

2 ASTROBIOLOGY
0 GRADUATE
1 CONFERENCE
7



CHARLOTTESVILLE, VA



1
00:00:00,790 --> 00:00:08,890

[Music]

2
00:00:15,410 --> 00:00:13,129

Thank You Jay with this wonderful title

3
00:00:16,970 --> 00:00:15,420

I better get right into it and start

4
00:00:19,849 --> 00:00:16,980

explaining what it is that I'm doing

5
00:00:22,519 --> 00:00:19,859

with texture analysis on geyser right

6
00:00:24,890 --> 00:00:22,529

deposits so the first thing what is the

7
00:00:26,750 --> 00:00:24,900

guy's right well guys rights are silica

8
00:00:29,000 --> 00:00:26,760

deposits that are found around silica

9
00:00:30,800 --> 00:00:29,010

depositing Hot Springs the top two

10
00:00:32,600 --> 00:00:30,810

images are geyser rights they were

11
00:00:35,000 --> 00:00:32,610

collected near the vent of a hot spring

12
00:00:37,490 --> 00:00:35,010

in Yellowstone National Park the one on

13
00:00:39,500 --> 00:00:37,500

the left is a specular geyser right that

14

00:00:42,740 --> 00:00:39,510

formed in the splash zone where wave

15

00:00:44,840 --> 00:00:42,750

activity be it from the surging of the

16

00:00:47,630 --> 00:00:44,850

hydrothermal fluid into the vent or when

17

00:00:49,580 --> 00:00:47,640

splashed water and it formed spicules

18

00:00:51,080 --> 00:00:49,590

which the image is cropped and so

19

00:00:54,080 --> 00:00:51,090

they're they're cut off a little bit but

20

00:00:57,139 --> 00:00:54,090

the next one on the right there at the

21

00:01:00,500 --> 00:00:57,149

top is a sub aqueous guys right that as

22

00:01:03,280 --> 00:01:00,510

the title implies it formed under the

23

00:01:06,230 --> 00:01:03,290

water under the hydrothermal fluid and

24

00:01:08,420 --> 00:01:06,240

with respect to this research we wanted

25

00:01:11,510 --> 00:01:08,430

laminated samples unfortunately we do

26
00:01:15,289 --> 00:01:11,520
have laminated samples the bottom image

27
00:01:17,630 --> 00:01:15,299
is an image of silica deposit from a

28
00:01:21,640 --> 00:01:17,640
hydrothermal vane inside of mine it too

29
00:01:24,440 --> 00:01:21,650
is composed down to the trace level of

30
00:01:26,690 --> 00:01:24,450
silica so the coloration differences

31
00:01:29,390 --> 00:01:26,700
within that sample are the trace element

32
00:01:33,109 --> 00:01:29,400
level why do we choose geyser rights

33
00:01:35,030 --> 00:01:33,119
though guys rights form as I said at the

34
00:01:37,880 --> 00:01:35,040
the high-temperature end of the hot

35
00:01:40,520 --> 00:01:37,890
spring system and evidence shows that

36
00:01:44,450 --> 00:01:40,530
life on Earth arose around hydrothermal

37
00:01:48,609 --> 00:01:44,460
systems so that was one reason also

38
00:01:50,870 --> 00:01:48,619

early life is keema autotroph

39

00:01:53,569 --> 00:01:50,880

photosynthesis had not evolved yet at

40

00:01:55,100 --> 00:01:53,579

that time and the microorganisms and the

41

00:01:58,370 --> 00:01:55,110

guys right samples are chemoautotrophs

42

00:02:00,950 --> 00:01:58,380

so that's plug number two for choosing

43

00:02:04,160 --> 00:02:00,960

geyser rights to start with also the

44

00:02:05,660 --> 00:02:04,170

microorganisms are very small they due

45

00:02:08,869 --> 00:02:05,670

to the environment they're in they're

46

00:02:11,300 --> 00:02:08,879

not able to develop large microbial mats

47

00:02:13,950 --> 00:02:11,310

which then take on characteristic

48

00:02:16,320 --> 00:02:13,960

microbial mat architectures which you

49

00:02:18,540 --> 00:02:16,330

further downstream or down the

50

00:02:20,550 --> 00:02:18,550

hydrothermal gradient so that was

51
00:02:23,640 --> 00:02:20,560
