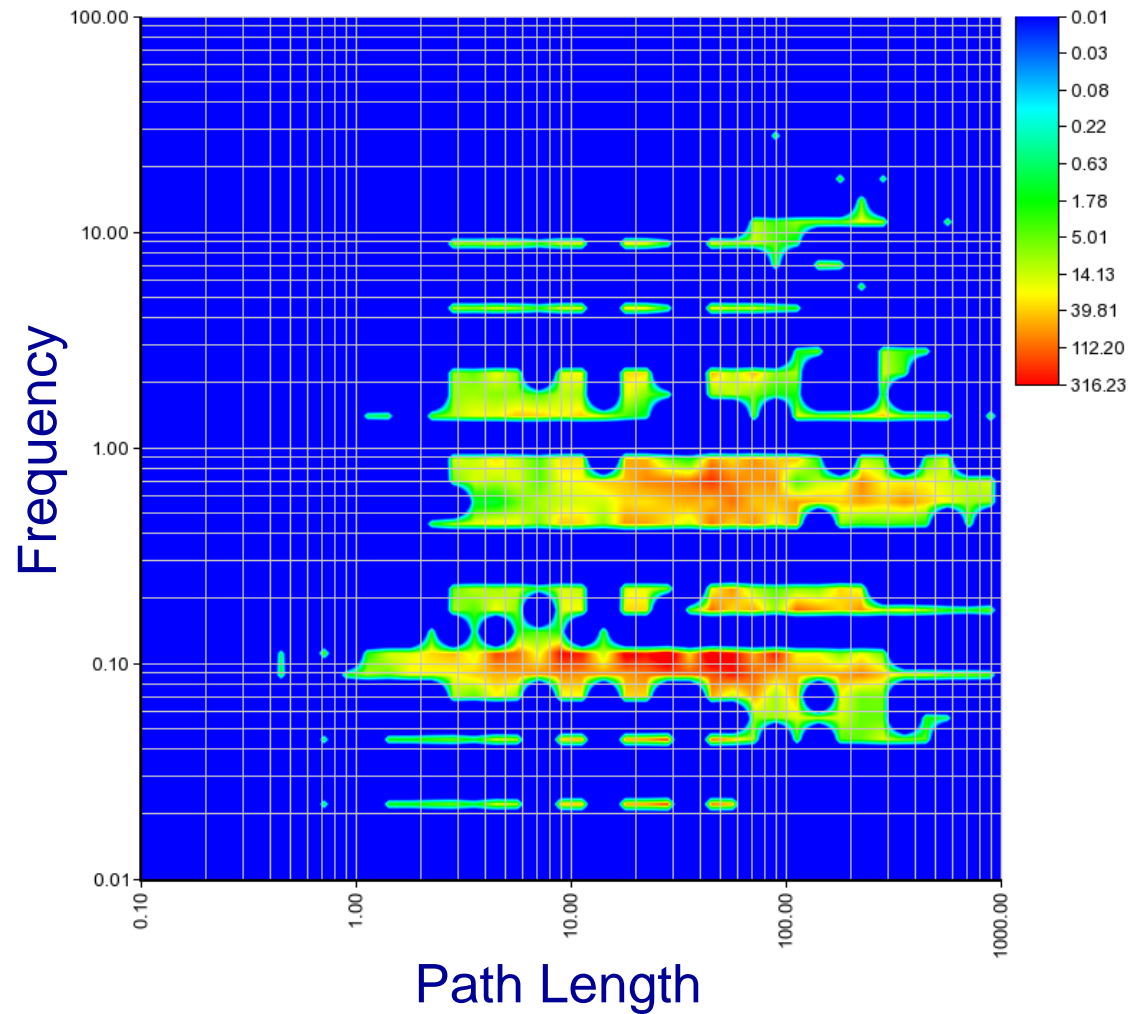


RX

(Green = Long Term data, Red = Short term data, Black = All data) 13



Joint parameter coverage





Science & Technology
Facilities Council

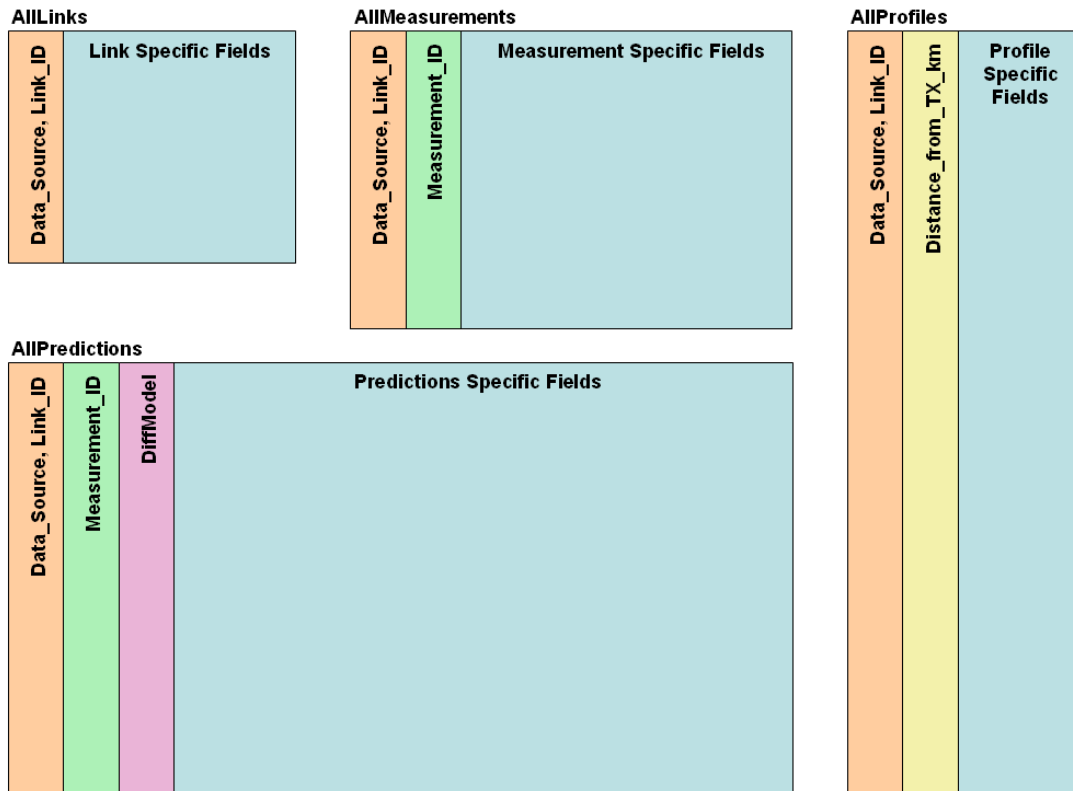
Accessing the data

(and the model results)



SQL Database

Data is stored in an SQL database



- Can output data in XML, KML, SG3 CSV



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

RCRU Link Database

Display all data sources in database
 Display all links
 Display link #2
 Display link with name

from source **RFR_Band_II**

Custom display
 Select * from AllLinks, AllMeasurements, AllProfiles, AllPredictions where
 TX_Latitude_deg == 40.9503 AND TX_Longitude_deg == -105.26 AND Data_Source == 'RFR_Band_II' AND DiffMode ==

Output to file? Inc measurements? Inc profile? File type: **XML**

Database Definition:
 Address Table
 Data_Source
 AllMeasurements Table
 Data_Source
 AllProfiles Table
 Data_Source
 AllPredictions Table
 Data_Source

Last query: `SELECT AllLinks.Data_Source, AllLinks.Link_ID, Link_Name, Path_Length_m, TX_Name, TX_CountryCode, TX_StationCode, TX_Latitude_deg, TX_Longitude_deg, TX_Height_m, TX_Radius, RX_Name, RX_CountryCode, RX_Latitude_deg, RX_Longitude_deg, RX_Height_m, RX_Radius, Central_Longitude_deg, Central_Longitude_deg, Max_Antenna_Height, Frequency, Channel, MeasurementDate, MeasurementDate, MeasurementDuration_sec, DataStartTime_min, DataEndTime_min, SamplingType, SamplingInterval_s, Interpolated, InModel, IsLongTerm