another reason why we wanted to choose

52
00:02:26,100 --> 00:02:23,650
these so with that we move on to texture

53
00:02:28,140 --> 00:02:26,110
analysis I do need to give you a brief

54
00:02:29,640 --> 00:02:28,150
highlights tour of the history we only

55
00:02:31,920 --> 00:02:29,650
have three stops not going to be too

56
00:02:35,130 --> 00:02:31,930
painful the first one we start way back

57
00:02:38,700 --> 00:02:35,140
in 1973 when Robert Herrick and his

58
00:02:42,150 --> 00:02:38,710
colleagues set out to not only define

59
00:02:44,490 --> 00:02:42,160
but characterize what human vision picks

60
00:02:47,700 --> 00:02:44,500
up on so we can identify the differences

61
00:02:50,190 --> 00:02:47,710
that we see after they determine the the

62
00:02:53,730 --> 00:02:50,200
characteristics of vision they then

63
00:02:56,130 --> 00:02:53,740

developed equations to quantify human

64

00:02:58,050 --> 00:02:56,140

vision and what allows us to see the

65

00:02:59,940 --> 00:02:58,060

differences that we see for example I in

66

00:03:02,310 --> 00:02:59,950

the carpet that we have in this room or

67

00:03:05,280 --> 00:03:02,320

in the image of the red leaves that I

68

00:03:07,470 --> 00:03:05,290

have up there on the screen so now we're

69

00:03:10,070 --> 00:03:07,480

going to dance all the way up to 1997

70

00:03:13,530 --> 00:03:10,080

and the National Institutes of Health

71

00:03:16,380 --> 00:03:13,540

released open source image processing

72

00:03:19,470 --> 00:03:16,390

software called image a while this in

73

00:03:21,720 --> 00:03:19,480

and of itself is not exactly may be

74

00:03:23,850 --> 00:03:21,730

considered a highlight it is important

75

00:03:27,720 --> 00:03:23,860

for our next step in the timeline and

76

00:03:31,770 --> 00:03:27,730

the last one which is 2003 when Julio

77

00:03:35,010 --> 00:03:31,780

Cabrera released a GL CM or the gray

78

00:03:39,390 --> 00:03:35,020

level co-occurrence matrix macro that

79

00:03:41,820 --> 00:03:39,400

works with the image J software so now

80

00:03:44,010 --> 00:03:41,830

we need to dive into the details why

81

00:03:45,980 --> 00:03:44,020

texture analysis and what questions are

82

00:03:48,090 --> 00:03:45,990

we trying to answer with this approach

83

00:03:51,030 --> 00:03:48,100

the two questions that we're trying to

84

00:03:53,190 --> 00:03:51,040

answer is first off does texture

85

00:03:54,660 --> 00:03:53,200

analysis which I will get more into more

86

00:03:56,310 --> 00:03:54,670

details don't worry you'll find out a

87

00:03:58,590 --> 00:03:56,320

little more about it but texture

88

00:04:00,690 --> 00:03:58,600

analysis we wanted to know can it

89

00:04:01,830 --> 00:04:00,700

distinguish the three different types of

90

00:04:04,710 --> 00:04:01,840

samples that we selected for this

91

00:04:06,780 --> 00:04:04,720

initial study is it going to is it going

92

00:04:10,110 --> 00:04:06,790

to clump them all together or will we

93

00:04:11,880 --> 00:04:10,120

get three distinct samples then the next

94

00:04:14,400 --> 00:04:11,890

thing is can it tell the difference

95

00:04:16,860 --> 00:04:14,410

between the different components that

96

00:04:17,490 --> 00:04:16,870

comprise a lamina and I'll also get more

97

00:04:21,840 --> 00:04:17,500

into that

98

00:04:23,790 --> 00:04:21,850

so with that for texture analysis this

99

00:04:24,770 --> 00:04:23,800

is a image that I took with a

100

00:04:26,760 --> 00:04:24,780

stereomicroscope

101
00:04:29,930 --> 00:04:26,770
take that of

102
00:04:32,969 --> 00:04:29,940
of your interest then I clip out

103
00:04:35,610 --> 00:04:32,979
sub-images is what I call them which are

104
00:04:39,570 --> 00:04:35,620
represented by those white square boxes

105
00:04:43,950 --> 00:04:39,580
up there those are 32 by 32 pixel sub

106
00:04:46,260 --> 00:04:43,960
images that were cut out after that the

107
00:04:48,480 --> 00:04:46,270
sub images need to be converted into

108
00:04:52,950 --> 00:04:48,490
grayscale I know it doesn't show very

109
00:04:54,809 --> 00:04:52,960
well on this screen it doesn't show up

110
00:04:57,360 --> 00:04:54,819
very well on my computer that I've

111
00:04:59,760 --> 00:04:57,370
actually converted it to gray level but

112
00:05:02,550 --> 00:04:59,770
it is converted to grayscale there are

113
00:05:06,240 --> 00:05:02,560

two caveats with working with the macro

114

00:05:08,550 --> 00:05:06,250

for texture analysis the one is it has

115

00:05:11,399 --> 00:05:08,560

to be a square that's based on part just

116

00:05:14,990 --> 00:05:11,409

because it's a matrix then the second

117

00:05:19,140 --> 00:05:15,000

one is the macro that Julio wrote for a

118

00:05:21,839 --> 00:05:19,150

texture analysis that is based entirely

119

00:05:24,809 --> 00:05:21,849

on the 1973 work done by Herrick and

120

00:05:27,390 --> 00:05:24,819

their equations so that's why it has to

121

00:05:29,820 --> 00:05:27,400

be grayscale because computers didn't

122

00:05:32,430 --> 00:05:29,830

handle color and we were just passed

123

00:05:35,149 --> 00:05:32,440

back in 1973 having to use punch cards

124

00:05:36,899 --> 00:05:35,159

to program computers you know so

125

00:05:37,409 --> 00:05:36,909

fortunately we've advanced a little bit

126
00:05:40,890 --> 00:05:37,419
since then

127
00:05:43,709 --> 00:05:40,900
okay so what next now we get to go to

128
00:05:46,439 --> 00:05:43,719
the macro and actually start using the

129
00:05:49,140 --> 00:05:46,449
texture analysis or the GL CM is what it

130
00:05:51,890 --> 00:05:49,150
comes up in if you install this macro on

131
00:05:54,029 --> 00:05:51,900
your computer and the image a software

132
00:05:57,899 --> 00:05:54,039
apartment primarily put this up here so

133
00:06:00,629 --> 00:05:57,909
you can see the variables that Harrell

134
00:06:03,059 --> 00:06:00,639
ik and his colleagues identified for

135
00:06:04,529 --> 00:06:03,069
looking at human vision I am NOT going

136
00:06:05,909 --> 00:06:04,539
to bore you with the details of this it

137
00:06:11,850 --> 00:06:05,919
would take the rest of the session to go

138
00:06:13,769 --> 00:06:11,860

through that so after you input what you

139

00:06:14,820 --> 00:06:13,779

want it to look at your step size you

140

00:06:17,219 --> 00:06:14,830

can choose different pixels that you

141

00:06:21,029 --> 00:06:17,229

want I always choose one because there

142

00:06:23,640 --> 00:06:21,039

are calculations that involve the

143

00:06:25,499 --> 00:06:23,650

nearest neighbor and so I choose a step

144

00:06:29,219 --> 00:06:25,509

of one just that way I don't introduce

145

00:06:32,610 --> 00:06:29,229

any artifacts into the data you click OK

146

00:06:35,040 --> 00:06:32,620

and you get a nice pop-up window with

147

00:06:37,740 --> 00:06:35,050

these values that I still have yet to

148

00:06:40,190 --> 00:06:37,750

figure out how they correlate so what is

149

00:06:42,230 --> 00:06:40,200

in the sub image but the value

150

00:06:44,300 --> 00:06:42,240

used to actually mean something I just

151
00:06:48,440 --> 00:06:44,310
haven't figured out with a high value in

152
00:06:51,230 --> 00:06:48,450
in one sub image is a low value in

153
00:06:53,990 --> 00:06:51,240
another I don't know why it does that