FROM AllLinks where AllLinks.Data_Source='RFR_Band_II' AND AllLinks.Link_ID=200000 BY Data_Source, Link_ID ORDER BY TX_Longitude_deg, AllLinks.Data_Source, AllLinks.Link_ID`

Link Information

Data_Source	RFR_Band_II
Link_ID	2
Link_Name	300014
Path_Length_m	964000
TX_Name	Kakuharu-Japan
TX_CountryCode	JP
TX_Latitude_deg	35.70670000
TX_Longitude_deg	139.46030000
TX_Height_m	75.0
RX_Name	M8 Bama Japan
RX_Latitude_deg	34.67900000
RX_Longitude_deg	136.66200000
RX_Height_m	600.0
Central_Longitude_deg	35.20603240
Central_Longitude_deg	137.67946209
Max_Antenna_Height	330.360707
Radius_of_Gradient	46.737365
Interpolated	0
InModel	1
IsLongTerm	1

[MainDatabase.html CSV XML EML Enduser CSV](#)
[LinkDatabase.html CSV XML EML Enduser CSV](#)
[RFR_Band_II.html CSV XML EML Enduser CSV](#)
[RFR_Band_III.html CSV XML EML Enduser CSV](#)

A javascript application facilitates access via the web interface

Links in to Google maps

Predictions

Go to [predictions analysis](#) page.

Downloads

Full database in [XML](#) or [Erydenk CSV](#) formats - 5540 files.

Link data with measurements for top RX heights only in [XML](#) or [Erydenk CSV](#) formats - 5540 files.

Link data with separate files for each RX height gain function [XML](#) or [Erydenk CSV](#) formats - 6107 files.

XML reader [sample code](#) - Note this requires libxml2 DLL.