154
00:06:55,880 --> 00:06:54,000
exactly but put that into a spreadsheet

155
00:06:58,040 --> 00:06:55,890
after you go through all your sub images

156
00:07:00,550 --> 00:06:58,050
and all the rotations of your sub image

157
00:07:04,970 --> 00:07:00,560
you get these huge spreadsheets and

158
00:07:07,160 --> 00:07:04,980
start performing statistics our first

159
00:07:09,890 --> 00:07:07,170
quick-and-dirty test that we did just to

160
00:07:11,540 --> 00:07:09,900
find out what results are we're going to

161
00:07:13,940 --> 00:07:11,550
get you know are we going to be able to

162
00:07:15,980 --> 00:07:13,950
see the different samples are they going

163
00:07:18,200 --> 00:07:15,990

to get separated out statistically so we

164

00:07:21,620 --> 00:07:18,210

did a hierarchical hierarchical cluster

165

00:07:25,550 --> 00:07:21,630

analysis and they separated out fairly

166

00:07:27,230 --> 00:07:25,560

nicely at the top in the red as the

167

00:07:30,890 --> 00:07:27,240

legend says we have this bicular guys

168

00:07:32,900 --> 00:07:30,900

right we have one anomalous little green

169

00:07:34,880 --> 00:07:32,910

data point that made its way into the

170

00:07:36,920 --> 00:07:34,890

sub aqueous guys right and then we have

171

00:07:39,050 --> 00:07:36,930

our control the hydrothermal vane infill

172

00:07:41,120 --> 00:07:39,060

we were pretty excited at these results

173

00:07:43,850 --> 00:07:41,130

because it showed that it's separated

174

00:07:45,470 --> 00:07:43,860

out of three samples really nicely so we

175

00:07:47,750 --> 00:07:45,480

wanted to take it up a level and put all

176

00:07:50,600 --> 00:07:47,760

the data points into a k-means

177

00:07:51,980 --> 00:07:50,610

clustering analysis which is a little

178

00:07:54,140 --> 00:07:51,990

more complicated because it takes all

179

00:07:56,720 --> 00:07:54,150

the data it plugs it in and plots it and

180

00:07:57,950 --> 00:07:56,730

then it the the algorithm is an

181

00:08:00,940 --> 00:07:57,960

iterative approach and it tries to

182

00:08:03,350 --> 00:08:00,950

cluster the data based on means and

183

00:08:06,740 --> 00:08:03,360

distance from the centroid value that it

184

00:08:08,450 --> 00:08:06,750

comes up with and we we did get fairly

185

00:08:11,180 --> 00:08:08,460

good separation on the specular guy's

186

00:08:13,580 --> 00:08:11,190

right but our sub aqueous and our

187

00:08:16,460 --> 00:08:13,590

control sample they didn't get resolved

188

00:08:18,020 --> 00:08:16,470

out very well and my thoughts on this I

189

00:08:20,540 --> 00:08:18,030

have yet to talk to my advisors about it

190

00:08:22,340 --> 00:08:20,550

but my thoughts are is they formed in

191

00:08:25,420 --> 00:08:22,350

similar environments the hydrothermal

192

00:08:29,450 --> 00:08:25,430

vane and Phil it's a fluidic environment

193

00:08:31,130 --> 00:08:29,460

the sub aqueous a fluidic environment so

194

00:08:33,650 --> 00:08:31,140

I can see the similarities you know

195

00:08:38,930 --> 00:08:33,660

where they didn't get flushed out quite

196

00:08:40,880 --> 00:08:38,940

as nicely so now we get back to what was

197

00:08:43,490 --> 00:08:40,890

it that we were actually or what were

198

00:08:46,340 --> 00:08:43,500

the sub images that I collected what was

199

00:08:49,640 --> 00:08:46,350

I looking at if we break it down into

200

00:08:52,040 --> 00:08:49,650

the anatomy of the lamina I discovered

201

00:08:54,020 --> 00:08:52,050

that I didn't collect all the data

202

00:08:56,570 --> 00:08:54,030

points that I really should have on this

203

00:09:00,500 --> 00:08:56,580

on this run of data I collected the

204

00:09:04,400 --> 00:09:00,510

abiotic which is a massive dumping event

205

00:09:08,270 --> 00:09:04,410

of silica so you get these thick silica

206

00:09:10,010 --> 00:09:08,280

horizons then and I did this for

207

00:09:11,450 --> 00:09:10,020

illustrative purposes this is not

208

00:09:14,710 --> 00:09:11,460

representative of the high-temperature

209

00:09:17,980 --> 00:09:14,720

end of the system I chose to use green

210

00:09:20,170 --> 00:09:17,990

photosynthetic like filamentous type

211

00:09:22,370 --> 00:09:20,180

microorganisms for illustrative purposes

212

00:09:28,160 --> 00:09:22,380

like I said this doesn't represent the

213

00:09:30,290 --> 00:09:28,170

high temp end but you have a biotic part

214

00:09:32,420 --> 00:09:30,300

of the lamina and then we have these two

215

00:09:33,980 --> 00:09:32,430

interfaces where we have an abiotic

216

00:09:36,350 --> 00:09:33,990

surface where we had that massive

217

00:09:37,640 --> 00:09:36,360

dumping event of silica that surface

218

00:09:40,670 --> 00:09:37,650

sent has to be colonized by

219

00:09:42,620 --> 00:09:40,680

microorganisms they thrive they flourish

220

00:09:43,790 --> 00:09:42,630

their living life and then until they're

221

00:09:47,180 --> 00:09:43,800

snuffed out by another abiotic

222

00:09:49,310 --> 00:09:47,190

deposition so these two interfaces

223

00:09:53,180 --> 00:09:49,320

they're going to be a little bit

224

00:09:57,890 --> 00:09:53,190

different for this study I only captured

225

00:10:00,140 --> 00:09:57,900

the abiotic - the biotic transition we

226
00:10:03,560 --> 00:10:00,150
then thought what if we do the k-means

227
00:10:07,130 --> 00:10:03,570
analysis looking at each part of the

228
00:10:11,090 --> 00:10:07,140
lamina so we plug in the abiotic data

229
00:10:15,560 --> 00:10:11,100
points only run the k-means they they

230
00:10:19,100 --> 00:10:15,570
got clump together rather nicely stain

231
00:10:22,640 --> 00:10:19,110
for the biotic and then the one

232
00:10:24,920 --> 00:10:22,650
interface that I did so what I found out

233
00:10:27,770 --> 00:10:24,930
from one of my advisors is that

234
00:10:30,800 --> 00:10:27,780
sometimes with statistical methods you

235
00:10:33,230 --> 00:10:30,810
can get some conflicting results but

236
00:10:36,380 --> 00:10:33,240
when you figure out what's going on in

237
00:10:38,720 --> 00:10:36,390
your system and what the statistics that

238
00:10:41,420 --> 00:10:38,730

you ran actually are doing you can

239

00:10:44,180 --> 00:10:41,430

figure out where the mistakes occur so

240

00:10:47,090 --> 00:10:44,190

we did end up being able to distinguish

241

00:10:51,050 --> 00:10:47,100

the different parts of the lamina and it

242

00:10:54,530 --> 00:10:51,060

also clumped them out distinctly based

243

00:10:56,630 --> 00:10:54,540

on sample type as well so why are we

244

00:10:58,760 --> 00:10:56,640

doing this what what is the goal you

245

00:11:00,710 --> 00:10:58,770

know where are we headed obviously I

246

00:11:03,680 --> 00:11:00,720

will be incorporating additional samples

247

00:11:07,550 --> 00:11:03,690

from different environments with the

248

00:11:09,410 --> 00:11:07,560

goal of developing a computer program

249

00:11:12,560 --> 00:11:09,420

that when coupled with spatially

250

00:11:15,950 --> 00:11:12,570

correlated chemical data can help

251
00:11:17,600 --> 00:11:15,960
determine whether or not a laminated

252
00:11:21,830 --> 00:11:17,610
sedimentary structure has been

253
00:11:23,269 --> 00:11:21,840
influenced by microorganisms or not and

254
00:11:25,430 --> 00:11:23,279