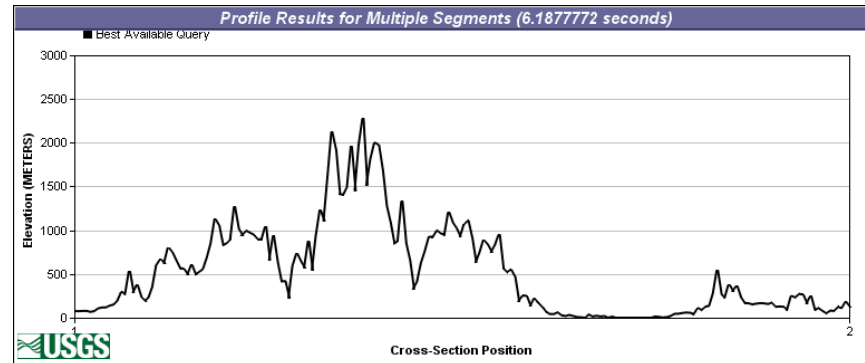
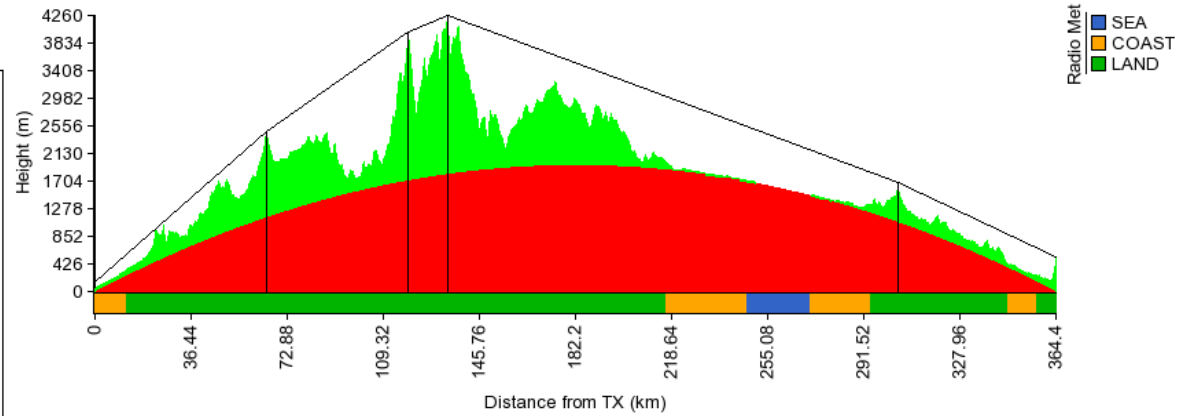
MinGW compiled [libxml2 DLL](#) and [include files](#) - Note this may not work on your system.



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	80.000	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

RADIO RESEARCH

RCRU Link Database Predictions Analysis

X Data: Suggested:
 Custom:

Y Data: Suggested:
 Custom:

Group data into sets by: Suggested: (applicable for Scatter, PDF and CDF plots)
 Custom:

SQL data constraint: Presets: ([SQLite expression syntax](#))
 Custom: (applicable for PDF, CDF and 2D PDF plots)

Number of bins:

Display key:
Display grid:
Display statistics:
Display X=Y line: (applicable for Scatter plot)

Marker style: (applicable for Scatter plot)

Marker size (pixels): (applicable for Scatter plot)

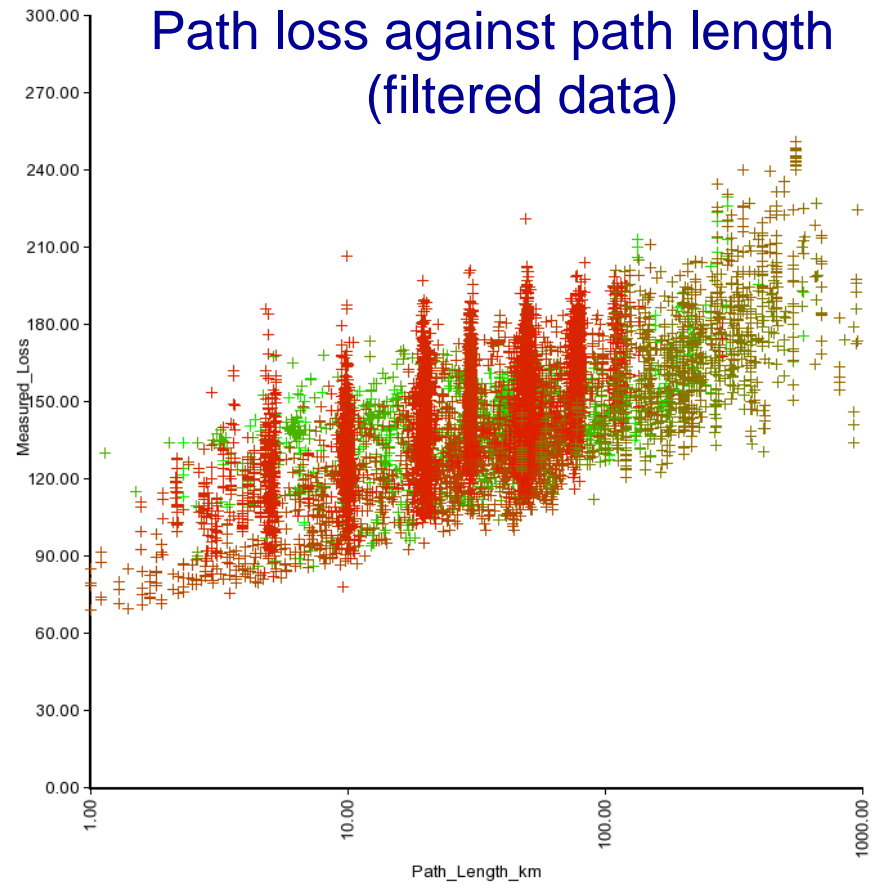
X-axis size (pixels): (applicable for Scatter plot)

Y-axis size (pixels):

Number of ticks on X-axis:
Number of ticks on Y-axis:

X-axis minimum: Automatic

X-axis maximum: Automatic



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



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
<code>DiffModel=1</code>	3-edge Deygout model as in ITU-R P.1812
<code>DiffModel=3</code>	Bullington model with LOS taper
<code>DiffModel=5</code>	3-edge Deygout model as in ITU-R P.1812 with some aspects of the US PTP cylindrical edge model
<code>DiffModel=6</code>	3-edge Deygout model as in ITU-R P.1812 with Chinese spherical Earth proposal detailed in ITU document 3K/150-E
<code>DiffModel=7</code>	Bullington model with LOS taper and Markus Liniger's distance correction [9 th order polynomial fit of $\log(\text{path_length})$ to 3-edge Deygout mean]
<code>DiffModel=8</code>	Bullington model with LOS taper and Markus Liniger's distance correction as additive term
<code>DiffModel=9</code>	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

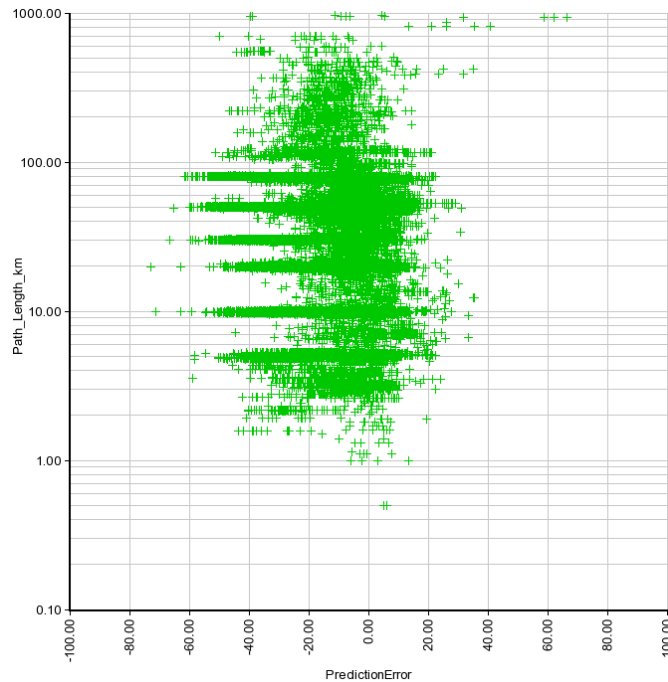
(by way of examples)