with that I leave you with the summary

255
00:11:27,620 --> 00:11:25,440
of our findings set with the combination

256
00:11:29,960 --> 00:11:27,630
of texture analysis and statistical

257
00:11:31,640 --> 00:11:29,970
methods we were able to distinguish

258
00:11:47,450 --> 00:11:31,650
between the different sample types and

259
00:11:49,280 --> 00:11:47,460
the different laminate components hello

260
00:11:50,870 --> 00:11:49,290
um so I had a quick question back on

261
00:11:53,440 --> 00:11:50,880
your clustering slide if you're not

262
00:11:59,269 --> 00:11:53,450
going there real quick yeah which one

263
00:12:01,370 --> 00:11:59,279

the very first tree you had sure yeah so

264

00:12:02,960 --> 00:12:01,380

um looking at the tree structure up

265

00:12:05,810 --> 00:12:02,970

first and foremost what are what are the

266

00:12:07,610 --> 00:12:05,820

support values for your nodes if you

267

00:12:10,850 --> 00:12:07,620

have thing I don't have that information

268

00:12:13,700 --> 00:12:10,860

okay secondly when you look at the

269

00:12:16,760 --> 00:12:13,710

topology of this tree I noticed that the

270

00:12:18,890 --> 00:12:16,770

top group of your sub aqueous guys

271

00:12:21,910 --> 00:12:18,900

rights are actually grouping with the

272

00:12:25,640 --> 00:12:21,920

rest of your specular guys rights

273

00:12:27,680 --> 00:12:25,650

underneath the very top ultimately

274

00:12:30,350 --> 00:12:27,690

successful I just I did not mention

275

00:12:33,200 --> 00:12:30,360

those my apologies okay so what's going

276

00:12:36,800 --> 00:12:33,210

on with those why are they they're

277

00:12:39,680 --> 00:12:36,810

similar when you at higher magnification

278

00:12:42,350 --> 00:12:39,690

when you look at the massive dumping of

279

00:12:44,420 --> 00:12:42,360

silica in the specular geyser right

280

00:12:47,360 --> 00:12:44,430

there are also massive dumpings of

281

00:12:49,850 --> 00:12:47,370

silica in the sub aqueous and texturally

282

00:12:56,509 --> 00:12:49,860

they can look very similar okay thank

283

00:13:00,059 --> 00:12:58,289

thank you for your talk

284

00:13:03,079 --> 00:13:00,069

have you looked at any of the chemical

285

00:13:08,699 --> 00:13:03,089

profiles yet utilizing something like

286

00:13:10,589 --> 00:13:08,709

ramen or MS like mass spec work maybe

287

00:13:12,119 --> 00:13:10,599

have you correlated your work yet with

288

00:13:15,599 --> 00:13:12,129

any of the other studies on your

289

00:13:20,369 --> 00:13:15,609

stromatolite not on these I've not but I

290

00:13:24,749 --> 00:13:20,379

have done some EDS work which I know is

291

00:13:26,459 --> 00:13:24,759

not quantitative but you do get carbon

292

00:13:28,949 --> 00:13:26,469

signatures that will come out they will

293

00:13:30,929 --> 00:13:28,959

plot distinctly as well as the silica

294

00:13:34,109 --> 00:13:30,939

and then there are other trace elements

295

00:13:36,569 --> 00:13:34,119

I did find in some different work that I

296

00:13:38,429 --> 00:13:36,579

did with filamentous microorganisms

297

00:13:40,739 --> 00:13:38,439

further down in the hot spring system

298

00:13:43,199 --> 00:13:40,749

there were salt deposits around the

299

00:13:45,329 --> 00:13:43,209

filaments and so there are trace

300

00:13:47,789 --> 00:13:45,339

elements in there we have to use higher

301

00:13:54,359 --> 00:13:47,799

resolution methods to resolve that

302

00:13:55,319 --> 00:13:54,369

better good question thank you any other

303

00:13:58,469 --> 00:13:55,329

questions

304

00:13:59,870 --> 00:13:58,479

I guess if not then let's think Shayna