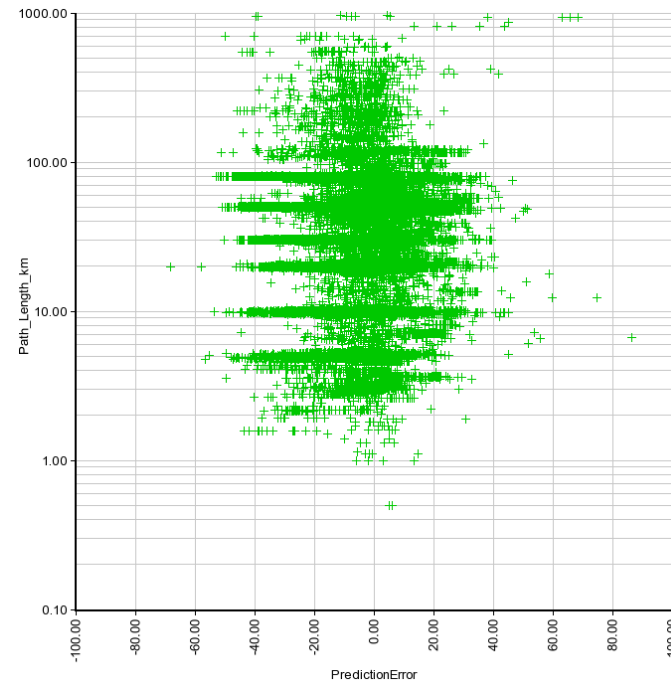
Take two models

Unfiltered data – prediction errors

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



Bullington Taper

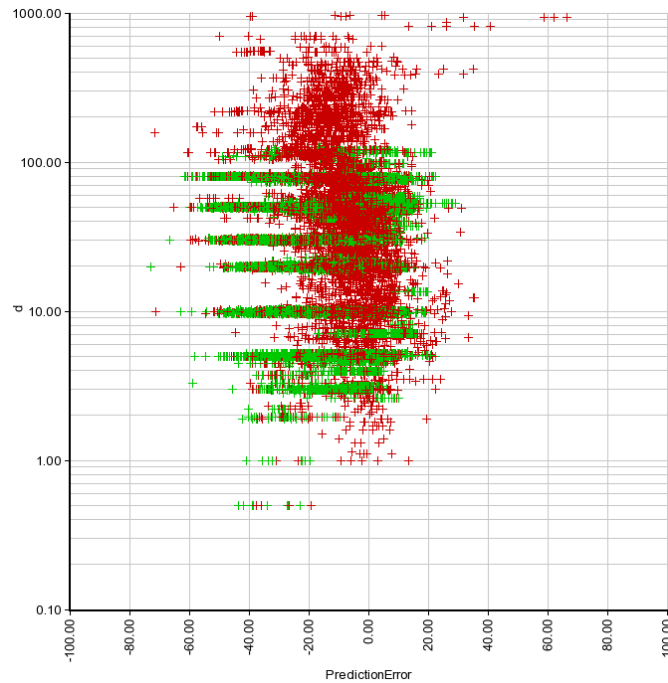


3-Edge Deygout

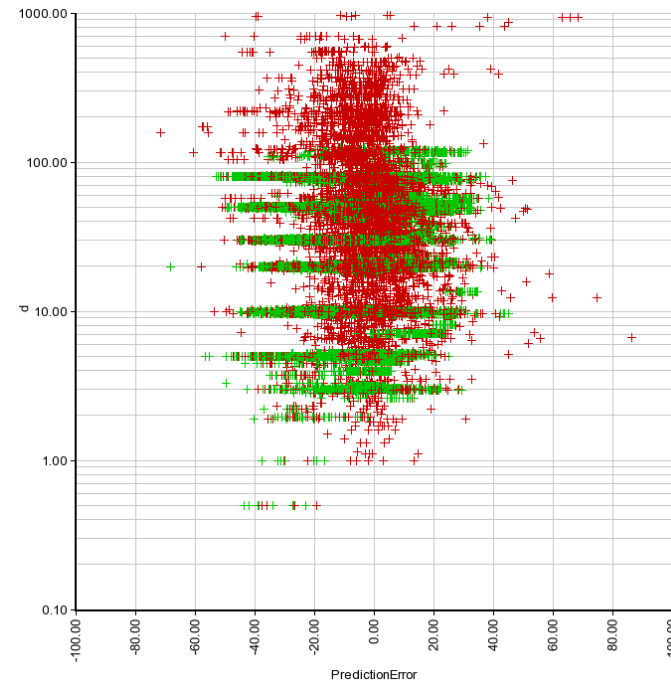


Take two models

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



Bullington Taper

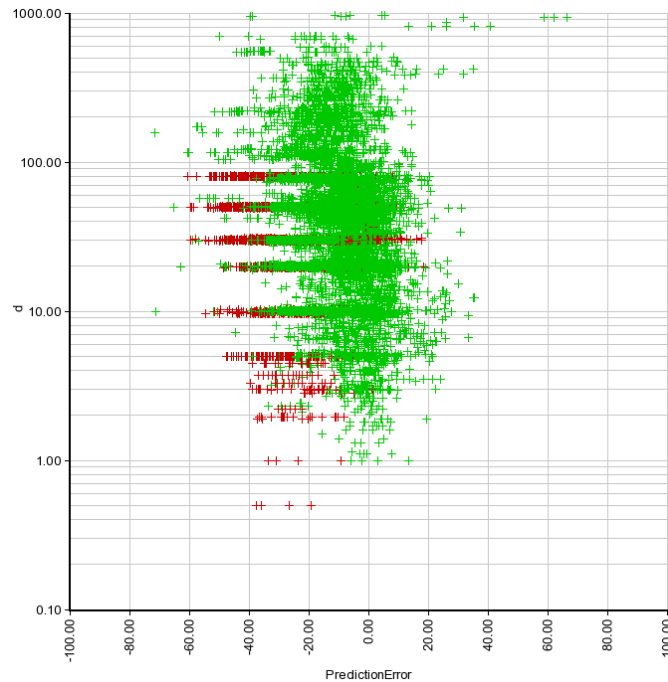


3-Edge Deygout

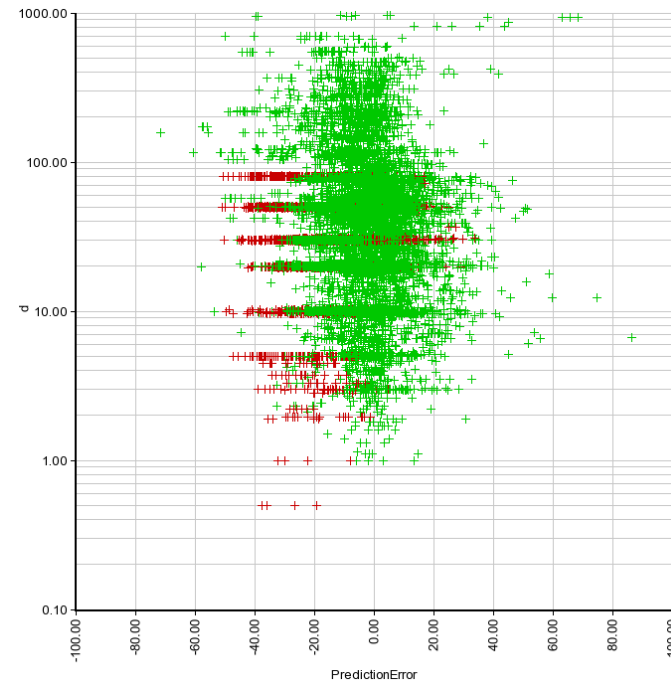


Take two models

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



Bullington Taper



3-Edge Deygout

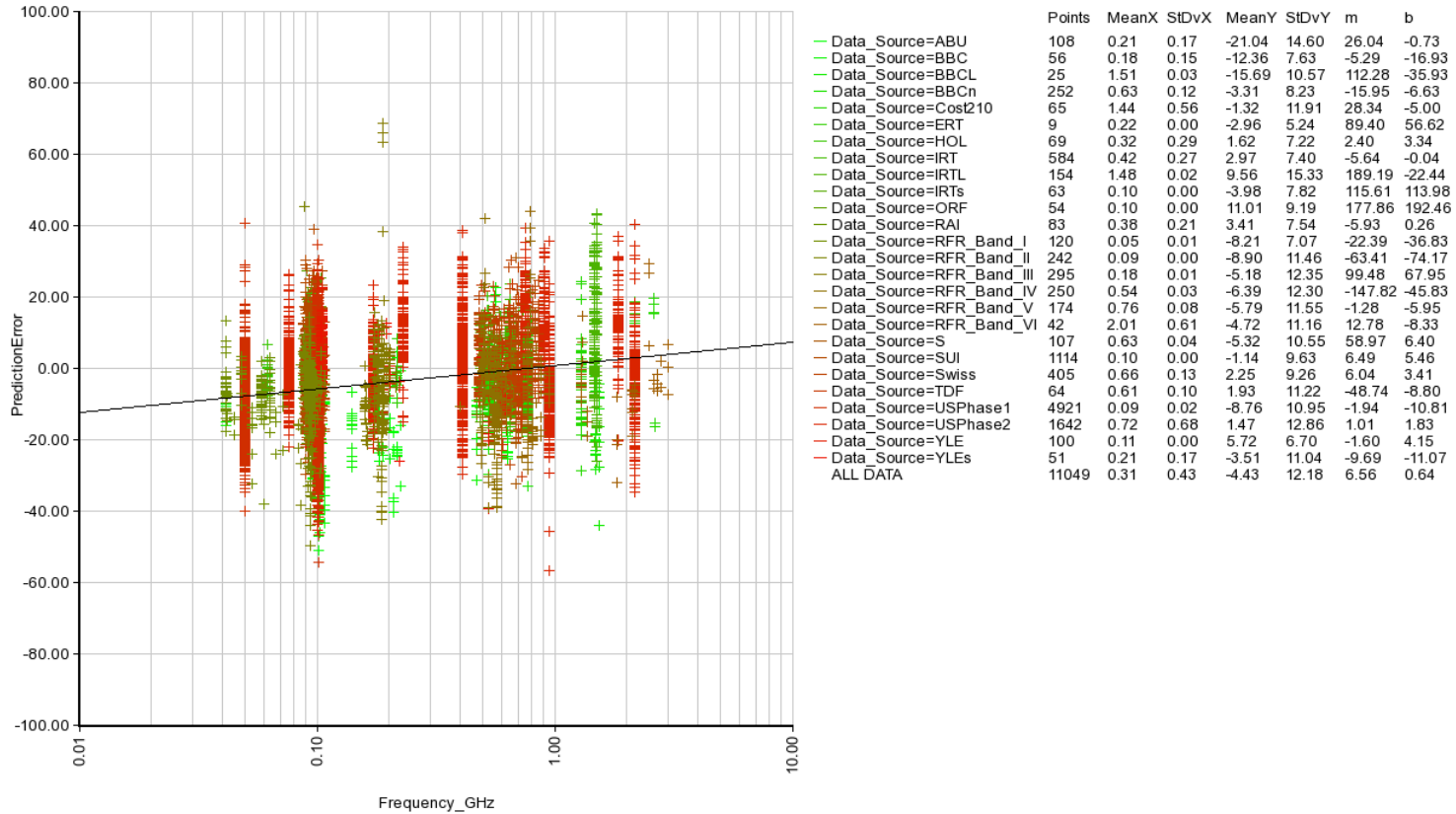
(we draw no conclusion here, but the differences are becoming clearer)



Regression fits

It is possible to make regression fits to the data

- Swiss 9 point distance correction to Bullington vs frequency

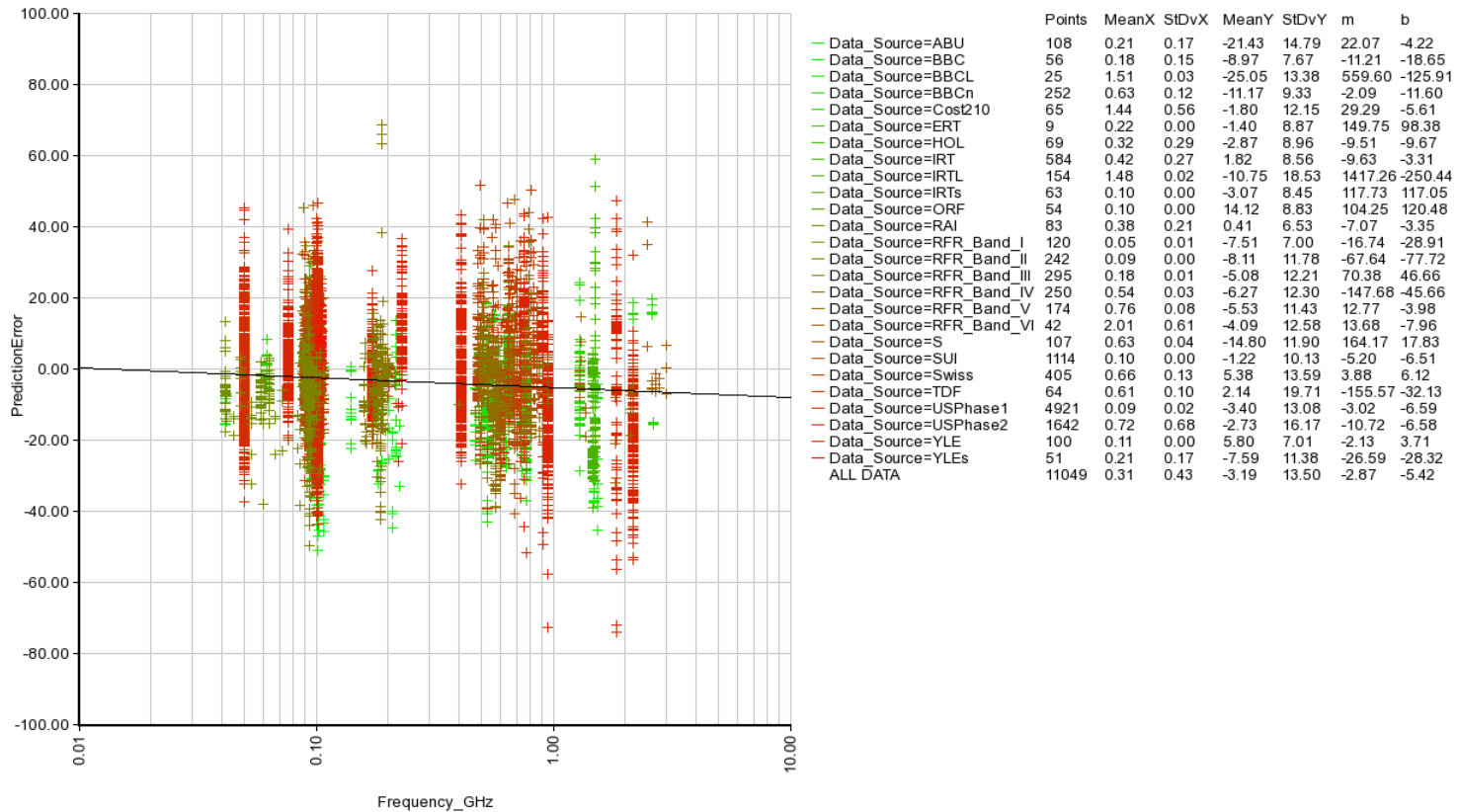




Regression fits

It is possible to make regression fits to the data

- Alternative distance correction to Bullington vs frequency



- Better at higher frequencies



Science & Technology
Facilities Council

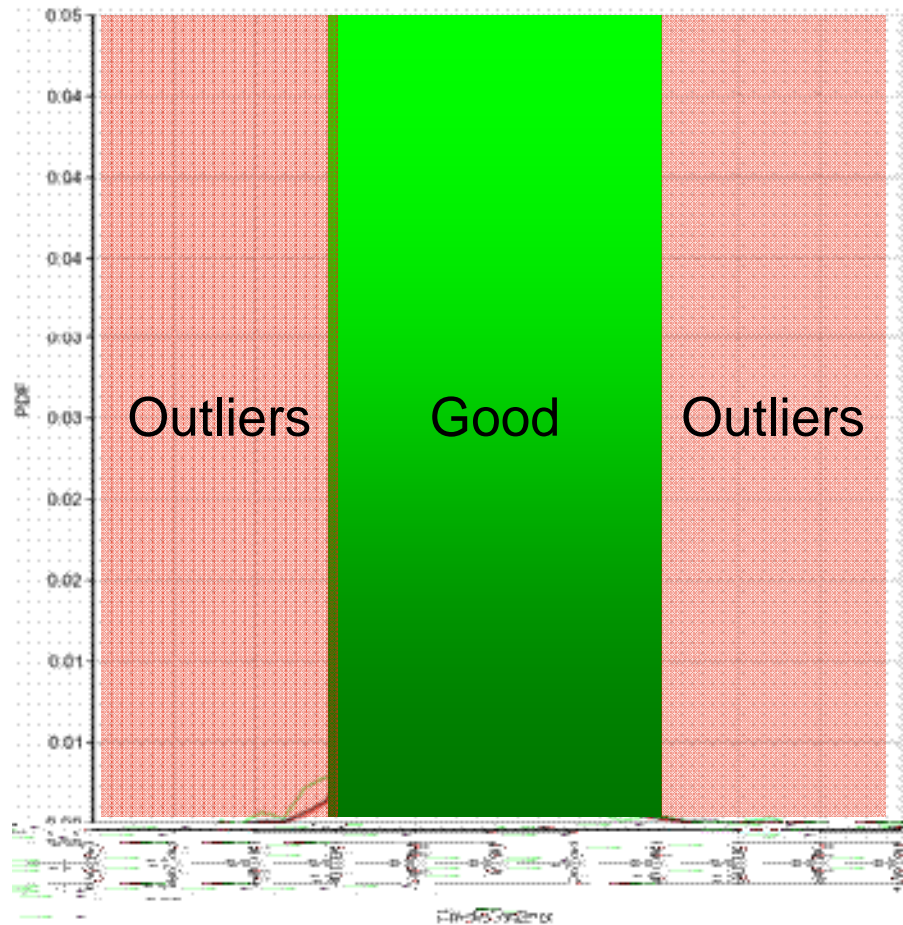
Data outliers

(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

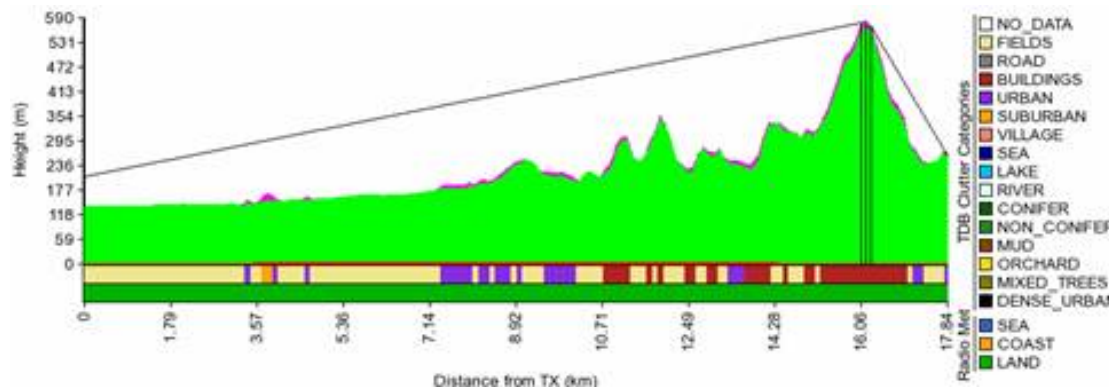


- Is the supplied profile right?
- We hope so but in this terrain it is very important, especially with low antennas at the receiver



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

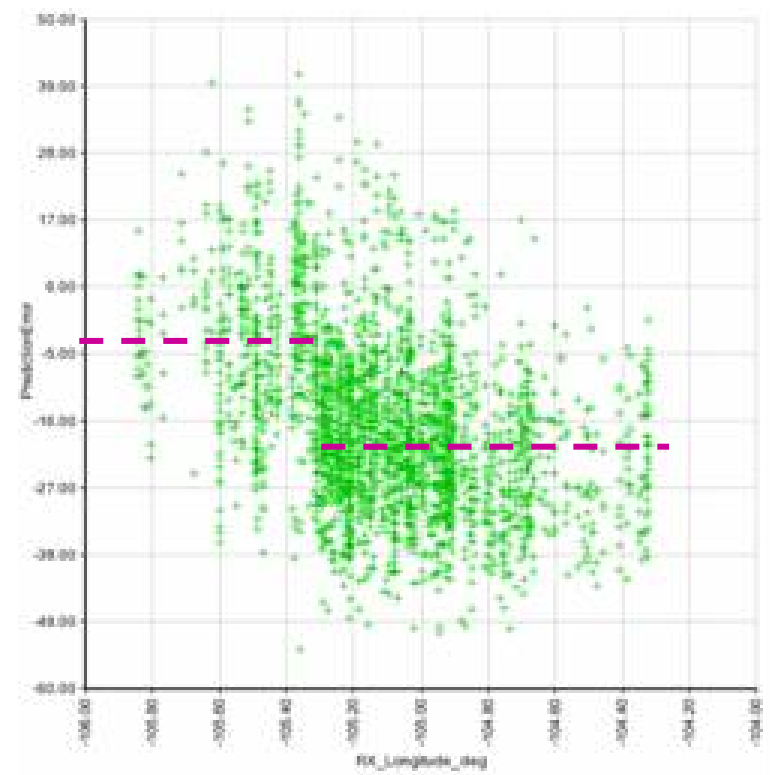
We noticed something odd with US Phase1 records

- 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



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

